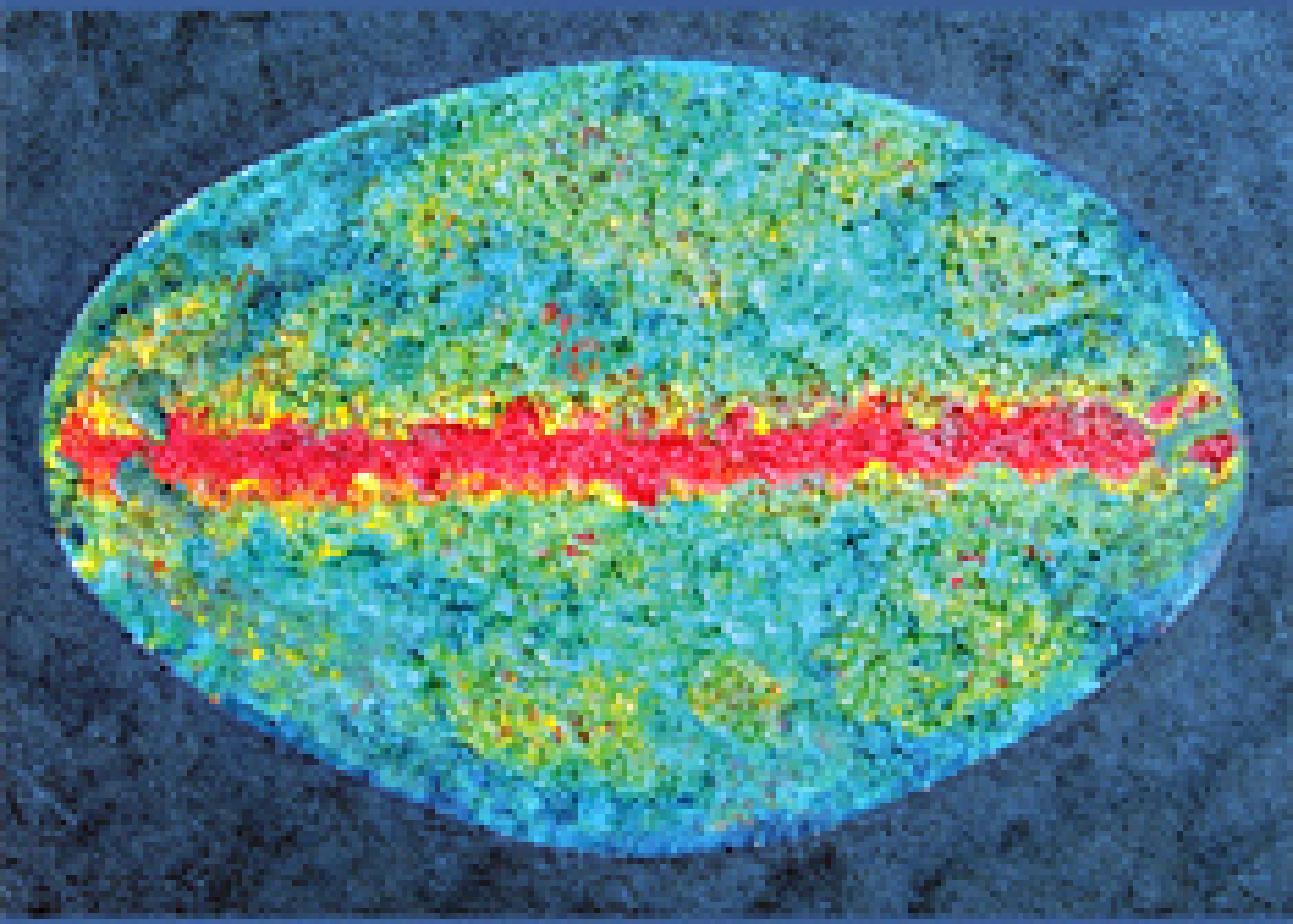
RUTH DURRER

THE Cosmic Microwave Background



CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521847049

This page intentionally left blank

THE COSMIC MICROWAVE BACKGROUND

The cosmic microwave background (CMB) is the radiation left over from the big bang. Recent analysis of the fluctuations in this radiation has given us valuable insights into our Universe and its parameters.

This textbook examines the theory of CMB and its recent progress. It starts with a brief introduction to modern cosmology and its main successes, followed by a thorough derivation of cosmological perturbation theory. It then explores the generation of initial fluctuations by inflation. In the following chapters the Boltzmann equation, which governs the evolution of CMB anisotropies, and polarization are derived using the total angular momentum method. Cosmological parameter estimation is discussed in detail. The lensing of CMB fluctuations and spectral distortions are also treated.

The book is the first to contain a full derivation of the theory of CMB anisotropies and polarization. Ideal for graduate students and researchers in this field, the textbook includes end-of-chapter exercises, and solutions to selected exercises are provided.

Ruth Durrer is Professor of Theoretical Physics at the Université de Genève. Her research focuses on the cosmic microwave background, cosmic magnetic fields and braneworld cosmology.

THE COSMIC MICROWAVE BACKGROUND

RUTH DURRER
Université de Genève

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521847049

First published in print format 2008

ISBN-13 978-0-511-42305-5 eBook (EBL)

ISBN-13 978-0-521-84704-9 hardback

To Martin, Florian, Melchior and Anna

Contents

Preface

Cosmology, the quest concerning the Universe as a whole, has been a primary interest of human study since the beginnings of mankind. For a long time our ideas about the Universe were dominated by religious beliefs – tales of creation. Only since the advent of general relativity in 1915 have we had a scientific theory at hand that might be capable of describing the Universe. Soon after Einstein's first attempt of a static universe, Hubble and collaborators (Hubble, 1929) discovered that the observable Universe is expanding. This together with the discovery of the cosmic microwave background (CMB) by Penzias and Wilson (Nobel prize 1978) has established the theory of an expanding and cooling universe which started in a 'big bang'.

For a long time observations that have led to the determination of cosmological parameters, such as the rate of expansion, the so-called Hubble parameter, the mean matter density of the Universe, or its curvature, have been very sparse and we could only determine the order of magnitude of these parameters.

During the last decade this situation has changed significantly and cosmology has entered an era of precision measurements. This major breakthrough is to a large extent due to precise measurement and analysis of the CMB. In this book I develop the theory which is used to analyse and understand measurements of the CMB, especially of its anisotropies and polarization, but also its frequency spectrum. The Nobel prize was awarded to George Smoot and John Mather, in 2006, for the discovery of these anisotropies and for precise measurements of the CMB spectrum.

The book is directed mainly towards graduate students and researchers who want to obtain an overview of the main developments in CMB physics, and who want to understand the state-of-the-art techniques which are used to analyse CMB data. I believe that the theory of CMB physics is now sufficiently mature for a book on this topic to be useful. I shall not enter into any details concerning CMB experiments. This is by no means because I consider them less interesting, but rather that they

are still in full development and will hopefully make significant progress in the near future. Of course, my background is also that of a theoretical physicist and my main interest lies in the theoretical aspects of CMB physics. I hope, however, that this book will also be useful to CMB experimentalists who want to know what happens inside their cosmic parameter estimation routines.

It is assumed that the reader is familiar with undergraduate physics including the basics of general relativity, and has an elementary knowledge of quantum field theory and particle physics. The beauty of cosmology lies in the fact that it employs more or less all fields of physics starting with general relativity over thermodynamics and statistical physics to electrodynamics, quantum mechanics and particle physics. In this book I do not want to present an introduction to these topics as well since, first of all, there exist wonderful textbooks on all of them and second you have learned them in your undergraduate physics courses.

Before we start, let me sketch the content of the different chapters and give you a guide on how to read this book.

The first chapter is an overview of the homogeneous and isotropic universe. We present and discuss the Friedmann equations, recombination, nucleosynthesis and inflation. Readers familiar with cosmology may skip this chapter or just skim it.

In Chapter 2 we develop cosmological perturbation theory. This is the basics of CMB physics. The main reason why the CMB allows such an accurate determination of cosmological parameters lies in the fact that its anisotropies are small and can be determined within first-order perturbation theory. In Fourier space the linear perturbation equations become a series of ordinary linear differential equations, which can be solved numerically to high precision without any difficulty. We derive the perturbations of Einstein's equations and the energy–momentum conservation equations and solve them for simple but relevant cases. We also discuss the perturbation equation for light-like geodesics. This is sufficient to calculate the CMB anisotropies in the so-called instant recombination approximation. The main physical effects which are missed in such a treatment are Silk damping on small scales and polarization. We then introduce the CMB power spectrum and draw our first conclusions for its dependence on cosmological and primordial parameters. For example, we derive an approximate formula for the position of the acoustic peaks. An experimentalist mainly interested in parameter estimation may jump, after Chapter 2, directly to Chapter 6 and skip the more theoretical parts between.

The third chapter is devoted to the initial condition. There we explain how the unavoidable quantum fluctuations are amplified during an inflationary phase and lead to a nearly scale-invariant spectrum of scalar and tensor perturbations. We also discuss the initial conditions for mixed adiabatic and iso-curvature perturbations.

In Chapter 4 we derive the perturbed Boltzmann equation for CMB photons. After a brief introduction to relativistic kinetic theory, we first derive the Liouville

equation, i.e. the Boltzmann equation without a collision term. We also discuss the connection between the distribution function and the energy–momentum tensor. We then derive the collision term, i.e. the right-hand side of the Boltzmann equation, due to Thomson scattering of photons and electrons. In this first attempt we neglect the polarization dependence of Thomson scattering. The chapter ends with a list of the full system of perturbation equations for a ΛCDM universe.

In Chapter 5 we discuss polarization. Here we derive the total angular momentum method that is perfectly adapted to the problem of CMB anisotropies and polarization, taking into account its symmetry, which allows a decomposition into modes with fixed total angular momentum. The representation theory of the rotation group and the spin weighted spherical harmonics which are extensively used in this chapter are deferred to an appendix. We interpret some results using the flat sky approximation, which is valid on small angular scales.

Chapter 6 is devoted to parameter estimation. We first discuss the physical dependence of CMB anisotropies on cosmological parameters. After a section on CMB data we then treat in some detail statistical methods for CMB data analysis. We discuss especially the Fisher matrix and explain Markov chain Monte Carlo methods. We also address degeneracies, combinations of cosmological parameters on which CMB anisotropies do not, or only very weakly, depend. Because of these degeneracies, cosmological parameter estimation also makes use of other, non-CMB related, observations. We summarize them in a separate section. We finish the chapter with a discussion of 'sources', i.e. inhomogeneously distributed contributions to the energy–momentum tensor, such as topological defects, which may also contribute to the CMB anisotropies and thereby affect the estimated cosmological parameters.

In Chapter 7 we treat lensing of CMB anisotropies and polarization. This second-order effect is especially important on small scales but also has to be taken into account for $\ell \gtrsim 500$ if we want to achieve an accuracy of better than 0.5%. We first derive the deflection angle and the lensing power spectrum. Then we discuss lensing of CMB fluctuations and polarization in the flat sky approximation, which is sufficiently accurate for angular harmonics with $\ell \gtrsim 50$. We conclude the chapter with an overview on other second-order effects.

In the final chapter spectral distortions of the CMB are discussed. We first introduce the three relevant collision processes in a universe with photons and non-relativistic electrons: elastic Compton scattering, Bremsstrahlung and double Compton scattering. We derive the corresponding collision terms and Boltzmann equations. For elastic Compton scattering this leads us to the Kompaneets equation for which we present a detailed derivation. We introduce the timescales corresponding to these three collision processes and determine at which redshift a given process freezes – becomes slower than cosmic expansion. Finally, we discuss the

possible generation of a chemical potential in the CMB spectrum and the Sunyaev–Zel'dovich effect.

All chapters are complemented with some exercises at the end.

In the appendices we collect useful constants and formulae, information on special functions and some more technical derivations. The solutions to a selection of exercises are also given in an appendix.

This book has grown out of a graduate course on CMB anisotropies that I have given on several occasions. Thanks are due to the students of these courses, who have motivated me to write it up in the form of a textbook. I am also indebted to many collaborators and colleagues with whom I have discussed various aspects of the book and who have helped me to clarify many issues. Especially I want to mention Chiara Caprini, Martin Kunz, Toni Riotto, Uros Seljak and Norbert Straumann. I am also immensely grateful to students and colleagues who have read parts of the draft and helped me correct numerous typographical errors and other mistakes: Camille Bonvin, Jean-Pierre Eckmann, Alice Gasparini, Sandro Scodeller and others. Of course all the remaining mistakes are entirely my responsibility. Marcus Ruser and Martin Kunz have also helped me with some of the figures. I also wish to thank Susan Staggs who provided me with a most useful dataset of the CMB spectrum.

Ruth Durrer

1

The homogeneous and isotropic universe

Notation

In this book we denote the derivative with respect to physical time by a prime, and the derivative with respect to conformal time by a dot,

$$\tau = \text{physical (cosmic) time} \qquad \frac{dX}{d\tau} \equiv X' \,, \tag{1.1}$$

$$t = \text{ conformal time} \qquad \frac{dX}{dt} \equiv \dot{X} \,. \tag{1.2}$$

Spatial 3-vectors are denoted by a bold face symbol such as $\mathbf{k}$ or $\mathbf{x}$ whereas four-dimensional spacetime vectors are denoted as $x = (x^\mu)$.

We use the metric signature $(-, +, +, +)$ throughout the book.

The Fourier transform is defined by

$$f(\mathbf{k}) = \int d^3x \; f(\mathbf{x})\, e^{i\mathbf{k}\cdot\mathbf{x}} \,, \tag{1.3}$$

so that

$$f(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3k \; f(\mathbf{k})\, e^{-i\mathbf{k}\cdot\mathbf{x}} \,. \tag{1.4}$$

We use the same letter for $f(\mathbf{x})$ and for its Fourier transform $f(\mathbf{k})$. The spectrum $P_f(k)$ of a statistically homogeneous and isotropic random variable f is given by

$$\langle f(\mathbf{k}) f^*(\mathbf{k}') \rangle = (2\pi)^3 \, \delta(\mathbf{k} - \mathbf{k}') P_f(k) \,. \tag{1.5}$$

Since it is isotropic, $P_f(k)$ is a function only of the modulus $k = |\mathbf{k}|$. If f is Gaussian, the Dirac delta function implies that different $\mathbf{k}$'s are uncorrelated.

Throughout this book we use units where the speed of light, c, Planck's constant, $\hbar$ and Boltzmann's constant, k_B are unity, $c = \hbar = k_B = 1$. Length and time therefore have the same units and energy, mass and momentum also have the same units, which are inverse to the unit of length. Temperature has the same units as energy.

We may use cm^{-1} to measure energy, mass, temperature, or eV^{-1} to measure distances or times. We shall use whatever unit is convenient to discuss a given problem. Conversion factors can be found in Appendix 1.

1.1 Homogeneity and isotropy

Modern cosmology is based on the hypothesis that our Universe is to a good approximation homogeneous and isotropic on sufficiently large scales. This relatively bold assumption is often called the 'cosmological principle'. It is an extension of the Copernican principle stating that not only should our place in the solar system not be a special one, but also that the position of the Milky Way in the Universe should be in no way statistically distinguishable from the position of other galaxies. Furthermore, no direction should be distinguished. The Universe looks statistically the same in all directions. This, together with the hypothesis that the matter density and geometry of the Universe are smooth functions of the position, implies homogeneity and isotropy on sufficiently large scales. Isotropy around each point together with analyticity actually already implies homogeneity of the Universe.[1] A formal proof of this quite intuitive result can be found in Straumann (1974).

But which scale is 'sufficiently large'? Certainly not the solar system or our galaxy. But also not the size of galaxy clusters. (In cosmology, distances are usually measured in Mpc (Megaparsec). 1 Mpc $= 3.2615 \times 10^6$ light years $= 3.0856 \times 10^{24}$ cm is a typical distance between galaxies, the distance between our neighbour Andromeda and the Milky Way is about 0.7 Mpc. These and other connections between frequently used units can be found in Appendix 1.)

It turns out that the scale at which the *galaxy distribution* becomes homogeneous is difficult to determine. From the analysis of the Sloan Digital Sky Survey (SDSS) it has been concluded that the irregularities in the galaxy density are still on the level of a few per cent on scales of 100 h^{-1} Mpc (Hogg *et al.*, 2005). Fortunately, we know that the *geometry* of the Universe shows only small deviations from the homogeneous and isotropic background, already on scales of a few Mpc. The geometry of the Universe can be tested with the peculiar motion of galaxies, with lensing, and in particular with the cosmic microwave background (CMB).

The small deviations from homogeneity and isotropy in the CMB are of uttermost importance since, most probably, they represent the 'seeds', which, via gravitational instability, have led to the formation of large-scale structure, galaxies and eventually solar systems with planets that support life in the Universe.

[1] If 'analyticity' is not assumed, the matter distribution could also be fractal and still statistically isotropic around each point. For a detailed elaboration of this idea and its comparison with observations see Sylos Labini *et al.* (1998).

Furthermore, we suppose that the initial fluctuations needed to trigger the process of gravitational instability stem from tiny quantum fluctuations that have been amplified during a period of inflationary expansion of the Universe. I consider this connection of the microscopic quantum world with the largest scales of the Universe to be of breathtaking philosophical beauty.

In this chapter we investigate the background Universe. We shall first discuss the geometry of a homogeneous and isotropic spacetime. Then we investigate two important events in the thermal history of the Universe. Finally, we study the paradigm of inflation. This chapter lays the basis for the following ones where we shall investigate *fluctuations* on the background, most of which can be treated in first-order perturbation theory.

1.2 The background geometry of the Universe

1.2.1 The Friedmann equations

In this section we assume a basic knowledge of general relativity. The notation and sign convention for the curvature tensor that we adopt are specified in Appendix A2.1.

Our Universe is described by a four-dimensional spacetime $(\mathcal{M}, g)$ given by a pseudo-Riemannian manifold $\mathcal{M}$ with metric g. A homogeneous and isotropic spacetime is one that admits a slicing into homogeneous and isotropic, i.e., maximally symmetric, 3-spaces. There is a preferred geodesic time coordinate τ, called 'cosmic time' such that the 3-spaces of constant time, $\Sigma_\tau = \{\mathbf{x}|(\tau, \mathbf{x}) \in \mathcal{M}\}$ are maximally symmetric spaces, hence spaces of constant curvature. The metric g is therefore of the form

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = -d\tau^2 + a^2(\tau)\gamma_{ij}\, dx^i\, dx^j\,. \tag{1.6}$$

The function $a(\tau)$ is called the scale factor and γ_{ij} is the metric of a 3-space of constant curvature K. Depending on the sign of K this space is locally isometric to a 3-sphere ($K > 0$), a three-dimensional pseudo-sphere ($K < 0$) or flat, Euclidean space ($K = 0$). In later chapters of this book we shall mainly use 'conformal time' t defined by $a\, dt = d\tau$, so that

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = a^2(t)\left(-dt^2 + \gamma_{ij}\, dx^i\, dx^j\right)\,. \tag{1.7}$$

The geometry and physics of homogeneous and isotropic solutions to Einstein's equations was first investigated mathematically in the early twenties by Friedmann (1922) and physically as a description of the observed expanding Universe in 1927

by Lemaître.[2] Later, Robertson (1936), Walker (1936) and others rediscovered the Friedmann metric and studied several additional aspects. However, since we consider the contributions by Friedmann and Lemaître to be far more fundamental than the subsequent work, we shall call a homogeneous and isotropic solution to Einstein's equations a 'Friedmann–Lemaître universe' (FL universe) in this book.

It is interesting to note that the Friedmann solution breaks Lorentz invariance. Friedmann universes are not invariant under boosts, there is a preferred cosmic time τ, the proper time of an observer who sees a spatially homogeneous and isotropic universe. Like so often in physics, the Lagrangian and therefore also the field equations of general relativity are invariant under Lorentz transformations, but a specific solution in general is not. In that sense we are back to Newton's vision of an absolute time. But on small scales, e.g. the scale of a laboratory, this violation of Lorentz symmetry is, of course, negligible.

The topology is not determined by the metric and hence by Einstein's equations. There are many compact spaces of negative or vanishing curvature (e.g. the torus), but there are no infinite spaces with positive curvature. A beautiful treatment of the fascinating, but difficult, subject of the topology of spaces with constant curvature and their classification is given in Wolf (1974). Its applications to cosmology are found in Lachieze-Rey & Luminet (1995).

Forms of the metric γ, which we shall often use are

$$\gamma_{ij}\,dx^i\,dx^j = \frac{\delta_{ij}\,dx^i\,dx^j}{(1+\frac{1}{4}K\rho^2)^2}\,, \tag{1.8}$$

$$\gamma_{ij}\,dx^i\,dx^j = dr^2 + \chi^2(r)\left(d\theta^2 + \sin^2(\theta)\,d\varphi^2\right)\,, \tag{1.9}$$

$$\gamma_{ij}\,dx^i\,dx^j = \frac{dR^2}{1-KR^2} + R^2\left(d\theta^2 + \sin^2(\theta)\,d\varphi^2\right)\,, \tag{1.10}$$

where in Eq. (1.8)

$$\rho^2 = \sum_{i,j=1}^{3}\delta_{ij}x^i x^j\,, \quad \text{and} \quad \delta_{ij} = \begin{cases} 1 & \text{if } i=j\,,\\ 0 & \text{else}\,, \end{cases} \tag{1.11}$$

and in Eq. (1.9)

$$\chi(r) = \begin{cases} r & \text{in the Euclidean case,} \quad K=0\,,\\ \frac{1}{\sqrt{K}}\sin(\sqrt{K}r) & \text{in the spherical case,} \quad K>0\,,\\ \frac{1}{\sqrt{|K|}}\sinh(\sqrt{|K|}r) & \text{in the hyperbolic case,} \quad K<0\,. \end{cases} \tag{1.12}$$

Often one normalizes the scale factor such that $K=\pm 1$ whenever $K\neq 0$. One has, however, to keep in mind that in this case r and K become dimensionless and the

[2] In the English translation of (Lemaître, 1927) from 1931 Lemaître's somewhat premature but pioneering arguments that the observed Universe is actually expanding have been omitted.

scale factor a has the dimension of length. If $K = 0$ we can normalize a arbitrarily. We shall usually normalize the scale factor such that $a_0 = 1$ and the curvature is not dimensionless. The coordinate transformations which relate these coordinates are determined in Ex. 1.1.

Due to the symmetry of spacetime, the energy–momentum tensor can only be of the form

$$(T_{\mu\nu}) = \begin{pmatrix} -\rho g_{00} & \mathbf{0} \\ \mathbf{0} & P g_{ij} \end{pmatrix} . \tag{1.13}$$

There is no additional assumption going into this ansatz, such as the matter content of the Universe being an ideal fluid. It is a simple consequence of homogeneity and isotropy and is also verified for scalar field matter, a viscous fluid or free-streaming particles in a FL universe. As usual, the energy density ρ and the pressure P are defined as the time- and space-like eigenvalues of (T^{μ}_{ν}).

The Einstein tensor can be calculated from the definition (A2.12) and Eqs. (A2.31)–(A2.38),

$$G_{00} = 3\left[\left(\frac{a'}{a}\right)^2 + \frac{K}{a^2}\right] \qquad \text{(cosmic time)}\,, \tag{1.14}$$

$$G_{ij} = -\left(2a''a + a'^2 + K\right)\gamma_{ij} \qquad \text{(cosmic time)}\,, \tag{1.15}$$

$$G_{00} = 3\left[\left(\frac{\dot{a}}{a}\right)^2 + K\right] \qquad \text{(conformal time)}\,, \tag{1.16}$$

$$G_{ij} = -\left(2\left(\frac{\dot{a}}{a}\right)^{\bullet} + \left(\frac{\dot{a}}{a}\right)^2 + K\right)\gamma_{ij} \qquad \text{(conformal time)}\,. \tag{1.17}$$

The Einstein equations relate the Einstein tensor to the energy–momentum content of the Universe via $G_{\mu\nu} = 8\pi G T_{\mu\nu} - g_{\mu\nu}\Lambda$. Here Λ is the so-called cosmological constant. In a FL universe the Einstein equations become

$$\left(\frac{a'}{a}\right)^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} \qquad \text{(cosmic time)}\,, \tag{1.18}$$

$$2\frac{a''}{a} + \frac{(a')^2}{a^2} + \frac{K}{a^2} = -8\pi G P + \Lambda \qquad \text{(cosmic time)}\,, \tag{1.19}$$

$$\left(\frac{\dot{a}}{a}\right)^2 + K = \frac{8\pi G}{3}a^2\rho + \frac{a^2\Lambda}{3} \qquad \text{(conformal time)}\,, \tag{1.20}$$

$$2\left(\frac{\dot{a}}{a}\right)^{\bullet} + \left(\frac{\dot{a}}{a}\right)^2 + K = -8\pi G a^2 P + a^2\Lambda \qquad \text{(conformal time)}\,. \tag{1.21}$$

Energy 'conservation', $T^{\mu\nu}_{;\mu} = 0$ yields

$$\dot{\rho} = -3(\rho + P)\left(\frac{\dot{a}}{a}\right) \quad \text{or, equivalently} \quad \rho' = -3(\rho + P)\left(\frac{a'}{a}\right). \tag{1.22}$$

This equation can also be obtained by differentiating Eq. (1.18) or (1.20) and inserting (1.19) or (1.21); it is a consequence of the contracted Bianchi identities (see Appendix A2.1). Eqs. (1.18)–(1.21) are the Friedmann equations. The quantity

$$H(\tau) \equiv \frac{a'}{a} = \frac{\dot{a}}{a^2} \equiv \mathcal{H}a^{-1}, \tag{1.23}$$

is called the Hubble rate or the Hubble parameter, where $\mathcal{H}$ is the comoving Hubble parameter. At present, the Universe is expanding, so that $H_0 > 0$. We parametrize it by

$$H_0 = 100\, h\ \mathrm{km\,s^{-1}\,Mpc^{-1}} \simeq 3.241 \times 10^{-18}\, h\ \mathrm{s^{-1}} \simeq 1.081 \times 10^{-28}\, h\ \mathrm{cm^{-1}}.$$

Observations show (Freedman *et al.*, 2001) that $h \simeq 0.72 \pm 0.1$. Eq. (1.22) is easily solved in the case $w = P/\rho =$ constant. Then one finds

$$\rho = \rho_0 (a_0/a)^{3(1+w)}, \tag{1.24}$$

where ρ_0 and a_0 denote the value of the energy density and the scale factor at present time, τ_0. In this book cosmological quantities indexed by a '0' are evaluated today, $X_0 = X(\tau_0)$. For non-relativistic matter, $P_m = 0$, we therefore have $\rho_m \propto a^{-3}$ while for radiation (or any kind of massless particles) $P_r = \rho_r/3$ and hence $\rho_r \propto a^{-4}$. A cosmological constant corresponds to $P_\Lambda = -\rho_\Lambda$ and we obtain, as expected $\rho_\Lambda =$ constant. If the curvature K can be neglected and the energy density is dominated by one component with $w =$ constant, inserting Eq. (1.24) into the Friedmann equations yields the solutions

$$a \propto \tau^{2/3(1+w)} \propto t^{2/(1+3w)} \qquad w = \text{constant} \neq -1, \tag{1.25}$$

$$a \propto \tau^{2/3} \propto t^2 \qquad w = 0, \quad \text{(dust)}, \tag{1.26}$$

$$a \propto \tau^{1/2} \propto t \qquad w = 1/3, \quad \text{(radiation)}, \tag{1.27}$$

$$a \propto \exp(H\tau) \propto 1/|t| \qquad w = -1, \quad \text{(cosmol. const.)}. \tag{1.28}$$

It is interesting to note that if $w < -1$, so-called 'phantom matter', we have to choose $\tau < 0$ to obtain an expanding universe and the scale factor diverges in finite time, at $\tau = 0$. This is the so-called 'big rip'. Phantom matter has many problems but it is discussed in connection with the supernova type 1a (SN1a) data, which are compatible with an equation of state with $w < -1$ or with an ordinary cosmological constant (Caldwell *et al.*, 2003). For $w < -\frac{1}{3}$ the time coordinate t has to be chosen as negative for the Universe to expand and spacetime cannot be

continued beyond $t = 0$. But $t = 0$ corresponds to a cosmic time, the proper time of a static observer, $\tau = \infty$; this is not a singularity. (The geodesics can be continued until affine parameter ∞.)

We also introduce the adiabatic sound speed c_s determined by

$$c_s^2 = \frac{P'}{\rho'} = \frac{\dot{P}}{\dot{\rho}} \,. \tag{1.29}$$

From this definition and Eq. (1.22) it is easy to see that

$$\dot{w} = 3\mathcal{H}(1+w)\big(w - c_s^2\big) \,. \tag{1.30}$$

Hence $w =$ constant if and only if $w = c_s^2$ or $w = -1$. Note that already in a simple mixture of matter and radiation $w \neq c_s^2 \neq$ constant (see Ex. 1.3).

Eq. (1.18) implies that for a critical value of the energy density given by

$$\rho(\tau) = \rho_c(\tau) = \frac{3H^2}{8\pi G} \tag{1.31}$$

the curvature and the cosmological constant vanish. The value ρ_c is called the critical density. The ratio $\Omega_X = \rho_X/\rho_c$ is the 'density parameter' of the component X. It indicates the fraction that the component X contributes to the expansion of the Universe. We shall make use especially of

$$\Omega_r \equiv \Omega_r(\tau_0) = \frac{\rho_r(\tau_0)}{\rho_c(\tau_0)} \,, \tag{1.32}$$

$$\Omega_m \equiv \Omega_m(\tau_0) = \frac{\rho_m(\tau_0)}{\rho_c(\tau_0)} \,, \tag{1.33}$$

$$\Omega_K \equiv \Omega_K(\tau_0) = \frac{-K}{a_0^2 H_0^2} \,, \tag{1.34}$$

$$\Omega_\Lambda \equiv \Omega_\Lambda(\tau_0) = \frac{\Lambda}{3H_0^2} \,. \tag{1.35}$$

1.2.2 The 'big bang' and 'big crunch' singularities

We can absorb the cosmological constant into the energy density and pressure by redefining

$$\rho_{\text{eff}} = \rho + \frac{\Lambda}{8\pi G} \,, \qquad P_{\text{eff}} = P - \frac{\Lambda}{8\pi G} \,.$$

Since Λ is a constant and $\rho_{\text{eff}} + P_{\text{eff}} = \rho + P$, the conservation equation (1.22) still holds. A first interesting consequence of the Friedmann equations is obtained

when subtracting Eq. (1.18) from (1.19). This yields

$$\frac{a''}{a} = -\frac{4\pi G}{3}(\rho_{\rm eff} + 3P_{\rm eff})\,. \tag{1.36}$$

Hence if $\rho_{\rm eff} + 3P_{\rm eff} > 0$, the Universe is decelerating. Furthermore, Eqs. (1.22) and (1.36) then imply that in an expanding and decelerating universe

$$\frac{\rho'_{\rm eff}}{\rho_{\rm eff}} < -2\frac{a'}{a}\,,$$

so that ρ decays faster than $1/a^2$. If the curvature is positive, $K > 0$, this implies that at some time in the future, $\tau_{\rm max}$, the density has dropped down to the value of the curvature term, $K/a^2(\tau_{\rm max}) = 8\pi G\rho_{\rm eff}(\tau_{\rm max})$. Then the Universe stops expanding and recollapses. Furthermore, this is independent of curvature, as a' decreases the curve $a(\tau)$ is concave and thus cuts the $a = 0$ line at some finite time in the past. This moment of time is called the 'big bang'. The spatial metric vanishes at this value of τ, which we usually choose to be $\tau = 0$; and spacetime cannot be continued to earlier times. This is not a coordinate singularity. From the Ricci tensor given in Eqs. (A2.31) and (A2.32) one obtains the Riemann scalar

$$R = 6\left[\frac{a''}{a} + \left(\frac{a'}{a}\right)^2 + \frac{K}{a^2}\right]\,,$$

which also diverges if $a \to 0$. Also the energy density, which grows faster than $1/a^2$ as $a \to 0$ diverges at the big bang.

If the curvature K is positive, the Universe contracts after $\tau = \tau_{\rm max}$ and, since the graph $a(\tau)$ is convex, reaches $a = 0$ at some finite time τ_c, the time of the 'big crunch'. The big crunch is also a physical singularity beyond which spacetime cannot be continued.

It is important to note that this behaviour of the scale factor can only be implied if the so-called 'strong energy condition' holds, $\rho_{\rm eff} + 3P_{\rm eff} > 0$. This is illustrated in Fig. 1.1.

1.2.3 Cosmological distance measures

It is notoriously difficult to measure distances in the Universe. The position of an object in the sky gives us its angular coordinates, but how far away is the object from us? This problem has plagued cosmology for centuries. It was only Hubble, who discovered around 1915–1920 that the 'spiral nebulae' are actually not situated inside our own galaxy but much further away. This then led to the discovery of the expansion of the Universe.

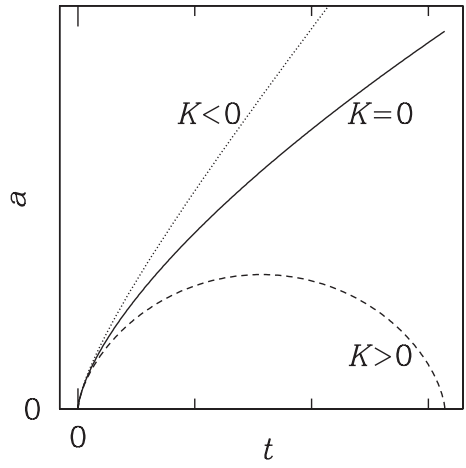


Fig. 1.1. The kinematics of the scale factor in a Friedmann–Lemaître universe which satisfies the strong energy condition, $\rho_{\rm eff} + 3P_{\rm eff} > 0$.

For cosmologically distant objects, a third coordinate, which is nowadays relatively easy to obtain, is the redshift z experienced by the photons emitted from the object. A given spectral line with intrinsic wavelength λ is redshifted due to the expansion of the Universe. If it is emitted at some time τ, it reaches us today with wavelength $\lambda_0 = \lambda a_0/a(\tau) = (1+z)\lambda$. This leads to the definition of the cosmic redshift

$$z(\tau) + 1 = \frac{a_0}{a(\tau)} \,. \tag{1.37}$$

On the other hand, an object at physical distance $d = a_0 r$ away from us, at redshift $z \ll 1$, recedes with speed $v = H_0 d$. To the lowest order in z, we have $\tau_0 - \tau \approx d$ and $a_0 \approx a(\tau) + a'(\tau_0 - \tau)$, so that

$$1 + z \approx 1 + \frac{a'}{a}(\tau_0 - \tau) \approx 1 + H_0 d \,.$$

For objects that are sufficiently close, $z \ll 1$ we therefore have $v \approx z$ and hence $H_0 = v/d$. This is the method usually applied to measure the Hubble constant.

There are different ways to measure distances in cosmology all of which give the same result in a Minkowski universe but differ in an expanding universe. They are, however, simply related as we shall see.

One possibility is to define the distance D_A to a certain object of given physical size Δ seen at redshift z_1 such that the angle subtended by the object is given by

$$\vartheta = \Delta/D_A \,, \qquad D_A = \Delta/\vartheta \,. \tag{1.38}$$

This is the angular diameter distance, see Fig. 1.2.

Fig. 1.2. The two ends of the object emit a flash simultaneously from A and B at z_1 which reaches us today. The angular diameter distance to A (or B) is defined by $D_A = \Delta/\vartheta$.

We now derive the expression

$$D_A(z) = \frac{1}{\sqrt{|\Omega_K|}H_0(1+z)}\chi\left(\sqrt{|\Omega_K|}H_0\int_0^z \frac{dz'}{H(z')}\right), \tag{1.39}$$

for the angular diameter distance to redshift z. In a given cosmological model, this allows us to express the angular diameter distance for a given redshift as a function of the cosmological parameters.

To derive Eq. (1.39) we use the coordinates introduced in Eq. (1.9). Without loss of generality we set $r = 0$ at our position. We consider an object of physical size Δ at redshift z_1 simultaneously emitting a flash at both of its ends A and B. Hence $r = r_1 = t_0 - t_1$ at the position of the flashes, A and B at redshift z_1. If Δ denotes the physical arc length between A and B we have $\Delta = a(t_1)\chi(r_1)\vartheta = a(t_1)\chi(t_0 - t_1)\vartheta$, i.e.,

$$\vartheta = \frac{\Delta}{a(t_1)\chi(t_0 - t_1)}\,. \tag{1.40}$$

According to Eq. (1.38) the angular diameter distance to t_1 or z_1 is therefore given by

$$a(t_1)\chi(t_0 - t_1) \equiv D_A(z_1)\,. \tag{1.41}$$

To obtain an expression for $D_A(z)$ in terms of the cosmic density parameters and the redshift, we have to calculate $(t_0 - t_1)(z_1)$.

Note that in the case $K = 0$ we can normalize the scale factor a as we want, and it is convenient to choose $a_0 = 1$, so that comoving scales become physical scales today. However, for $K \neq 0$, we have already normalized a such that $K = \pm 1$ and $\chi(r) = \sin r$ or $\sinh r$. In this case, we have no normalization constant left and a_0 has the dimension of a length. The present spatial curvature of the Universe then is $\pm 1/a_0^2$.

The Friedmann equation Eq. (1.20) reads

$$\dot{a}^2 = \frac{8\pi G}{3}a^4\rho + \frac{1}{3}\Lambda a^4 - Ka^2, \tag{1.42}$$

where $\dot{a} = da/dt$. To be specific, we assume that ρ is a combination of dust, cold, non-relativistic 'matter' of $P_m = 0$ and radiation of $P_r = \rho_r/3$.

Since $\rho_r \propto a^{-4}$ and $\rho_m \propto a^{-3}$, we can express the terms on the r.h.s. of (1.42) as

$$\frac{8\pi G}{3} a^4 \rho = H_0^2 \left(a_0^4 \Omega_r + \Omega_m a a_0^3\right) , \tag{1.43}$$

$$\frac{1}{3}\Lambda a^4 = H_0^2 \Omega_\Lambda a^4 , \tag{1.44}$$

$$-K a^2 = H_0^2 \Omega_K a^2 a_0^2 . \tag{1.45}$$

The Friedmann equation then implies

$$\frac{da}{dt} = H_0 a_0^2 \left(\Omega_r + \frac{a}{a_0}\Omega_m + \frac{a^4}{a_0^4}\Omega_\Lambda + \frac{a^2}{a_0^2}\Omega_K\right)^{1/2} , \tag{1.46}$$

so that

$$t_0 - t_1 = \frac{1}{H_0 a_0} \int_0^{z_1} \frac{dz}{\left[\Omega_r (z+1)^4 + \Omega_m (z+1)^3 + \Omega_\Lambda + \Omega_K (z+1)^2\right]^{1/2}} . \tag{1.47}$$

Here we have used $z + 1 = a_0/a$ so that $da = -dz a_0/(1+z)^2$.

In principle, we could of course also add other matter components like, e.g. 'quintessence' (Caldwell *et al.*, 1998), which would lead to a somewhat different form of the integral (1.47), but for definiteness, we remain with matter, radiation and a cosmological constant.

From $-K/H_0^2 a_0^2 = \Omega_K$ we obtain $H_0 a_0 = 1/\sqrt{|\Omega_K|}$ for $\Omega_K \neq 0$. The expression for the angular diameter distance thus becomes

$$D_A(z) = \begin{cases} \frac{1}{\sqrt{|\Omega_K|} H_0 (z+1)} \chi\left(\sqrt{|\Omega_K|} \int_0^z \frac{dz'}{\left[\Omega_r (z'+1)^4 + \Omega_m (z'+1)^3 + \Omega_\Lambda + \Omega_K (z'+1)^2\right]^{1/2}}\right) \\ \qquad \text{if } K \neq 0 \\ \\ \frac{1}{H_0 (z+1)} \int_0^z \frac{dz'}{\left[\Omega_r (z'+1)^4 + \Omega_m (z'+1)^3 + \Omega_\Lambda\right]^{1/2}} \\ \qquad \text{if } K = 0 . \end{cases} \tag{1.48}$$

Using the Friedmann equation, this formula can also be written in the more general form of Eq. (1.39).

In general, the above integral has to be solved numerically. It determines the angle $\vartheta(\Delta, z) = \Delta/D_A(z)$ under which an object of size Δ placed at redshift z is seen (see Figs. 1.3 and 1.4).

If we are able to measure the redshifts and the angular extensions of a certain class of objects at different redshifts, of which we know the intrinsic size Δ, comparing

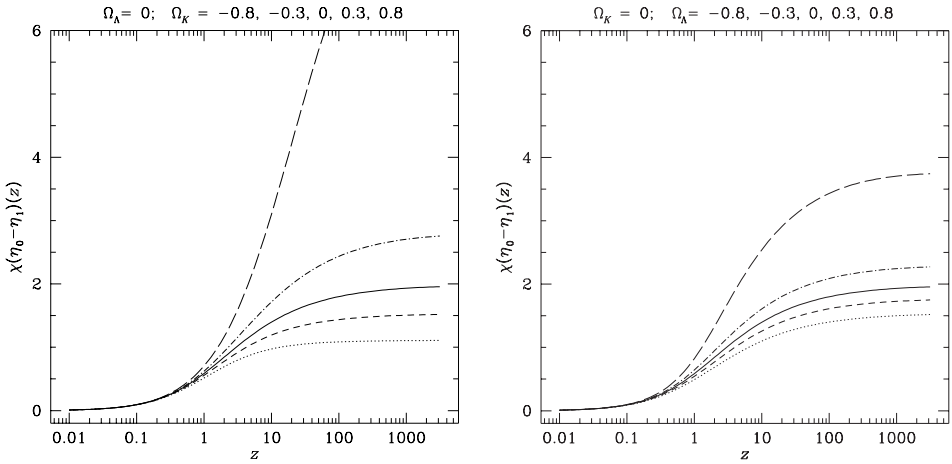


Fig. 1.3. The function $\chi(t_0 - t_1)$ as a function of the redshift z for different values of the cosmological parameters Ω_K (left, with $\Omega_\Lambda = 0$) and Ω_Λ (right, with $\Omega_K = 0$), namely -0.8 (dotted), -0.3 (short-dashed), 0 (solid), 0.3 (dot-dashed), 0.8 (long-dashed).

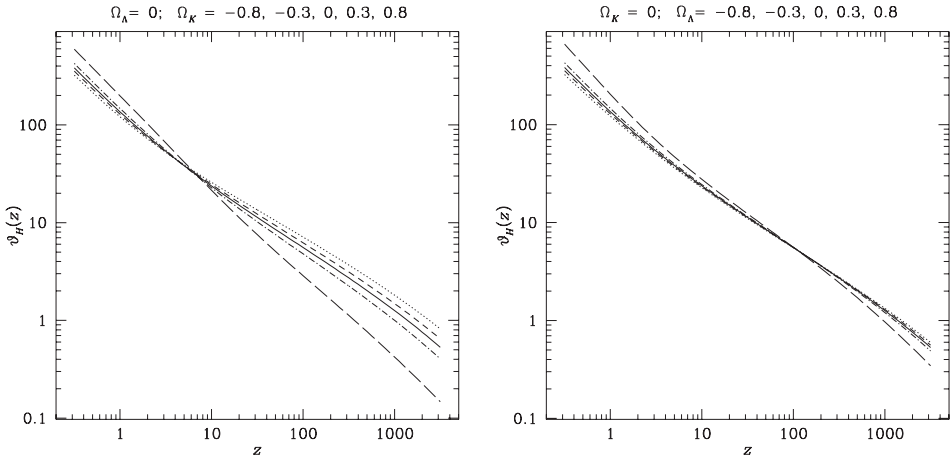


Fig. 1.4. $\vartheta_H(z_1)$ (in degrees) for different values of the cosmological parameters Ω_K and Ω_Λ the line styles are as in Fig. 1.3.

with Eq. (1.48) allows us, in principle, to determine the parameters Ω_m, Ω_Λ, Ω_K and H_0.

Observationally we know for certain that $10^{-5} < \Omega_r \leq 10^{-4}$ as well as $0.1 \leq \Omega_m \lesssim 1$, $|\Omega_\Lambda| \lesssim 1$ and $|\Omega_K| \lesssim 1$.

If we are interested in small redshifts, $z_1 \lesssim 10$, we may therefore safely neglect Ω_r. In this region, Eq. (1.48) is very sensitive to Ω_Λ and provides an excellent mean to constrain the cosmological constant.

At high redshift, $z_1 \gtrsim 1000$, neglecting radiation is no longer a good approximation.

We shall later also need the opening angle of the *horizon* distance,

$$\vartheta_H(z_1) = \frac{t_1}{\chi(t_0 - t_1)}\,, \tag{1.49}$$

$$t_1 = \frac{1}{H_0 a_0} \int_{z_1}^{\infty} \frac{dz}{\left[\Omega_r (z+1)^4 + \Omega_m (z+1)^3 + \Omega_\Lambda + \Omega_K (z+1)^2\right]^{1/2}}\,. \tag{1.50}$$

(Clearly this integral diverges if $\Omega_r = \Omega_m = 0$. This is exactly what happens during an inflationary period and leads there to the solution of the horizon problem, see Section 1.5.)

Neglecting Ω_r, for $\Omega_\Lambda = 0$ and small curvature, $0 < |\Omega_K| < \Omega_m z_1$ at high enough redshift, $z_1 \geq 10$, one has $t_0 - t_1 \simeq 2\sqrt{|\Omega_K|/\Omega_m} = 2/(H_0 a_0 \sqrt{\Omega_m})$. With $\chi(x) \simeq x$, which is valid for small curvature, this yields $\vartheta(\Delta, z_1) \simeq \sqrt{\Omega_m} H_0 a_0 \Delta/(2a_1) = \frac{1}{2}\sqrt{\Omega_m} H_0 \Delta/(z_1+1)$ (see also Ex. 1.8).

Another important distance measure in cosmology is the luminosity distance. It is defined as follows. Let L be the luminosity (energy emitted per second) of a source at redshift z_1 and F its flux (energy received per second per square centimetre) arriving at the observer position. We define the luminosity distance to the source by

$$D_L(z_1) \equiv \left(\frac{L}{4\pi F}\right)^{1/2}\,. \tag{1.51}$$

We now want to show that $D_L(z_1) = (1+z_1)^2 D_A(z_1)$.

In a proper time interval of the emitter, $d\tau_1 = a(t_1)\,dt$, the source emits the energy $L a(t_1)\,dt$. This energy is redshifted by a factor of $(1+z_1)^{-1} = a(t_1)/a(t_0)$. It is then distributed over a sphere with radius $a(t_0)\chi(t_0 - t_1)$. So that the flux per proper time of the observer $d\tau_0 = a(t_0)\,dt$ becomes

$$F = \frac{L a^2(t_1)}{4\pi a^4(t_0)\chi^2(t_0 - t_1)}\,,$$

leading to

$$D_L(z_1) = \frac{a(t_0)^2}{a(t_1)}\chi(t_0 - t_1) = (1+z_1)^2 D_A(z_1)\,. \tag{1.52}$$

The luminosity distance hence contains two additional factors $(1+z)$ compared to the angular diameter distance. One of them is due to the 'redshift' of proper time and the other is due to the redshift of photon energy.

1.3 Recombination and decoupling

We assume that, at sufficiently early times, reaction rates for particle interactions are much faster than the expansion rate, so that the cosmic fluid is in thermal equilibrium. During its expansion, the Universe then cools adiabatically. At early times, it is dominated by a relativistic radiation background with

$$\rho = C/a^4 = g_{\rm eff} a_{SB} T^4 \,. \tag{1.53}$$

This behaviour implies that $T \propto a^{-1}$. Here $g_{\rm eff}$ is the effective number of degrees of freedom, which we define below and a_{SB} is the Stefan–Boltzmann constant, $a_{SB} = \pi^2/30$ in our units. For massless (or extremely relativistic) fermions and bosons in thermal equilibrium at temperature T with N_b respectively N_f spin degrees of freedom we have (remember that we use units such that $\hbar = k_B = c = 1$)

$$\begin{aligned}
\rho_b &= \frac{N_b 4\pi}{(2\pi)^3} \int_0^\infty \frac{p^3\, dp}{\exp(p/T) - 1} = \frac{N_b T^4}{2\pi^2} \int_0^\infty \frac{x^3\, dx}{\exp(x) - 1} \\
&= \frac{N_b T^4}{2\pi^2} \Gamma(4)\zeta(4) = \frac{N_b T^4 \pi^2}{30} \,, \qquad (1.54) \\
\rho_f &= \frac{N_f 4\pi}{(2\pi)^3} \int_0^\infty \frac{p^3\, dp}{\exp(p/T) + 1} = \frac{N_f T^4}{2\pi^2} \int_0^\infty \frac{x^3\, dx}{\exp(x) + 1} \\
&= \frac{N_f T^4}{2\pi^2} \Gamma(4)\zeta(4)\frac{7}{8} = \frac{7}{8}\frac{N_f T^4 \pi^2}{30} \,, \qquad (1.55)
\end{aligned}$$

where Γ denotes the Gamma-function and ζ is the Riemann zeta-function and we make use of the integrals (Gradshteyn & Ryzhik, 2000)

$$I_b(\alpha) = \int_0^\infty \frac{x^\alpha\, dx}{\exp(x) - 1} = \Gamma(\alpha + 1)\zeta(\alpha + 1) \,, \tag{1.56}$$

$$I_f(\alpha) = \int_0^\infty \frac{x^\alpha\, dx}{\exp(x) + 1} = \left[1 - \left(\frac{1}{2}\right)^\alpha\right] \Gamma(\alpha + 1)\zeta(\alpha + 1) \,. \tag{1.57}$$

Furthermore, $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, and $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$, see Abramowitz & Stegun (1970).

Hence $\rho = \rho_b + \rho_f = g_{\rm eff} a_{SB} T^4$ for $a_{SB} = \pi^2 k_B^4/(30\ \hbar^3 c^2) = \pi^2/30$ and $g_{\rm eff} = N_b + 7/8 N_f$, if all the particles are at the same temperature T. If the temperatures are different, like e.g., the neutrino temperature after electron–positron annihilation, this has to be taken into account with a factor $(T_\nu/T_\gamma)^4$.

At temperatures below the electron mass, $T < m_e \sim 0.5$ MeV only neutrinos and photons are still relativistic. Very recently, $T \lesssim 0.01$ eV the neutrinos also become non-relativistic so that the density parameter of relativistic particles today

is given only by the photon density,

$$\Omega_{\rm rel} = \Omega_\gamma = \frac{16\pi G}{3H_0^2} a_{SB} T_0^4 = 2.49 \times 10^{-5} h^{-2} \,. \tag{1.58}$$

Here we have set $T_0 = 2.725\,\mathrm{K}$ (see Particle Data Group, 2006).

The pressure of relativistic particles is given by $P = T_i^i/3 = \rho/3$. The thermodynamic relation $dE = T\,dS - P\,dV$ therefore gives for entropy density $s = dS/dV$

$$s = \frac{dS}{dV} = \frac{1}{T}\left(\frac{dE}{dV} + P\right) = \frac{\rho + P}{T} = \frac{4\rho}{3T} \,. \tag{1.59}$$

Using the expression for the energy density (1.54) and (1.55) this gives for each particle species X

$$s_X = \begin{cases} \frac{2\pi^2}{45} N_X T^3 & \text{for bosons} \,, \\ \frac{7\pi^2}{180} N_X T^3 & \text{for fermions} \,. \end{cases} \tag{1.60}$$

The particle density for relativistic particles is given by

$$n_X = \frac{N_X}{2\pi^2} \int \frac{p^2}{\exp(p/T) \pm 1}\, dp = \begin{cases} T^3 \frac{N_X}{\pi^2} \zeta(3) & \text{for bosons} \,, \\ T^3 \frac{N_X}{\pi^2} \zeta(3) \frac{3}{4} & \text{for fermions} \,. \end{cases} \tag{1.61}$$

The particle and entropy densities both scale like T^3. Using $\zeta(3) \simeq 1.202\,057$ we obtain

$$s_X \simeq \begin{cases} 3.6 \cdot n_X & \text{for bosons} \,, \\ 4.2 \cdot n_X & \text{for fermions} \,. \end{cases} \tag{1.62}$$

The photons obey a Planck distribution ($\epsilon = ap =$ the photon energy),

$$f(\epsilon) = \frac{1}{e^{\epsilon/T} - 1} \,. \tag{1.63}$$

At a temperature of about $T \sim 4000\,\mathrm{K} \sim 0.4$ eV, the number density of photons with energies above the hydrogen ionization energy ($= \Delta = 1\,\mathrm{Ry} = 13.6$ eV) drops below the baryon density of the Universe, and the protons begin to (re)combine to neutral hydrogen. (Helium has already recombined earlier.) Photons and baryons are tightly coupled before (re)combination by Thomson scattering of electrons. During recombination the free electron density drops sharply and the mean free path of the photons grows larger than the Hubble scale. At the temperature $T_{\rm dec} \sim 3000\,\mathrm{K}$ (corresponding to the redshift $z_{\rm dec} \simeq 1100$ and the physical time $t_{\rm dec} \simeq a_{\rm dec}\eta_{\rm dec} \simeq 10^5$ yr) photons decouple from the electrons and the Universe becomes transparent. We now want to study this process in somewhat more detail.

1.3.1 The physics of recombination

As we have seen above, the photon entropy is given by

$$s_\gamma = \frac{4\pi^2}{45} T^3 \simeq 3.6 n_\gamma \,.$$

The conserved baryon number n_B satisfies $a^3 n_B = \text{constant}$, hence $n_B \propto a^{-3} \propto T^3$. The entropy per baryon is therefore a constant,

$$\sigma = s_\gamma / n_B = \frac{\frac{4\pi^2}{45} T_0^3}{\Omega_B \rho_c(\tau_0)/m_p} = 1.4 \times 10^8 \frac{T_{2.7}^3}{\Omega_B h^2} \,. \tag{1.64}$$

Here we have used (see Appendix 1)

$$\begin{aligned} \rho_c(\tau_0) &= 1.88 h^2 \times 10^{-29}\,\mathrm{g\,cm^{-3}} = 8.1 h^2 \times 10^{-11}\,(\mathrm{eV})^4 \,, \\ m_p &= 9.38 \times 10^8\ \mathrm{eV} \,, \qquad \text{(proton mass)}, \\ T(\tau_0) &= 2.3 T_{2.7} \times 10^{-4}\ \mathrm{eV} \,, \quad T_{2.7} = T(\tau_0)/2.7\,\mathrm{K} \,. \end{aligned}$$

As we shall see in the next section, the baryon density is approximately $\Omega_B h^2 \simeq 2 \times 10^{-2}$ so that $\sigma \simeq 10^{10}$. Correspondingly the ratio between photon and baryon density is

$$\eta_B = n_B / n_\gamma = 2.7 \times 10^{-8} \left(\frac{\Omega_B h^2}{T_{2.7}^3} \right). \tag{1.65}$$

As long as hydrogen is ionized, the timescale of interaction between photons and electrons (Thomson scattering) and between electrons and protons (Rutherford scattering) is much faster than expansion and we may therefore consider the latter as adiabatic. At every moment, the electron, proton, photon plasma is in thermal equilibrium. As long as the temperature is above the ionization energy of neutral hydrogen, $T > 1\,\mathrm{Ry} = \Delta = \alpha^2 m_e/2 = 13.6$ eV all hydrogen atoms that form are rapidly dissociated. Most electrons and protons are free and the neutral hydrogen density is very low. At some sufficiently low temperature, however, there will no longer be sufficiently many energetic photons around to disrupt neutral hydrogen and the latter becomes more and more abundant. To determine the temperature at which this transition, called 'recombination',[3] happens, we apply the standard rules of equilibrium statistical mechanics to the reaction

$$e^- + p \longleftrightarrow \mathrm{H} + \gamma \ (13.6\ \mathrm{eV}) \,. \tag{1.66}$$

Supposing that pressure and temperature are fixed and only the number of free electrons, N_e, free protons, N_p, hydrogen atoms, N_H, and photons, N_γ, can change,

[3] The expression 'combination' would be more adequate, since this is the first time that neutral hydrogen forms, but it is difficult to change historical mis-namings. . . .

the second law of thermodynamics implies that the Gibbs potential G is constant,

$$0 = dG = \mu_p\, dN_p + \mu_e\, dN_e + \mu_H\, dN_H + \mu_\gamma\, dN_\gamma\ ,$$

Here μ_X denotes the chemical potential of species X. The different dN_X are not independent. Particle number conservation implies

$$dN_p + dN_H = dN_e + dN_H = 0\,. \tag{1.67}$$

As there is no conservation of photons, the chemical potential of photons is thus $\mu_\gamma = 0$. With this and Eq. (1.67) the Gibbs equation, $dG = 0$ implies

$$\mu_e + \mu_p - \mu_H = 0\ . \tag{1.68}$$

In this discussion, where we are more interested in the basic concepts than in accuracy we neglect helium which has recombined earlier. We shall set $n_p + n_H = n_B$ which induces an error of about 25%. For an accurate calculation of the final ionization fraction, one would have to take into account both, the recombination of helium and the recombination into excited states of hydrogen. It is actually interesting to note that recombination into the ground state (1S) is not efficient at all since the ionization cross section is very high for resonant Lyα photons so that most of these just ionize another hydrogen atom leading to no net recombination. The same is true for recombination into the 2P excited state. It is more efficient if electrons are captured into the 2S level from which they can decay into the ground state via the emission of two photons. By angular momentum conservation, the emission of a single photon is not possible. The inverse process, excitation from 1S to 2S is a three-body process and therefore highly unlikely. Even though the rate of the transition $(e,\, p) \to \mathrm{H}_{2\mathrm{S}} \to \mathrm{H}_{1\mathrm{S}}$ is relatively low, it wins against direct recombination into the ground state and subsequent cosmological redshifting of the photon before the next ionization can take place. More details are found in Peebles (1993), Mukhanov (2005), Rubino-Martin *et al.* (2006) and Wong *et al.* (2006). Despite this fact, a discussion of recombination into the ground state captures the main features of the process and the correct recombination and decoupling redshifts do not significantly differ from those obtained here.

In thermal equilibrium, electrons, protons and hydrogen atoms obey a Maxwell–Boltzmann distribution. Their number densities are given by (see Ex. 1.5)

$$n_e = \frac{2}{(2\pi)^3}(2\pi m_e T)^{3/2}\, \exp\left(-\frac{m_e - \mu_e}{T}\right), \tag{1.69}$$

$$n_p = \frac{2}{(2\pi)^3}(2\pi m_p T)^{3/2}\, \exp\left(-\frac{m_p - \mu_p}{T}\right), \tag{1.70}$$

$$n_H = \frac{4}{(2\pi)^3}(2\pi m_H T)^{3/2}\, \exp\left(-\frac{m_H - \mu_H}{T}\right). \tag{1.71}$$

We now make use of the fact that the Universe is globally neutral, $n_e = n_p$. Furthermore, the binding energy of hydrogen $\Delta = \alpha^2 m_e/2$ (here $\alpha \simeq 1/137$ is the fine structure constant) is given by $\Delta = m_e + m_p - m_H$. With this we obtain

$$\frac{n_e^2}{n_H} = \frac{n_e n_p}{n_H} = \left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-\Delta/T}. \tag{1.72}$$

Here we have neglected the small difference between the hydrogen and the proton mass in the second factor of Eqs. (1.70) and (1.71) but not in the exponential.

We now define the ionization fraction x_e by $x_e \equiv n_e/(n_e + n_H)$. Eq. (1.72) then leads to

$$\frac{x_e^2}{1 - x_e} = \frac{n_e^2}{n_H(n_p + n_H)} = \frac{1}{n_B}\left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-\Delta/T}. \tag{1.73}$$

Inserting the entropy per baryon, $\sigma = (4\pi^2/45)T^3/n_{\rm B}$, in this equation yields

$$\frac{x_e^2}{1 - x_e} = \frac{45\sigma}{4\pi^2}\left(\frac{m_e}{2\pi T}\right)^{3/2} e^{-\Delta/T}. \tag{1.74}$$

This is the *Saha equation*. At very high temperatures, $T \gg \Delta$, the ionization fraction x_e is close to 1. Recombination happens roughly when $\sigma \exp(-\Delta/T)$ is of the order of unity. If $\sigma \sim 1$ this corresponds to $T \sim \Delta$. The fact that the entropy per baryon is very large, $\sigma = 1.4 \times 10^8 (\Omega_B h^2)^{-1} \sim 10^{+10}$ delays recombination significantly. Since there are so many more photons than baryons in the Universe, even at a temperature much below $\Delta = 13.6$ eV there are still enough photons in the high-energy tail of the Planck distribution to keep the Universe ionized.

To be more specific we define the recombination temperature $T_{\rm rec}$ as the temperature when $x_e = 0.5$ (as we shall see, the precise value is of little importance). Eq. (1.74) then leads to

$$\left(\frac{T_{\rm rec}}{1\,{\rm eV}}\right)^{-3/2} e^{-\Delta/T_{\rm rec}} = 1.3 \times 10^{-16}\, \Omega_B h^2. \tag{1.75}$$

For $\Omega_B h^2 \simeq 0.02$ we obtain

$$T_{\rm rec} = 3757\,{\rm K} = 0.32\,{\rm eV}, \quad z_{\rm rec} = 1376\,.$$

The function $x_e(T)$ is shown in Fig. 1.5. Clearly, this function grows very steeply from $x_e \sim 0$ to $x_e \sim 1$ at $T \sim 3700$ K and $T_{\rm rec}$ depends only weakly on the value chosen for $x_e(T_{\rm rec})$.

Interestingly, at temperature $T_{\rm rec}$ the baryon and photon densities are of the same order, $\rho_\gamma(T_{\rm rec}) \simeq \rho_B(T_{\rm rec})$. This seems to be a complete coincidence. More

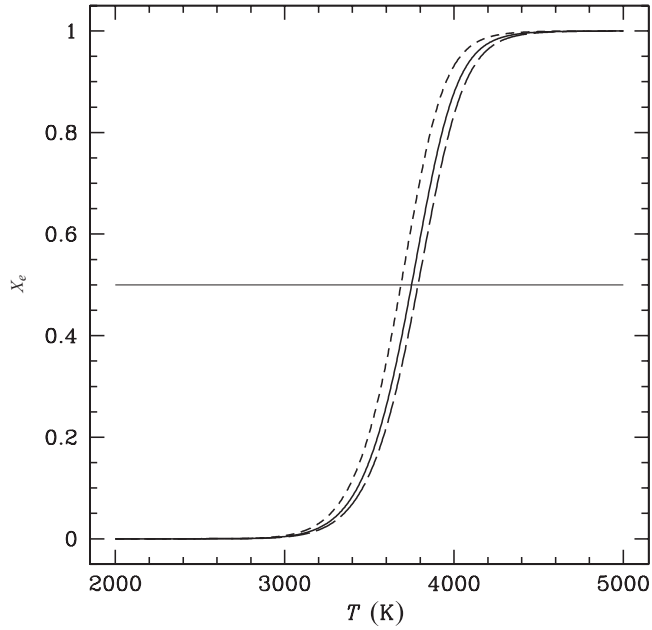


Fig. 1.5. The ionization fraction x_e as a function of the temperature is obtained via the Saha equation for $\Omega_B h^2 = 0.02$ (solid curve), for $\Omega_B h^2 = 0.03$ (long-dashed curve) and for $\Omega_B h^2 = 0.01$ (short-dashed curve). Our definition of recombination, $x_{\rm rec} = 0.5$, is indicated. Note that x decays from $x_e \simeq 1$ to $\simeq 0$ between $T = 4000$ and 3400 K.

precisely, the ratio of these two densities is given by

$$\frac{\rho_\gamma}{\rho_B} = \frac{(\pi^2/15)T^4}{n_B m_p} = \frac{\pi^2 T_0^4}{15 n_B(t_0) m_p}(z+1)$$
$$\simeq 2 \times 10^{-5} \left(\Omega_B h^2\right)^{-1} (z+1)\,. \tag{1.76}$$

This ratio is equal to 1 at redshift z_{rb} given by

$$(1+z_{rb}) = 10^3 \left(\frac{\Omega_B h^2}{2\times 10^{-2}}\right) \simeq 10^3 \sim 1 + z_{\rm rec}\,. \tag{1.77}$$

1.3.2 Final ionization and photon decoupling

We have determined the temperature at which electrons and protons recombine to neutral hydrogen. As the free electron fraction drops, the interaction rate between electrons and protons decreases and at some point, the remaining free electrons and protons are too sparse to find each other, so that the number of free electrons remains constant. But also the photon–electron interaction rate decreases. Whenever

an interaction rate Γ drops below the expansion rate of the Universe,

$$\Gamma < H \,,$$

one considers the correponding reaction as 'frozen'. It becomes negligible. When the recombination rate drops below the expansion rate, recombination freezes and the ionization fraction remains constant. When the scattering rate of photons on electrons falls below the expansion rate of the Universe, photons become free to propagate without further scattering. We want to calculate both, the final ionization, x_R, and the redshift $z_{\rm dec}$ of the decoupling of photons. Let us first determine the temperature T_g at which the process of reionization freezes out. The cross section of the reaction $p^+ + e^- \rightarrow \mathrm{H} + \gamma$ is (see, e.g. Rybicki & Lightman, 1979)

$$\langle \sigma_R v \rangle \simeq 4.7 \times 10^{-24} \left(\frac{T}{1\,\mathrm{eV}} \right)^{-1/2} \mathrm{cm}^2 \,. \tag{1.78}$$

Here v is the thermal electron velocity and we have used the fact that $3T = m_e v^2$. The reaction rate is therefore

$$\begin{aligned} \Gamma_R &= n_p \langle \sigma_R v \rangle = x_e \left(\frac{n_B}{n_\gamma} \right) n_\gamma \langle \sigma_R v \rangle \\ &\simeq 2.4 \times 10^{-10}\,\mathrm{cm}^{-1} \left(\frac{T}{1\,\mathrm{eV}} \right)^{7/4} \exp(-\Delta/2T)(\Omega_B h^2)^{1/2} \,, \end{aligned}$$

where we have inserted the Saha equation, assuming that the ionization fraction is much smaller than 1, i.e.,

$$x_e \simeq (\sqrt{45\sigma}/2\pi)(m_e/2\pi T)^{3/4} \exp(-\Delta/2T) \ll 1 \,.$$

We have also used Eq. (1.65).

To determine the expansion rate $H(T)$, we neglect curvature or a possible cosmological constant, which is certainly a good approximation for all redshifts larger than, say, 5. We also assume that the Universe is matter dominated at freeze-out, which induces an error of about 15% in H. The Friedmann equation (1.18) then gives

$$\begin{aligned} H^2 &\simeq \frac{8\pi G}{3} \rho \simeq \frac{8\pi G}{3} \rho_0 (a_0/a)^3 \\ &= \frac{8\pi G}{3} \Omega_m \rho_c(t_0)(T/T_0)^3 \,, \end{aligned}$$

so that

$$H \simeq 3 \times 10^{-23}\,\mathrm{cm}^{-1} (\Omega_m h^2)^{1/2} \left(\frac{T}{1\,\mathrm{eV}} \right)^{3/2} . \tag{1.79}$$

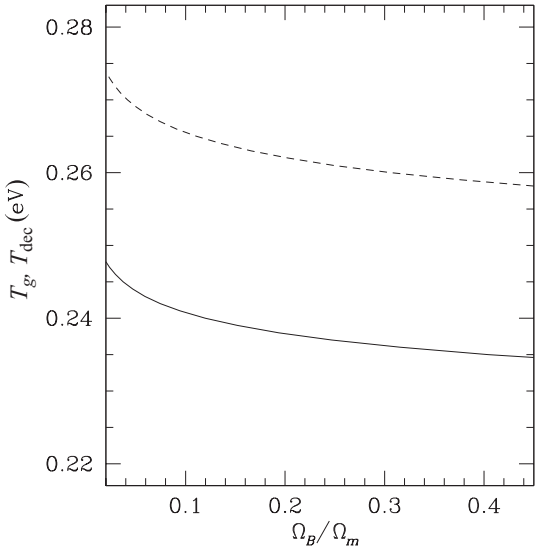


Fig. 1.6. The freeze-out temperatures of recombination (solid curve) and of Thomson scattering (dashed curve) as functions of Ω_B/Ω_m.

Eq. (1.79) is a very useful formula, valid whenever the Universe is dominated by non-relativistic matter, dust, $P \ll \rho$, and curvature or a cosmological constant are negligible.

The temperature T_g is defined by $\Gamma_R(T_g) = H(T_g)$, which finally leads to

$$\left(\frac{T_g}{1\,\mathrm{eV}}\right)^{1/4} e^{-\Delta/2T_g} = 1.2 \times 10^{-13} \left(\frac{\Omega_m}{\Omega_B}\right)^{1/2} . \tag{1.80}$$

This result is independent of h. For $\Omega_m \simeq 7\Omega_B$ (the value inferred from observations (Spergel *et al.*, 2003)), we obtain $T_g \simeq 0.24$ eV and $z_g \simeq 1010$ (see Fig. 1.6). T_g depends only weakly on the ratio Ω_B/Ω_m.

The final ionization fraction is given by

$$x_R \simeq x_e(T_g) \simeq 7.3 \times 10^{-6} \left(\frac{T_g}{1\,\mathrm{eV}}\right)^{-1} \Omega_m^{1/2}/(\Omega_B h) \simeq 3 \times 10^{-5} \Omega_m^{1/2}/(\Omega_B h) . \tag{1.81}$$

A more detailed numerical analysis, taking into account the contribution from radiation to the expansion rate and the recombination into excited states of the hydrogen atoms and the presence of helium (see next section) gives $x_R \sim 1.2 \times 10^{-5} \Omega_m^{1/2}/(\Omega_B h)$ (Peebles, 1993; Mukhanov, 2005). We can use this result to calculate the optical depth τ to Thomson scattering of photons by free electrons up to a redshift $z < z_g$ in a recombined universe. The optical depth to z is the scattering probability of a photon integrated from z until today. With the Thomson

cross section

$$\sigma_T = \frac{8\pi}{3}\alpha^2 m_e^{-2} \simeq 6.65 \times 10^{-25}\,\mathrm{cm}^2\,, \tag{1.82}$$

one finds

$$\tau(z) \equiv \int_{t(z)}^{t_0} \sigma_T n_e \frac{dt}{a} \simeq 0.046 x_R (1+z)^{3/2} \Omega_B \Omega_m^{-1/2} h\,. \tag{1.83}$$

With the residual ionization above we find $\tau(z = 800) \simeq 0.01$. As we shall see in Section 6.3 the Universe is reionized at low redshift $z \sim 10$, which increases the optical depth by roughly a factor of 10. This rescattering of CMB photons is relevant for the evolution of fluctuations as we shall discuss in Section 6.3.

As long as the temperature is larger than T_g, the reaction $p + e \longleftrightarrow \mathrm{H} + \gamma$ is in thermal equilibrium. When the temperature drops below T_g, the recombination process freezes out and the degree of ionization remains nearly constant.

Let us also note that in deriving the Saha equation (1.74), we used the fact that the process of recombination is in thermal equilibrium, which we have verified only now since freeze-out happens after recombination, $T_g < T_{\mathrm{rec}}$.

We finally calculate the redshift of the decoupling of photons. The process which remains effective longest is elastic Thomson scattering. Its rate is given by

$$\begin{aligned}\Gamma_T &= \sigma_T n_e = \sigma_T x_e \left(\frac{n_B}{n_\gamma}\right) n_\gamma \\ &\simeq 3.4 \times 10^{-11}\,\mathrm{cm}^{-1} (\Omega_B h^2)^{1/2} \left(\frac{T}{1\,\mathrm{eV}}\right)^{9/4} \exp(-\Delta/2T)\,.\end{aligned} \tag{1.84}$$

Comparing it to the expansion rate, we find T_{dec} which is defined by $H(T_{\mathrm{dec}}) = \Gamma_T(T_{\mathrm{dec}})$. A rough estimate gives $T_{\mathrm{dec}} \sim 0.26$ eV (see Fig. 1.6) which corresponds to $z_{\mathrm{dec}} \sim 1100$. Again we have assumed $x_e \ll 1$ in Eq. (1.84) which is justified since $T_{\mathrm{dec}} \sim 3000\,\mathrm{K}$ (see Fig. 1.5).

Even though after z_{dec} photons decouple from electrons, the latter are still coupled to photons. The scattering rate of electrons, given by $\Gamma_e = \sigma_T x_e n_\gamma \gg \sigma_T n_e$, is sufficient to keep the electrons and with them the matter in thermal equilibrium with the photons until very low redshift. Therefore, even after recombination the matter temperature is equal to the temperature of the CMB and does not decay like $1/a^2$ as would be expected from a pure thermal gas of massive particles (see page 25).

1.3.3 Propagation of free photons and the CMB

After $t_{\rm dec}$, photons cease any interaction with the cosmic fluid and propagate freely. It is straight forward to estimate that the cross section for Rayleigh scattering with hydrogen atoms is much too weak to be relevant.

The free propagation of photons after decoupling is described with the Liouville equation for the photon distribution function, which we now develop. Since photons do not interact anymore, they simply move along geodesics. The Liouville equation translates this to a differential equation for the 1-particle distribution function f of the photons. The function f describes the particle density in the phase space P_0, the photon mass-shell, given by

$$P_0 = \{(x,\, p) \in T\mathcal{M} \mid g_{\mu\nu}(x) p^\mu p^\nu = 0\}, \quad f : P_0 \to \mathbb{R}\,.$$

The distribution function f gives the number of particles per phase space volume $|g|\, d^3x\, d^3p$ at fixed time t. In some general geometry a specific space-like hypersurface Σ has to be chosen and one then has to show that f does not depend on this choice (more details are found in Ehlers (1971) and Stewart (1971)). In cosmology, due to the symmetries present, we simply use the hypersurfaces of constant time, $\Sigma = \Sigma_t$.

We choose the coordinates $(x^\mu,\, p^i)$ on the seven-dimensional mass-shell ($0 \le \mu \le 3$ and $1 \le i \le 3$). The energy p^0 is then determined by the mass-shell condition $g_{\mu\nu}(x) p^\mu p^\nu = 0$. Liouville's equation now says that the 1-particle distribution remains unchanged if we follow the geodesic motion of the particles, i.e.,

$$0 = \frac{df}{dt} = \dot{x}^\mu \partial_\mu f + \dot{p}^i \frac{\partial f}{\partial p^i}\,,$$

$$0 = p^\mu \partial_\mu f - \Gamma^i_{\mu\nu} p^\mu p^\nu \frac{\partial f}{\partial p^i} \equiv L_{X_g} f\,. \tag{1.85}$$

A particle distribution obeying this equation is often also called a geodesic spray (see Abraham & Marsden, 1982). If the particles are not free, but collisions are so rare that an equilibrium description is not adequate, one uses the Boltzmann equation,

$$L_{X_g} f = C[f]\,, \tag{1.86}$$

where $C[f]$ is the so-called 'collision integral' which depends on the details of the interactions.

It may be disturbing to some readers that we take over these concepts from non-relativistic physics so smoothly to the relativistic case. In cosmology, this does not cause any problems. But in general, it is true that the collision integral is not always

well defined and certain conditions have to be posed to the nature of the spacetime and of the interaction. This problem has been studied in detail by Ehlers (1971).

Since the photons are massless, $|\mathbf{p}|^2 = \gamma_{ij}p^i p^i = (p^0)^2$. Here p^0 is the 0-component of the momentum 4-vector in *conformal* time so that $\epsilon = ap^0$ is the physical photon energy. Isotropy of the distribution implies that f depends on p^i only via $p \equiv |\mathbf{p}| = p^0$, and so

$$\frac{\partial f}{\partial p^i} = \frac{\partial p}{\partial p^i}\frac{\partial f}{\partial p} = \frac{p_i}{p}\frac{\partial f}{\partial p}\,. \tag{1.87}$$

Furthermore, f depends on x^i only through $p = \sqrt{\gamma_{ij}p^i p^i}$. Spatial derivatives are therefore given by

$$\begin{aligned} p^i\partial_i f &= p^i\gamma_{lm,i}\frac{p^l p^m}{p}\frac{\partial f}{\partial p} = p_j\gamma^{ij}\gamma_{lm,i}\frac{p^l p^m}{p}\frac{\partial f}{\partial p} \\ &= \frac{1}{2}\gamma^{ij}\left(\gamma_{li,m}+\gamma_{mi,l}-\gamma_{lm,i}\right)\frac{p_j p^l p^m}{p}\frac{\partial f}{\partial p} \\ &= \Gamma^j_{lm}\frac{p^l p^m p_j}{p}\frac{\partial f}{\partial p}\,. \end{aligned}$$

This leads to

$$p^i\partial_i f - \Gamma^i_{\mu\nu}\frac{p^\mu p^\nu p_i}{p}\frac{\partial f}{\partial p} = -\left(\Gamma^i_{j0}+\Gamma^i_{0j}\right)\frac{p^j p p_i}{p}\frac{\partial f}{\partial p} = -2p^2\frac{\dot a}{a}\frac{\partial f}{\partial p}\,,$$

where we have used the expressions in Appendix A2.3 for $\Gamma^i_{\mu\nu}$ and $p = p^0$. Inserting this result into (1.85) we obtain, with Eq. (1.87),

$$\partial_t f - 2p\frac{\dot a}{a}\frac{\partial f}{\partial p} = 0\,, \tag{1.88}$$

which is satisfied by an arbitrary function $f = f(pa^2) = f(a\epsilon)$. Hence the distribution of free-streaming photons changes only by redshifting the *physical energy* $\epsilon = ap^0$ or the *physical momentum* $a|\mathbf{p}| = \epsilon$. Therefore, setting $T \propto a^{-1}$ even after recombination, the blackbody shape of the photon distribution remains unchanged. This radiation of free photons with a perfect blackbody spectrum is the CMB. Its physics, especially its fluctuation and polarization are the main topic of this book.

The same result is also obtained for massive particles,

$$\partial_t f - 2p\frac{\dot a}{a}\frac{\partial f}{\partial p} = 0\,, \tag{1.89}$$

where $p = |\mathbf{p}|$; hence the momentum is simply redshifted. Therefore, massive particles which decouple when they are still relativistic, keep their extremely relativistic Fermi–Dirac (or Bose–Einstein) distribution, $f = (\exp(ap/T) \pm 1)$, with a

temperature which simply scales as $T \propto 1/a$. This is especially important for the cosmic neutrinos which probably have masses in the range of a few eV $> m_\nu \gtrsim$ 0.01 eV. But, as we shall see in the next section, they decouple at $T \sim 1.4$ MeV. We therefore expect them to be distributed according to an extremely relativistic Fermi–Dirac distribution.

Note however, that after decoupling the particles are no longer in thermal equilibrium and the T in their distribution function is not a temperature in the thermodynamical sense but merely a parameter, representing a measure of the mean kinetic energy.

The situation is different for the electron–proton–hydrogen plasma. The free electrons still scatter with photons and keep the same temperature as the latter. In other words: even though most photons are no longer interacting with the electrons, the latter are still interacting with the photons. (To have one collision with all the remaining electrons, only a fraction of about 10^{-14} of the photons have to be involved!)

Soon after recombination, the baryon energy density exceeds the photon energy density and one might expect that this would change the evolution of the temperature. To investigate this we use the energy conservation equation of the baryon–photon system. We neglect the tiny number of free electrons. The energy density and pressure are then given by

$$\rho = n_B m_B + (3/2) n_B T + \frac{\pi^2}{15} T^4 , \tag{1.90}$$

$$p = n_B T + \frac{\pi^2}{45} T^4 . \tag{1.91}$$

The energy conservation equation, $d\rho/da = -3(\rho + p)/a$ now gives

$$\frac{a}{T}\frac{dT}{da} = -\frac{3n_B + \frac{4\pi^2}{15} T^3}{(3/2) n_B + \frac{4\pi^2}{15} T^3} = -\frac{\sigma + 1}{\sigma + 1/2} . \tag{1.92}$$

Since $\sigma \gg 1$, the photons are so much more numerous than the baryons that the latter have no influence on the temperature which keeps evolving as $1/a$. Note, however, that in the absence of photons, the temperature of a mono-atomic gas would decrease like $1/a^2$ (just consider the limit $\sigma \to 0$).

The blackbody spectrum of the CMB photons is extremely well verified observationally (see Fig. 1.7 and Chapter 8). The limits on deviations are often parametrized in terms of three parameters: the chemical potential μ, the Compton-y parameter (which quantifies a well defined change in the spectrum arising from interactions with a non-relativistic electron gas at a different temperature, see Chapter 8) and $Y_{\rm ff}$ (describing a contamination by free–free emission).

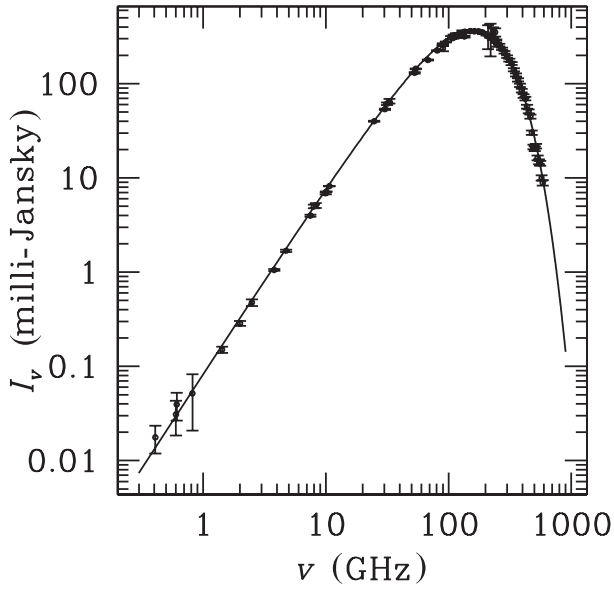


Fig. 1.7. The spectrum of the cosmic background radiation. The data are from many different measurements which are all compiled in Kogut *et al.* (2006). The points around the top are the measurements from the FIRAS experiment on COBE (Fixsen *et al.*, 1996). The line traces a blackbody spectrum at a temperature of 2.728 K (the data are courtesy of Susan Staggs).

The present 95% confidence limits on these parameters are (Particle Data Group, 2006)

$$|\mu| < 9 \times 10^{-5}, \quad |y| < 1.2 \times 10^{-5}, \quad |Y_{\rm ff}| < 1.9 \times 10^{-5}. \tag{1.93}$$

The CMB photons not only have a very thermal spectrum, but they are also distributed very isotropically, apart from a dipole which is (most probably) mainly due to our motion relative to the surface of last scattering.

Indeed, an observer moving with velocity $\mathbf{v}$ relative to a source in direction $\mathbf{n}$ emitting a photon with proper momentum $\mathbf{p} = -\epsilon\mathbf{n}$ sees this photon redshifted with frequency

$$\epsilon' = \gamma\epsilon\,(1 - \mathbf{nv})\,, \tag{1.94}$$

where $\gamma = 1/\sqrt{1 - v^2}$ is the relativistic γ-factor. For an isotropic emission of photons coming from all directions $\mathbf{n}$ this leads to a dipole anisotropy to first order in $\mathbf{v}$. This dipole anisotropy, which is of the order of

$$\left(\frac{\Delta T}{T}\right)_{\rm dipole} \simeq 1.2 \times 10^{-3}\,,$$

has already been discovered in the seventies (Conklin, 1969; Henry, 1971). Interpreting it as due to our motion with respect to the last scattering surface implies a velocity for the solar system barycentre of $v = 371 \pm 0.5\,\mathrm{km\,s^{-1}}$ at 68% CL (Particle Data Group, 2006).

In addition to the dipole, the COBE[4] DMR experiment (differential microwave radiometer) has found fluctuations of order

$$\sqrt{\left\langle \left(\frac{\Delta T}{T}\right)^2 \right\rangle} \sim (\text{a few}) \times 10^{-5}\,, \tag{1.95}$$

on all angular scales $\theta \geq 7^\circ$ (Smoot *et al.*, 1992). On smaller angular scales many experiments found fluctuations (we shall describe the experimental results in more detail later), but all of them satisfy $|\Delta T/T| \lesssim 10^{-4}$.

As we shall see in Chapter 2, the CMB fluctuations on large scales provide a measure for the deviation of the geometry from the Friedmann–Lemaître one. The geometry perturbations are thus small, and we may calculate their effects by *linear perturbation theory*. On smaller scales, $\Delta T/T$ reflects the fluctuations in the energy density in the baryon/radiation plasma prior to recombination. Their amplitude is just about right to allow the formation of the presently observed non-linear structures (like galaxies, clusters, etc.) by gravitational instability.

These findings strongly support our hypothesis that the large-scale structure (i.e., the galaxy distribution) observed in the Universe has been formed by gravitational instability from small ($\sim 10^{-4}$) initial fluctuations. As we shall see in Chapters 2, 4 and 5, such initial fluctuations leave an interesting 'fingerprint' on the cosmic microwave background.

1.4 Nucleosynthesis

1.4.1 Expansion dynamics at $T \sim$ a few MeV

At high temperatures, $T > 30$ MeV, none of the light nuclei (deuterium, ^{2}H, helium-4, ^{4}He, helium-3, ^{3}He or lithium, ^{7}Li) are stable. At these temperatures, we expect the baryons to form a simple mixture of protons and neutrons in thermal equilibrium with each other and with electrons, photons and neutrinos. The highest binding energy is the one of ^{4}He which is about 28 MeV. Nevertheless, ^{4}He cannot form at this temperature since the baryon density of the Universe is not high enough for three- or even four-body interactions to occur in thermal equilibrium. Therefore, before any nucleosynthesis can occur, the temperature has to drop below the binding energy of deuterium which is about 2.2 MeV. But even at this temperature there

[4] Cosmic Background Explorer, NASA satellite launched 1990.

are still far too many high-energy photons around for deuterium to be stable. This is due to the very low baryon to photon ratio $\eta_B \simeq 10^{-10}$. Just as recombination is delayed from the naively expected temperature $T = 13.7\,\mathrm{eV}$ to about $T_{\mathrm{rec}} \sim 0.3\,\mathrm{eV}$, nucleosynthesis does not happen at $T \sim 2.2\,\mathrm{MeV}$ but around $T_{\mathrm{nuc}} \sim 0.1\,\mathrm{MeV}$. Most of the neutrons present at that temperature are converted into ^{4}He. Only small traces remain as deuterium or are burned into ^{3}He and ^{7}Li.

Let us study this in some more detail. At the time of recombination, the relativistic particle species are the photon and, probably three types of neutrinos. As we shall see in the next paragraph, the neutrino temperature is actually a factor of $(4/11)^{1/3}$ lower than the temperature of the photons. With Eqs. (1.54) and (1.55), the energy density of these particles while they are relativistic is given by

$$\rho_{\mathrm{rel}}(t) = \left[\rho_\gamma(t) + \rho_\nu(t)\right] = \left[1 + 3\frac{7}{8}(4/11)^{4/3}\right]\frac{\pi^2}{15}T^4\,, \tag{1.96}$$

$$\simeq 10^{-33}\,\mathrm{g\,cm^{-3}}\left(\frac{T}{T_0}\right)^4\,, \tag{1.97}$$

$$\simeq \rho_c(t_0)\Omega_{\mathrm{rel}}h^2(1+z)^4\,,\ \text{where}$$

$$\Omega_{\mathrm{rel}}h^2 \simeq 4.4\times 10^{-5}\,. \tag{1.98}$$

Note that at temperatures below the highest neutrino mass, this is no longer the energy density of relativistic particles, therefore Ω_{rel} is not the density parameter of relativistic particles today. Above the neutrino mass threshold and below the electron mass threshold we have

$$\frac{\rho_{\mathrm{rel}}}{\rho_m} = \frac{\Omega_{\mathrm{rel}}}{\Omega_m}(1+z) \simeq 4.4\times 10^{-5}\left(\frac{1}{\Omega_m h^2}\right)(1+z)\,, \tag{1.99}$$

Since $\Omega_m h^2 \simeq 0.15$, the redshift z_{eq} above which the Universe is dominated by relativistic particles is about

$$z_{\mathrm{eq}} \simeq 3.4\times 10^3\,, \quad T_{\mathrm{eq}} \simeq 1\,\mathrm{eV}\,. \tag{1.100}$$

At temperatures significantly above T_{eq}, we can also neglect a possible contribution from curvature or a cosmological constant to the expansion of the Universe, so that for

$$z \gg z_{\mathrm{eq}} \qquad P = \frac{1}{3}\rho\,, \quad a \propto \tau^{1/2} \propto t\,. \tag{1.101}$$

At these high temperatures the energy density of the Universe is given by

$$\rho = g_{\mathrm{eff}}\frac{\pi^2}{30}T^4 \quad \text{where} \quad g_{\mathrm{eff}} = N_B(T) + \frac{7}{8}N_F(T)\,. \tag{1.102}$$

Here, N_B and N_F denote the number of bosonic and fermionic degrees of freedom of relativistic particles (i.e., particles with mass $m < T$) which are in thermal equilibrium at temperature T.

To discuss the physical processes at work at some temperature T, we need to know the spectrum of relativistic particles and their interactions at this temperature. Here, we shall study the Universe at $10\,\mathrm{keV} < T < 100\,\mathrm{MeV}$ where the physics is well known. The only relativistic particles present at these temperatures are electrons, positrons, photons and three types of neutrinos. Even if the individual neutrino masses are not very well constrained, the oscillation experiments (Particle Data Group, 2004) imply that their masses are below 1 eV if there is no degeneracy. Therefore, we may neglect the neutrino masses in our treatment. The baryon number is well conserved at these temperatures, so that we may set η_B equal to its present value, $\eta_B = n_B/n_\gamma \simeq 2.7 \times 10^{-8}\Omega_B h^2 = \text{constant}$. We neglect the small contribution from muon/anti-muon pairs which decay exponentially $\propto \exp(-m_\mu/T)$ via the reaction

$$\mu + \bar{\nu}_\mu \to e + \bar{\nu}_e\,.$$

Thermal equilibrium between photons and electron/positrons is maintained mainly via the process $e^- + e^+ \longleftrightarrow 2\gamma$ (or $3\gamma\ldots$). The conservation of the chemical potential during this reaction implies

$$\mu_e + \bar{\mu}_e = 2\mu_\gamma = 0\,. \tag{1.103}$$

The last equals sign comes from the fact that photons can be generated and destroyed, their number is not conserved, and hence their chemical potential vanishes in thermal equilibrium. Here we use the notation $e^+ = \bar{e}$ and $\mu_{\bar{e}} = \bar{\mu}_e$. The difference in the density of electrons and positrons is therefore

$$n_e - \bar{n}_e = \frac{1}{\pi^2}\int p^2\,dp\left[\frac{1}{\exp\left(\frac{E-\mu_e}{T}\right)+1} - \frac{1}{\exp\left(\frac{E+\mu_e}{T}\right)+1}\right]. \tag{1.104}$$

At low temperatures this number is dictated by the neutrality of the Universe, $n_e - \bar{n}_e \sim n_B$ is much smaller than $n_e + \bar{n}_e \sim n_\gamma$. Therefore, the chemical potential is much smaller than the electron mass, $\mu_e \ll m_e$. At high temperatures, $T \gg m_e$ we may therefore approximate the electron number density by

$$n_e - \bar{n}_e \simeq \frac{2\mu_e}{\pi^2 T}\int p^2\,dp\frac{\exp(p/T)}{\left[\exp(p/T)+1\right]^2} = \frac{2\mu_e T^2}{\pi^2}\zeta(2)\,. \tag{1.105}$$

With $n_\gamma = 2T^3\zeta(3)/\pi^2$ this yields

$$\frac{n_e - \bar{n}_e}{n_\gamma} \simeq 1.4\frac{\mu_e}{T} \sim \frac{n_B}{n_\gamma} \simeq 1.4 \times 10^{-8}\Omega_B h^2. \tag{1.106}$$

We can therefore neglect the small chemical potential of the electrons and positrons. The interaction $e + \bar{e} \longleftrightarrow \nu + \bar{\nu}$ also implies that $\mu_\nu = -\bar{\mu}_\nu$. But unfortunately, the number $n_\nu - \bar{n}_\nu$ which determines, together with $n_e - \bar{n}_e$, the lepton number of the Universe is not known from observations. We suppose that the lepton number, like the baryon number, is small and that we may also neglect the chemical potential of the neutrinos. Comparing our results with observations, we can check this hypothesis later.

At $T \lesssim 100$ MeV photons, electron/positrons and neutrinos are still relativistic, so that $N_B = 2$ and $N_F = 4 + 6$, hence

$$g_{\text{eff}}(T \sim 100\,\text{MeV}) = \frac{43}{4} = 10.75\,. \tag{1.107}$$

The Hubble parameter is therefore given by

$$\left(\frac{a'}{a}\right)^2 = H^2 = \frac{1}{4\tau^2} = \frac{8\pi G}{3}\rho = \frac{8\pi^3 G}{90} g_{\text{eff}} T^4\,.$$

With the Planck mass, m_P defined by $G = 1/m_P^2 = 1/(1.22 \times 10^{19}\,\text{GeV})^2$, this gives

$$H^2(T) \simeq 2.76 g_{\text{eff}}(T)\left(\frac{T^2}{m_P}\right)^2, \tag{1.108}$$

$$H \simeq 0.21\sqrt{g_{\text{eff}}}\left(\frac{T}{1\,\text{MeV}}\right)^2 \text{s}^{-1}, \tag{1.109}$$

$$\tau = \frac{1}{2H} \simeq 0.3 g_{\text{eff}}(T)^{-1/2}\left(\frac{m_P}{T^2}\right) \simeq 2.3\,\text{s}\left(\frac{1\,\text{MeV}}{T}\right)^2 g_{\text{eff}}(T)^{-1/2}\,. \tag{1.110}$$

The temperature of $T \sim 100$ MeV corresponds thus to an age of $\tau \sim 7 \times 10^{-5}$ s, and $T = 1$ MeV corresponds to $\tau \sim 0.7$ s. The relations (1.109) and (1.110) can be applied as long as the Universe is dominated by relativistic particles.

1.4.2 Neutrino decoupling

Neutrinos are kept in thermal equilibrium via the exchange of a W-boson, $e + \bar{\nu} \longleftrightarrow e + \bar{\nu}$ and $\nu + \bar{e} \longleftrightarrow \nu + \bar{e}$, or a Z-boson, $e + \bar{e} \longleftrightarrow \nu + \bar{\nu}$. At low energies $E \ll m_{Z,W} \sim 100$ GeV, we can determine the cross sections within the 4-fermion theory of weak interaction. Within this approximation, the effective

interaction Langrangian is given by

$$L_{\rm int} = \frac{G_F}{\sqrt{2}} J_\mu^\dagger J^\mu + \text{hermitean conjugate}$$
$$= \frac{G_F}{\sqrt{2}} \left(\bar{u}_e \gamma_\mu \frac{1}{2}(1-\gamma^5) u_\nu \right) \left(u *_\nu \gamma^\mu \frac{1}{2}(1-\gamma^5) u_e \right) + \text{h.c.}\,, \quad (1.111)$$

where the coupling parameter, G_F is the Fermi constant,

$$G_F = 1.166 \times 10^{-5}\,\text{GeV}^{-2} = (293\,\text{GeV})^{-2}. \quad (1.112)$$

The fermion $V-A$ current J_μ is expressed in terms of the electron and neutrino spinors $u_{e,\nu}$ and the Dirac γ-matrices, γ_μ and γ_5.

The cross section of the different processes above are identical within this approximation and they are given by

$$\sigma_F \simeq G_F^2 E^2 \sim G_F^2 T^2\,,$$

The involved particle density is $n_F(T) = g_F(T)\zeta(3)T^3/\pi^2 \sim 1.3T^3$ where we have set $g_F(T) = 3/4 N_F(T) = 30/4$ for the three types of left-handed neutrinos and the $e^\pm$ s. Since the particles are relativistic, we can set $v \sim 1$ so that we obtain an interaction rate of

$$\Gamma_F = \langle \sigma_F v \rangle n_F \simeq 1.3 G_F^2 T^5\,.$$

Comparing this with the expansion rate H obtained in (1.109), we find

$$\frac{\Gamma_F}{H} \simeq 0.24 T^3 m_P G_F^2 \simeq \left(\frac{T}{1.4\,\text{MeV}} \right)^3 . \quad (1.113)$$

At temperatures below $T_F \sim 1.4$ MeV the probability for a neutrino to interact within one Hubble time, H^{-1}, becomes less than unity and the neutrinos effectively decouple. The plasma becomes transparent to neutrinos which are no longer in thermal equilibrium with electrons and positrons and hence photons and baryons.

As we have discussed in the previous section, even at temperatures far below their mass $m_\nu \gtrsim 0.01$ eV, their particle distribution remains an extremely relativistic Fermi–Dirac distribution with temperature

$$T_\nu = T_F \frac{a_F}{a}\,,$$

since they are no longer in thermal equilibrium and their distribution is affected solely by redshifting of the momenta.

As long as the photon/electron/baryon temperature also scales like $1/a$, the neutrinos conserve the same temperature as the thermal plasma, but when the number of degrees of freedom, $g_{\rm eff}$, changes, the plasma temperature decays for

a brief period of time less rapidly than $1/a$ and therefore remains higher than the neutrino temperature. This is exactly what happens at the electron–positron mass threshold, $T = m_e \simeq 0.5$ MeV. Below that temperature, only the process $e + \bar{e} \to 2\gamma$ remains in equilibrium while $2\gamma \to e + \bar{e}$ is exponentially suppressed. We calculate the reheating of the photons gas by electron–positron annihilation assuming that the process takes place in thermal equilibrium and that the entropy remains unchanged. This is well justified since the cross section of this process is very high. Denoting the entropy inside a volume of size Va^3 before and after electron–positron annihilation by S_i and S_f we therefore have $S_i = S_f$. Hence

$$S_i = \frac{4}{3} a_{SB} g_{\mathrm{eff},i} (Ta)_i^3 V, \quad S_f = \frac{4}{3} a_{SB} g_{\mathrm{eff},f} (Ta)_f^3 V \,.$$

The electron–positron degrees of freedom disappear in this process so that $g_{\mathrm{eff},f} = 2$ while $g_{\mathrm{eff},i} = 2 + 4(\frac{7}{8}) = 11/2$. From $S_i = S_f$ we therefore conclude

$$(Ta)_f = (Ta)_i \left(\frac{11}{4}\right)^{1/3} .$$

The neutrino temperature is not affected by $e^{\pm}$ annihilation, so that $(T_\nu a)_f = (T_\nu a)_i = (Ta)_i$. For the last equals sign we have used that the neutrino and photon temperatures are equal before $e^{\pm}$ annihilation. At temperatures $T \ll m_e$ we therefore have

$$T = \left(\frac{11}{4}\right)^{1/3} T_\nu \,. \tag{1.114}$$

Since there are no further annihilation processes, this relation remains valid until today and the present Universe not only contains a thermal distribution of photons, but also a background of cosmic neutrinos which have an extremely relativistic Fermi–Dirac distribution with temperature

$$T_\nu(\tau_0) = (4/11)^{1/3} T_0 = 1.95\,\mathrm{K} \,. \tag{1.115}$$

We set

$$g_0 = 2 + \frac{7}{8} 6 \left(\frac{4}{11}\right)^{4/3} \simeq 3.36 \,, \quad \text{and} \tag{1.116}$$

$$g_{0S} = 2 + \frac{7}{8} 6 \left(\frac{4}{11}\right) \simeq 3.91 \,. \tag{1.117}$$

These are respectively the effective degrees of freedom of the energy and entropy densities as long as all the neutrinos are relativistic. Until then we therefore have

$$\rho_{\mathrm{rel}}(T) = \frac{\pi^2}{30} g_0 T^4 \simeq 8.1 \times 10^{-34}\,\mathrm{g\,cm^{-3}} \left(\frac{T}{T_0}\right)^4 , \tag{1.118}$$

$$s(T) = \frac{2\pi^2}{45} g_{0S} T^3 \simeq 3 \times 10^3\,\mathrm{cm^{-3}} \left(\frac{T}{T_0}\right)^3 . \tag{1.119}$$

The neutrino cross section at low energies is extremely weak, and so far the neutrino background has not been observed directly (see Ex. 1.7).

1.4.3 The helium abundance

The observed abundance of helium is universally about

$$\frac{n_{He}\, m_{He}}{n_H\, m_H} \equiv Y \simeq 0.24. \tag{1.120}$$

It is well known that this amount of helium cannot have been produced in stars. We now want to investigate how much helium is produced in the primordial Universe. At temperatures of a few MeV nuclei and baryons are non-relativistic and the equilibrium distribution for a nucleus with atomic mass (i.e., number of protons and neutrons) A and proton number Z is given by

$$n_A = N_A \left(\frac{m_A T}{2\pi}\right)^{3/2} \exp\left(-\frac{m_A - \mu_A}{T}\right). \tag{1.121}$$

The proton density is given in Eq. (1.70). The neutron density is correspondingly

$$n_n = 2\left(\frac{m_B T}{2\pi}\right)^{3/2} \exp\left(-\frac{m_n - \mu_n}{T}\right). \tag{1.122}$$

Here, we neglect the small difference $Q = m_n - m_p = 1.293$ MeV in the pre-factor, setting $m_n \sim m_p \sim m_B$. The conservation of the chemical potentials in nuclear reactions implies

$$\mu_A = Z\mu_p + (A - Z)\mu_n\,,$$

so that

$$\begin{aligned}\exp\left(-\frac{m_A - \mu_A}{T}\right) &= \left(e^{\mu_p/T}\right)^Z \left(e^{\mu_n/T}\right)^{(A-Z)} e^{-m_A/T}\,, \\ &= \frac{1}{2^A}\left(\frac{2\pi}{m_B T}\right)^{3A/2} \exp(B_A/T) n_p^Z n_n^{A-Z}\,.\end{aligned}$$

Here, $B_A = Zm_p + (A - Z)m_n - m_A$ is the binding energy of the nucleus (A, Z). In thermal equilibrium, the density of this ion is then given by

$$n_A = \frac{N_A}{2^A} A^{3/2} \left(\frac{2\pi}{m_B T}\right)^{3(A-1)/2} n_p^Z n_n^{A-Z} \exp(B_A/T). \tag{1.123}$$

Here we have again neglected the nucleon mass difference Q and the binding energy B_A in the pre-factor by setting $m_A \sim Am_B$, but not in the exponential.

We define the various mass abundances by

$$Y_A \equiv \frac{An_A}{n_B} = \frac{An_A}{\eta_B\, n_\gamma}\,,$$
$$Y_p \equiv \frac{n_p}{n_B} = \frac{n_p}{\eta_B\, n_\gamma}\,,$$
$$Y_n \equiv \frac{n_n}{n_B} = \frac{n_n}{\eta_B\, n_\gamma}\,.$$

Hence the thermal abundance of the nucleus (A, Z) is given by

$$Y_A = F(A)\left(\frac{T}{m_B}\right)^{3(A-1)/2} \eta_B^{A-1} Y_p^Z Y_n^{A-Z} e^{B_A/T}\,, \qquad (1.124)$$

$$\text{where} \quad F(A) = N_A A^{5/2} \zeta(3)^{A-1} \pi^{-(A-1)/2} 2^{(3A-5)/2}\,. \qquad (1.125)$$

This equation shows nicely the influence of the radiation entropy on nucleosynthesis. If we had $\eta_B \sim 1$, the nucleus (A, Z) would become stable and relatively abundant at $T \sim B_A$. At this temperature the formation of (A, Z) (controlled by the factor $\exp(B_A/T)$) is sufficiently important to counterbalance photo-dissociation (controlled by the factor η_B^{A-1}). In equilibrium, the exponential $\exp(B_A/T)$ is then of the order of $\eta_B^{1-A} \sim 1$ and the ratio Y_A then approaches the value $Y_A \sim Y_p^Z Y_n^{A-Z}$. However, if η_B is very small, the equilibrium between production of (A, Z) and photo-dissociation is delayed until $\exp(-B_A/T) \sim \eta_B^{A-1} \ll 1$, i.e., to much lower temperatures. Neglecting the numerical factor $F(A)$, the temperature T_A, defined by $Y_A(T_A) \sim Y_p(T_A)^Z Y_n(T_A)^{A-Z}$, is

$$T_A \sim \frac{B_A}{(A-1)\left[\ln(\eta_B^{-1}) + 3/2 \ln(m_B/T_A)\right]}\,.$$

For the deuteron with binding energy $B_2 = 2.22$ MeV we find

$$T_2 \sim 0.085\,\text{MeV}\,. \qquad (1.126)$$

The reaction rate Γ_{np} of the process $n + p \longleftrightarrow {}^2\text{H} + \gamma$ is given by

$$\Gamma_{np} = \langle \sigma_{np} v \rangle n_p \simeq 1.8 \times 10^{-17} (T/T_0)^3 \eta_B\, \text{s}^{-1} \simeq 10^{12} \eta_B \left(\frac{T}{\text{MeV}}\right)^3 \text{s}^{-1}\,,$$

where we have used $\langle \sigma_{np} v \rangle = \text{constant} = 4.55 \times 10^{-20}\,\text{cm}^3\,\text{s}^{-1}$ at temperatures $1\,\text{keV} \leq T \leq 10\,\text{MeV}$, and $n_p = \eta_B n_\gamma \simeq 420 \eta_B (T/T_0)^3\,\text{cm}^{-3}$. Using $H \simeq 0.4 (T/\text{MeV})^2\,\text{s}^{-1}$, we conclude that this interaction remains in thermal equilibrium as long as $T \gtrsim 0.004$ MeV. So the assumption of a thermal deuterium abundance is justified. As already mentioned, three-body interactions are not in thermal equilibrium, their reaction rate contains an additional factor $n_B/n_\gamma = \eta_B \ll 1$.

Therefore, at temperature T_2 only deuterium can form and subsequently virtually all the neutrons present are burned into ^{4}He. To determine the helium abundance, we have to determine the neutron density at this temperature. Let us first determine the temperature at which β and inverse β processes drop out of equilibrium,

$$\nu + n \longleftrightarrow p + e, \quad \bar{e} + n \longleftrightarrow p + \bar{\nu}, \quad n \to p + e + \bar{\nu}\,.$$

On one hand, particle conservation imposes

$$\mu_n - \mu_p = \mu_e - \mu_\nu\,.$$

On the other hand, the neutrality of the Universe requires $n_p = n_e$. Since $m_e \ll m_p$, the Eqs. (1.69) and (1.70) imply $\mu_e \ll \mu_p$. Finally, setting $\mu_\nu \sim 0$, the chemical potentials of the neutron and the proton are approximately equal, i.e., $\mu_n \simeq \mu_p$. The ratio of their densities is thus simply given by the mass difference $Q = m_n - m_p$,

$$\frac{n_n}{n_p} = \frac{Y_n}{Y_p} = \exp(-Q/T)\,.$$

This ratio remains constant as long as the reactions $n \longleftrightarrow p$ are sufficiently rapid. At the decoupling temperature of these reactions,

$$\Gamma(T_D) = H(T_D) \simeq 3\frac{T_D^2}{m_P}\,,$$

the ratio (n_n/n_p) is hence given by

$$\left(\frac{n_n}{n_p}\right)(T_D) = \exp(-Q/T_D)\,.$$

Afterwards, the neutron density decays exponentially by β-decay, $n \to p + e + \bar{\nu}$,

$$n_n(\tau) = n_n(\tau_D)\exp\left(-\frac{\tau - \tau_D}{\tau_n}\right) \quad \text{for } \tau > \tau_D, \tag{1.127}$$

where $\tau_n \simeq 886$ s is the neutron lifetime.

We now want to determine the temperature T_D. We can again use Fermi theory to determine the different cross sections. For nucleons, the pure $V - A$ current, $\bar{\psi}\gamma_\mu(1 - \gamma_5)\psi$, is replaced by $\bar{\psi}\gamma_\mu(g_V + g_A\gamma_5)\psi$ which takes into account the internal structure of the nucleons. In the Born approximation the cross section becomes, see e.g., Maggiore (2005)

$$\sigma(\nu + n \to p + e) = \frac{G_F^2}{\pi}(g_V^2 + 3g_A^2)v_e E_e^2\,.$$

The constants g_V and g_A are determined experimentally (e.g., by measuring the neutron lifetime), $g_V \simeq 1.00$ and $g_A \simeq 1.25$. The interaction rate per neutron is

obtained by multiplying the above result with $v_\nu n_\nu$,

$$\Gamma(\nu + n \to p + e) = \langle \sigma v_\nu \rangle n_\nu = \frac{1}{2\pi^2} \int \frac{p_\nu^2 \, dp_\nu}{e^{p_\nu/T_\nu} + 1} v_\nu \sigma \left(1 - \frac{1}{e^{E_e/T} + 1}\right).$$

The factor $1 - 1/[\exp(E_e/T) + 1]$ is the probability that the electron state with energy E_e is free (it implements the Pauli principle). To simplify the integral we first use energy conservation, $E_\nu + E_n = E_p + E_e$. Since all the energies involved are of the order of MeV, we can set $E_n - E_p \sim m_n - m_p = Q = 1.293\,\text{MeV}$ and $E_e = p_\nu + Q$. Furthermore, $E_e = m_e\gamma = m_e/\sqrt{1 - v_e^2}$ which implies $v_e = \sqrt{(p_\nu + Q)^2 - m_e^2}/E_e$. Inserting these simplifications, we obtain finally

$$\Gamma(\nu + n \to p + e) = \frac{G_F^2(g_V^2 + 3g_A^2)m_e^5}{2\pi^3} \times \int_0^\infty \frac{e^{\alpha(x+q)} x^2 (x+q)\sqrt{(x+q)^2 - 1}}{(1 + e^{\alpha(x+q)})(1 + e^{\beta x})} \, dx \,, \tag{1.128}$$

where we have set $x = p_\nu/m_e, \alpha = m_e/T_\gamma, \beta = m_e/T_\nu$ and $q = Q/m_e \simeq 2.5$. To compute the other processes we note that the matrix element $\mathcal{M}(p_\nu, p_n, p_p, p_e)$ which appears in the amplitude for $\nu + n \longleftrightarrow p + e$ is invariant under the transformations $(\mathbf{p}_\nu, p_n, p_p, \mathbf{p}_e) \to (-\mathbf{p}_\nu, p_n, p_p, -\mathbf{p}_e)$ and $(\mathbf{p}_\nu, p_n, p_p, \mathbf{p}_e) \to (-\mathbf{p}_\nu, p_n, p_p, \mathbf{p}_e)$, where p_ν, p_n, p_p and p_e are the momenta of the neutrino, neutron, proton and electron respectively,

$$\mathcal{M}(p_\nu, p_n, p_p, p_e) = \mathcal{M}(-p_\nu, p_n, p_p, -p_e) \,,$$
$$\mathcal{M}(p_\nu, p_n, p_p, p_e) = \mathcal{M}(-p_\nu, p_n, p_p, p_e) \,.$$

This observation allows us immediately to determine the reaction rates of the other processes. We simply have to take into account the different phase space constraints. With $x = E_e/m_e$, (the other parameters as above) we obtain

$$\Gamma(e + p \to n + \nu) = \frac{G_F^2(g_V^2 + 3g_A^2)m_e^5}{2\pi^3} \times \int_q^\infty \frac{e^{\beta(x-q)} x (x-q)^2 \sqrt{x^2 - 1}\, dx}{(1 + e^{\beta(x-q)})(1 + e^{\alpha x})} \,, \tag{1.129}$$

and

$$\Gamma(n \to p + e + \bar{\nu}) \simeq \frac{G_F^2(g_V^2 + 3g_A^2)m_e^5}{2\pi^3} \times \int_1^q \frac{e^{\alpha x} e^{\beta(q-x)} (x-q)^2 x \sqrt{x^2 - 1}\, dx}{(1 + e^{\beta(q-x)})(1 + e^{\alpha x})} \,, \tag{1.130}$$

$$\Gamma(n \to p + e + \bar{\nu})|_{T \ll m_e} \simeq 1.6 \frac{G_F^2}{2\pi^3}(g_V^2 + 3g_A^2)m_e^5 = \tau_n^{-1} \tag{1.131}$$

$$\tau_n^{-1} = \frac{1}{886\ \text{s}}$$

for the β-decay of the neutron at low temperature.

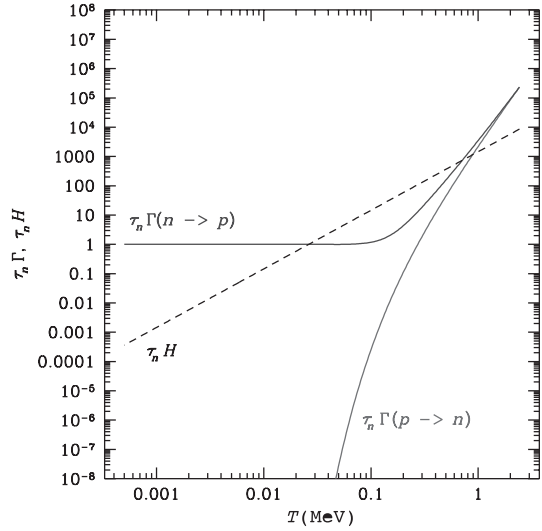


Fig. 1.8. The weak interaction rates, $\tau_n\Gamma(p \to n)$ and $\tau_n\Gamma(n \to p)$, are shown as functions of the temperature. The expansion rate, $\tau_n H$, is also indicated.

The products $\tau_n\Gamma$ are functions of the temperature T. When $T \gg Q$, the kinetic energy in the system $e + \bar{\nu}$ is much higher than the electron mass. Hence $x \pm q \simeq x$ at the positions which contribute most to the above integrals and the reaction rates go like

$$\left.\begin{array}{l}\tau_n\Gamma(n \to p)\\ \tau_n\Gamma(p \to n)\end{array}\right\} \propto T^5\,, \qquad \text{for } T \gg Q\,.$$

In the regime $0.1\,\text{MeV} \leq T \leq 1\,\text{MeV}$, the product $\tau_n\Gamma(n \to p)$ is roughly proportional to $T^{4.4}$. The same is true for $\tau_n\Gamma(p \to n)$. But the phase space for β-decay is larger than for the reaction $p \to n$, so that $\tau_n\Gamma(n \to p) > \tau_n\Gamma(p \to n)$. Once the temperature drops below about 0.1 MeV, $\tau_n\Gamma(p \to n)$ decays exponentially while $\tau_n\Gamma(n \to p)$ converges to 1 (see Fig. 1.8, where $\tau_n\Gamma(n \to p)$, $\tau_n\Gamma(p \to n)$ and the expansion rate $\tau_n H$ are shown as functions of the temperature).

According to Fig. 1.8, the line $\tau_n H$ intersects the lines $\tau_n\Gamma(n \to p)$ and $\tau_n\Gamma(p \to n)$ around $T = 0.8$ MeV. A more detailed analysis gives a decoupling temperature of $T_D \simeq 0.7$ MeV, below which the three reactions are no longer in thermal equilibrium.

Another way to see this dropping out of the thermal equilibrium of weak interaction is to compare the true neutron abundance, Y_n, with the one obtained in thermal equilibrium. A semi-analytical calculation gives (see Bernstein *et al.*, 1989) the behaviour plotted in Fig. 1.9.

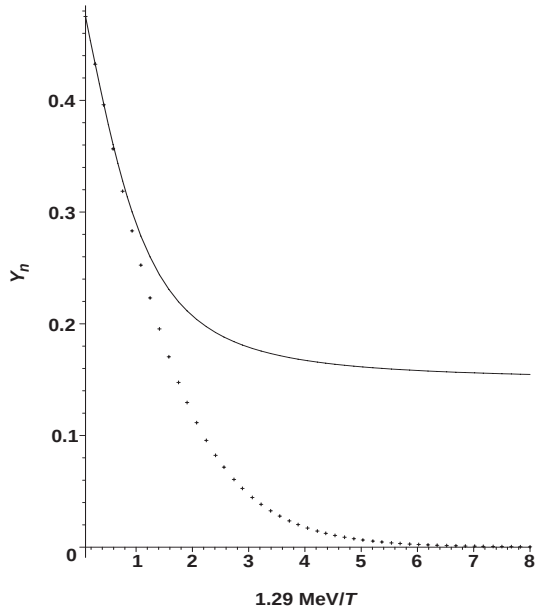


Fig. 1.9. The true neutron abundance as a function of $\Delta m/T$ (solid line) is compared with the equilibrium abundance (dotted line). Clearly, weak interaction freezes out around $T \sim 0.6 \times \Delta m \sim 0.7$ MeV.

At decoupling, the ratio of the neutron to proton density is

$$\left(\frac{n_n}{n_p}\right)(T_D) = \exp(-Q/T_D) \simeq 1/6\,, \tag{1.132}$$

so that

$$Y_n = 1/7 \quad \text{and} \quad Y_p = 6/7\,. \tag{1.133}$$

Since T_2, the temperature of deuterium formation, is lower than T_D, in the interval $T_D > T > T_2$, neutrons simply β-decay. At τ_2 given by $T_2 = T(\tau_2) = 0.085$ MeV their density is

$$\left(\frac{n_n}{n_p}\right)(T_2) = e^{-Q/T_D} \exp(-\tau_2/\tau_n) \simeq 0.8/6 \simeq 1/7\,, \tag{1.134}$$

and therefore

$$Y_n = 1/8 \quad \text{and} \quad Y_p = 7/8\,. \tag{1.135}$$

For this we have used $t_2 \simeq 1.3\,\mathrm{s}\,(1/0.085)^2 \simeq 180$ s. Once deuterium is formed, helium-4 is very rapidly synthesized via the reactions

$$
\begin{aligned}
{}^2\mathrm{H} + {}^2\mathrm{H} &\longrightarrow n + {}^3\mathrm{He} \\
{}^3\mathrm{He} + {}^2\mathrm{H} &\longrightarrow p + {}^4\mathrm{He}
\end{aligned}
$$

$$
\begin{aligned}
{}^2\mathrm{H} + {}^2\mathrm{H} &\longrightarrow \mathrm{p} + {}^3\mathrm{H} \\
{}^3\mathrm{H} + {}^2\mathrm{H} &\longrightarrow n + {}^4\mathrm{He}
\end{aligned}
$$

$$
{}^2\mathrm{H} + {}^2\mathrm{H} \longrightarrow \gamma + {}^4\mathrm{He}
$$

and essentially all deuterium is transformed in ^{4}He. The helium abundance is thus in good approximation, given by half the neutron abundance at temperature $T_2 \simeq 0.085$ MeV. With this approximation we obtain a helium-4 abundance of

$$
Y_{^4He} = \frac{4(n_n/2)}{n_n + n_p} = \frac{2(n_n/n_p)}{n_n/n_p + 1} \simeq \frac{1}{4}. \tag{1.136}
$$

In this expression we have used the neutron abundance from Eq. (1.135). Considering that t_2 scales like $\sqrt{\log \eta_B}$ while T_D depends strongly on the expansion rate H which is proportional to $\sqrt{g_{\mathrm{eff}}} \propto \sqrt{N_\nu (4/11)^{4/3} + 1}$. We conclude that the helium-4 abundance is very sensitive on the number of neutrino families, but does not change very rapidly with η_B. Historically, the cosmological helium-4 abundance has been the first experimental data to determine the number of (light) neutrino families in the range $N_\nu = 3.24 \pm 1.2$, when allowing for very generous error bars in the measurements (Fields & Sarkar, 2006). Presently, the Z-boson decay width, which has been measured very accurately with the LEP accelerator at CERN, gives the tightest value (see Particle Data Group, 2006), $N_\nu = 3.07 \pm 0.12$ at 95% confidence.

1.4.4 Deuterium, helium-3 and lithium-7

Nucleosynthesis starts at $T \sim 0.1$ MeV, corresponding to $t \sim 130$ s and terminates after a few minutes. Apart from ^{4}He very small amounts of all other elements up to lithium-7 are formed (some deuterium, tritium and helium-3 remain unprocessed). All these elements except deuterium, helium-3 and lithium-7 decay radioactively and their primordial abundance can no longer be observed today.

The amount of deuterium and helium-3 which is not burned into helium-4 is a steep function of the baryon abundance in the Universe. The higher the baryon density, the more efficient is the conversion of deuterium and helium-3 into helium-4 (see Fig. 1.10). This can be used to determine the baryon density in the Universe very accurately. Measuring the primordial deuterium abundance is an art by itself

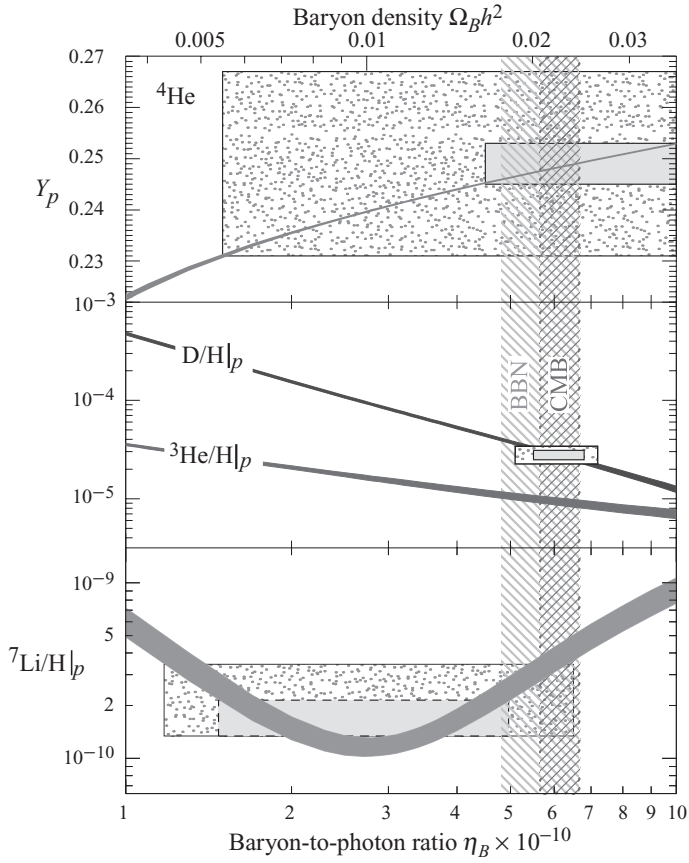


Fig. 1.10. The primordial element abundance as function of the parameter $\eta_B = n_B/n_\gamma$. The bands compatible with the observations of the different nuclei are indicated. The wide vertical band shows the range of η_B (or equivalently $\Omega_B h^2$) compatible with the nucleosynthesis data while the narrower dashed range is compatible with CMB anisotropies (see Chapter 6). Figure from Fields & Sarkar (2006).

on which we shall not dwell here. Most recent results are obtained by measuring it from the absorption lines in hydrogen (Ly-α) clouds intervening in the line of sight between us and quasars. Within generous error bars one obtains $2 \times 10^{-5} < Y_{^2H}/Y_p < 2 \times 10^{-4}$. This gives $4.7 \times 10^{-10} < \eta_B < 6.5 \times 10^{-10}$ (for more details see Olive *et al.* (2000), Burles *et al.* (2001) and Particle Data Group (2006)).

As one sees in Fig. 1.10, the lithium abundance is not a monotone function of η_B. This is so since, depending on the value of η_B, two different processes lead to lithium formation. If the baryon density is small, $\eta_B < 3 \times 10^{-10}$, lithium abundance is determined by the competition between the production process ^{4}He +

$^3\mathrm{H} \to {}^7\mathrm{Li} + \gamma$ and the destruction process ${}^7\mathrm{Li} + p \to {}^4\mathrm{He} + {}^4\mathrm{He}$. In this regime, the abundance decays with growing η_B. For $\eta_B > 3 \times 10^{-10}$, the dominant channel goes over beryllium production ${}^4\mathrm{He} + {}^3\mathrm{He} \to {}^7\mathrm{Be} + \gamma$ which is then converted into lithium-7 via the reaction ${}^7\mathrm{Be} + e \to {}^7\mathrm{Li} + \gamma$. The destruction process is the same as at low density. Since the conversion of beryllium into lithium increases with increasing baryon density, lithium abundance grows with η_B, for $\eta_B > 3 \times 10^{-10}$. The lithium abundance has a minimum around $\eta_B \simeq 3 \times 10^{-10}$. Inference of the primordial lithium abundance is still a matter of considerable debate. It nevertheless allows us to constrain $10^{-10} < \eta_B < 10^{-9}$.

Finally, in the regime $10^{-10} < \eta_B < 10^{-9}$ the helium-4 abundance is well approximated by the formula

$$Y_{^4He} = 0.23 + 0.011 \ln(\eta_{10}) + 0.013(N_\nu - 3) \,, \tag{1.137}$$

where we have introduced $\eta_{10} = \eta_B / 10^{-10}$. All the present observations of light elements taken together limit $4.7 < \eta_{10} < 6.5$, leading to $0.017 < \Omega_B h^2 < 0.024$ (a constantly updated review can be found in Particle Data Group (2006)). It is remarkable that this value is in very good agreement with the result obtained from measurements of the fluctuations in the CMB which are based on completely different physics (see Chapter 6).

This value is much larger than the density of luminous baryons which make up the stars and gas in the galaxies, and which lead only to $\Omega_L h^2 \simeq 0.004$. Hence most baryons in the Universe are not luminous. On the other hand, dynamical measurements and, more accurately the anisotropies in the CMB (see Chapter 6) require an energy density of non-relativistic matter today of about $\Omega_m h^2 \simeq 0.13$. To satisfy both constraints, the matter density of the Universe has to be dominated to about 80% by non-baryonic, so-called dark matter (dark in this context means that this matter does not interact with photons). So far, this dark matter has not been observed directly, but many experiments are underway and are starting to reach promising sensitivities. There are several candidates for dark matter particles. Most notably the lightest super-symmetric particle, but also the gravitino, axion or primordial black holes are viable candidates.

The good agreement of N_ν and $\Omega_B h^2$ obtained from the study of primordial nucleosynthesis with other experiments, confirms that the Universe has been in a thermal state, expanding adiabatically back to temperatures of the order of $T \sim 1$ MeV. For earlier times we have no experimental evidence. However, if the Universe has been in a thermal state at a temperature of $T \sim 200$ MeV, $\tau \sim 0.1$ s, it has then undergone a confinement transition leading from a quark gluon plasma at higher temperatures to baryons (such as the proton and neutron) and mesons (such as pions). If it has also been in thermal equilibrium at temperatures of up to $T \sim 200$

GeV, $\tau \sim 0.001$ s, it has then undergone the electroweak transition giving masses to the $W^{\pm}$ and Z bosons. At even higher temperatures we have no experimentally confirmed theory of fundamental interactions. Maybe, at $T \sim$ a few TeV the Universe becomes super-symmetric. Maybe, at $T \sim 10^{16}$ GeV a phase transition from a previous grand unified symmetry to the (super-symmetric) standard model symmetries took place. At this or higher energies the Universe may also have gone through (or emerged from) a super-string phase. To date such questions remain entirely speculative. Their quantitative investigation, especially possible observable signatures of a super-string phase is an active field of research.

1.5 Inflation

1.5.1 Cosmological problems

We first discuss the motivation for, and some consequences of a so-called 'inflationary phase'. We then exemplify the idea with a cosmology dominated by a scalar field. It is, however, clear that this realization has to be regarded as a toy model since the actual physical degrees of freedom relevant in the very early Universe, where such a period has most probably to be situated (see Chapter 3), are not known. In that sense this section is on a different level from the previous ones. We do not have any direct evidence that an inflationary phase has taken place in our Universe. Such a period just addresses several otherwise mysterious initial conditions of the observed Universe. The most significant observed 'prediction' of inflation is a nearly scale-invariant spectrum of initial fluctuations which we shall discuss in Chapter 3. What is more serious is that we have no 'direct' experimental evidence of the existence of an 'inflaton field'.

We include a possible cosmological constant into the energy density and the pressure, so that Eqs. (1.20) and (1.21) reduce to

$$\mathcal{H}^2 = \frac{8\pi G}{3} a^2 \rho - K \,, \tag{1.138}$$

$$\dot{\mathcal{H}} = -\frac{4\pi G}{3} a^2 (\rho + 3P) = \left(\frac{\ddot{a}}{a}\right) - \mathcal{H}^2 \,. \tag{1.139}$$

If $\rho + 3P > 0$ at all times, the homogeneous and isotropic cosmological model has several important problems.

First, as we have discussed in Section 1.2.2, there is the big bang singularity in the finite past, $t = 0$. At this time $a = 0$ and the curvature diverges.

Furthermore, the causal horizon at (conformal) time t, i.e., the distance a photon has travelled from $t = \tau = 0$ until time t, is given by $a(t)t = a(t) \int_0^{\tau(t)} a^{-1}\, d\tau$. Since

for $\rho + 3P > 0$, a grows slower than linear in τ, the above integral converges, is finite. As we have seen [Eq. (1.25)], $a(\tau) \propto \tau^{\frac{2}{3(1+w)}}$ if $w = P/\rho$ is constant.

For example, the size of the causal horizon at recombination is seen today under the angle of about 1°, if the Universe was radiation ($w = 1/3$) and matter ($w = 0$) dominated up to recombination, see Ex. 1.8. It is therefore very mysterious that we see the same microwave background temperature on patches separated by much more than 1°, which have never been in causal contact before the microwave photons have been emitted. This is the '**horizon problem**'.

Another problem is the following: the Friedmann equations, (1.138) and (1.139), allow us to derive an evolution equation for $\Omega(t) \equiv 8\pi G\rho a^4/3\dot{a}^2 \equiv 1 + K/\mathcal{H}^2$,

$$\frac{d}{dt}(\Omega(t) - 1) = (\Omega(t) - 1)\frac{8\pi G a^2}{3}\left(\frac{\rho + 3P}{\mathcal{H}}\right). \tag{1.140}$$

This shows, that in an expanding universe with $\rho + 3P > 0$, $\Omega = 1$ is an unstable fixed point of evolution: If $\Omega(t) > 1$, the derivative is positive and $\Omega(t)$ increases while for $\Omega(t) < 1$, the derivative is negative and $\Omega(t)$ decreases. For a present value of $0.1 < \Omega < 2$ we need $|\Omega(\eta_{\rm nuc}) - 1| \sim (z_{\rm eq}/z_{\rm nuc}^2)|\Omega_0 - 1| \leq 10^{-15}$ at nucleosynthesis, or $|\Omega(t_P) - 1| \leq 10^{-60}$ at the Planck time, $\tau_P = \sqrt{\hbar G/c^5} \simeq 5.4 \times 10^{-44}$ s. Why is $\Omega(t)$ still of order unity so long after the only timescale in the problem which is τ_P?

This '**flatness problem**' can also be formulated as an '**entropy problem**'. The entropy inside the curvature radius is already of the order of $S_K \geq 10^{88}$ at the Planck time.

Another problem is the '**monopole problem**' or more generically the problem of unwanted 'relics'. Most particle physics models produce some stable 'relics' at very high temperatures, which are not observed in the present Universe. A very rapid phase of expansion can help to dilute such relics.

To resolve these problems one introduces an 'inflationary phase'. Inflation is a phase during which the strong energy condition, $\rho + 3P > 0$, is violated and expansion can therefore be much more rapid than linear in τ.

1.5.2 Scalar field inflation

We now study the most common solution of the above mentioned problems, namely the introduction of a period where the dynamics of the Universe is dominated by a scalar field, ϕ which is usually called the 'inflaton'. The scalar field Lagrangian is given by

$$\mathcal{L}_\phi = -\frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi - W(\phi)\,. \tag{1.141}$$

The sign of the kinetic term in the above Lagrangian differs from what we are used to from quantum field theory. This comes from the fact that we use the metric signature $(-,+,+,+)$.

The field ϕ can, in principle, interact with other fields such as fermions, gauge bosons, etc., but we assume that this interaction can be neglected during inflation, and that energy and pressure are dominated by the contribution from the inflaton. The energy–momentum tensor of ϕ is given by

$$T_{\mu\nu} = \frac{-2}{\sqrt{-g}} \frac{\partial}{\partial g^{\mu\nu}} (\sqrt{-g}\mathcal{L}_\phi) \,,$$

where $g = \det(g_{\mu\nu})$. This yields

$$\begin{aligned} T_{\mu\nu} &= \partial_\mu\phi\, \partial_\nu\phi - g_{\mu\nu}\mathcal{L}_\phi \\ &= \partial_\mu\phi\, \partial_\nu\phi - \frac{1}{2} g_{\mu\nu}\, \partial_\lambda\phi\, \partial^\lambda\phi - g_{\mu\nu} W(\phi) \,. \end{aligned}$$

Here we have used that the derivatives of the determinant A of an arbitrary matrix A_{ab} with respect to the elements of its inverse, A^{ab}, are given by $\partial A/\partial A^{ab} = A A_{ab}$.

For the energy density and pressure we thus obtain

$$\rho_\phi = -T^0_0 = \frac{1}{2a^2}\dot{\phi}^2 + \frac{1}{2a^2}(\nabla\phi)^2 + W(\phi) \,, \tag{1.142}$$

and

$$P_\phi = \frac{1}{3} T^i_i = \frac{1}{2a^2}\dot{\phi}^2 - \frac{1}{6a^2}(\nabla\phi)^2 - W(\phi) \,. \tag{1.143}$$

We now assume that there exists some region of space within which we may neglect the spatial derivatives of ϕ, at some initial time τ_i, and the temporal derivative is much smaller than the potential,

$$\nabla\phi(\mathbf{x}, \tau_i) \ll \dot{\phi}(\mathbf{x}, \tau_i) \ll W(\phi) \,. \tag{1.144}$$

Furthermore, we assume that the potential is positive,

$$W(\phi(\mathbf{x}, \tau_i)) > 0 \,. \tag{1.145}$$

We then have,

$$\frac{3H^2}{8\pi G} = \rho = \rho_\phi = \frac{1}{2a^2}\dot{\phi}^2 + W(\phi) \simeq W(\phi) \,, \tag{1.146}$$

$$P = P_\phi = \frac{1}{2a^2}\dot{\phi}^2 - W(\phi) \simeq -W(\phi) \,. \tag{1.147}$$

so that $P_\phi \simeq -\rho_\phi$ and $\rho_\phi + 3P_\phi \simeq -2W(\phi) < 0$. (We have neglected a possible curvature term. Qualitatively nothing changes if we include it, since it soon becomes subdominant.)

This is the basic idea of inflation: at some early time, in some sufficiently large patch, the Universe is dominated by the potential of a slowly varying (slow rolling) scalar field, and hence it is in an inflationary phase. During inflation this patch expands rapidly, the causal horizon becomes very large and $\Omega(t)$ tends to 1, so that the curvature term is soon negligible. As time goes on, the scalar field starts evolving faster and inflation eventually comes to an end when the time derivative ϕ'^2 grows to the order of W. The scalar field then soon reaches the minimum of the potential and starts to oscillate. We suppose that at large values of $a^{-1}\dot{\phi}$, the coupling of the inflaton to other fields becomes significant so that it decays into a thermal mix of elementary particles, leading to a radiation dominated universe. There are many detailed realizations of this basic picture which can be found in the literature, see e.g., Liddle & Lyth (2000). It is, however very difficult to deduce them from a serious high-energy physics theory such as string theory.

Let us study slow roll inflation in somewhat more detail. When neglecting spatial derivatives, the equation of motion of the scalar field becomes ($W_{,\phi} \equiv dW/d\phi$)

$$\phi'' + 3\left(\frac{a'}{a}\right)\phi' + W_{,\phi} = 0\,. \tag{1.148}$$

During slow rolling, the first term of this equation is negligible with respect to the two others, so that

$$3\left(\frac{a'}{a}\right)\phi' \simeq -W_{,\phi}\,. \tag{1.149}$$

The slow roll conditions are therefore

$$\frac{1}{2}\phi'^2 \ll W \quad \text{and} \quad |\phi''| \ll 3H|\phi'|\,. \tag{1.150}$$

With $H = a'/a$, slow rolling also implies that $H' \ll H^2$. Taking the time derivative of Eq. (1.146) and replacing ϕ' by (1.149), this yields the slow roll conditions

$$\epsilon_1 \equiv -\frac{H'}{H^2} = \frac{\mathcal{H}^2 - \dot{\mathcal{H}}}{\mathcal{H}^2} = \frac{m_P^2}{16\pi}\left(\frac{W_{,\phi}}{W}\right)^2 \simeq \frac{3}{2}\frac{\phi'^2}{W} \ll 1\,. \tag{1.151}$$

The second condition of Eq. (1.150) gives

$$\left|\frac{\phi''}{3H\phi'}\right| \ll 1\,.$$

We now set

$$\epsilon_2 \equiv -\frac{m_P^2}{24\pi}\left(\frac{W_{,\phi\phi}}{W}\right) \qquad \text{and require} \qquad |\epsilon_2| \ll 1\,. \tag{1.152}$$

Note that ϵ_1 is always positive while ϵ_2 can have either sign. With $H^2 \simeq 8\pi/(3m_P^2)W$, and the derivative of $\phi' = -W_{,\phi}/(3H)$, one finds that the inequalities (1.151) and (1.152) are equivalent to the slow roll conditions (1.150). The parameters ϵ_1 and ϵ_2 are the slow roll parameters. Inflation terminates when ϵ_1 approaches unity.

Taking the derivative (w.r.t. t) of Eq. (1.151) in the last equals sign, one obtains,

$$\dot{\epsilon}_1 = 2\epsilon_1(3\epsilon_2 + 2\epsilon_1)\mathcal{H}\,, \qquad \epsilon_2 = \frac{1}{6}\left(\frac{\dot{\epsilon}_1}{\epsilon_1}\mathcal{H}^{-1} - \epsilon_1\right)\,. \tag{1.153}$$

The last equation can also be used as a definition of ϵ_2. The advantage of this definition is its independence of the realization of slow roll inflation by means of a scalar field. A more systematic procedure is to define $\tilde{\epsilon}_1 \equiv \epsilon_1$ and $\tilde{\epsilon}_2 = (\dot{\tilde{\epsilon}}_1/\tilde{\epsilon}_1)\mathcal{H}^{-1}$, $\tilde{\epsilon}_3 = (\dot{\tilde{\epsilon}}_2/\tilde{\epsilon}_2)\mathcal{H}^{-1}$ and so forth. Our parameter ϵ_2 is related to $\tilde{\epsilon}_2$ via

$$\epsilon_2 = -\frac{1}{3}\epsilon_1 + \frac{1}{6}\tilde{\epsilon}_2\,. \tag{1.154}$$

While ϵ_2 is usually of the same order of magnitude as ϵ_1, we expect $\tilde{\epsilon}_2$ to be significantly smaller.

As an example we consider power law expansion, $a \propto t^q$. In this case we have

$$\mathcal{H} = \frac{q}{t}, \qquad \epsilon_1 = 1 + \frac{1}{q}\,, \qquad \epsilon_2 = -\frac{1}{3}\epsilon_1\,, \qquad \tilde{\epsilon}_2 = \tilde{\epsilon}_n = 0\,. \tag{1.155}$$

During slow roll inflation, $q \sim -1$, the parameters ϵ_1 and ϵ_2 are small. Also note that $\epsilon_2 = -(1/3)\epsilon_1$ during power law expansion. The parameters $\tilde{\epsilon}_i, i > 1$ describe the deviation from power law expansion. They have been used in the literature to derive a systematic slow roll expansion to higher orders (Schwarz *et al.*, 2001). In this book we shall not go beyond the first order and we use the standard parameters ϵ_1 and ϵ_2 to make contact with the standard literature.

There are two principally different possibilities for slow roll inflation.

(i) We first consider a potential which is simply $\propto \phi^n$, so that $W_{,\phi\phi}/W \sim (W_{,\phi}/W)^2 \sim \phi^{-2}$. The slow roll conditions then require $\phi \gg m_P$ and inflation stops when the inflaton becomes of order the Planck mass. These models are termed **large-field inflation**. Setting $W = (\lambda/n)m_P^4(\phi/m_P)^n$, during the inflationary phase Eq. (1.149) together with Eq. (1.146) implies

$$\sqrt{\frac{24\pi\lambda}{n}}m_P\,(\phi/m_P)^{n/2}\,\phi' = -\lambda m_P^3\left(\frac{\phi}{m_P}\right)^{n-1}\,. \tag{1.156}$$

Dividing by ϕ^{n-1}, if $n \neq 4$ the left-hand side becomes the derivative of $(\phi/m_P)^{2-n/2}$, which hence is a constant. If $n = 4$, the left-hand side is $\propto 1/\phi$, i.e., the derivative of $\log(\phi/m_P)$. The general solution is therefore given by

$$\phi(\tau)^{(4-n)/2} = \phi_i^{(4-n)/2} + \frac{4-n}{2}\sqrt{n\lambda/24\pi}\, m_P(\tau - \tau_i) \quad \text{if } n \neq 4\,, \tag{1.157}$$

$$\phi(\tau) = \phi_i \exp\left(-\sqrt{\frac{\lambda}{6\pi}} m_P(\tau - \tau_i)\right) \qquad \text{if } n = 4\,. \tag{1.158}$$

Inserting now $\phi' = -\sqrt{\lambda n}/24\pi m_P^2(\phi/m_P)^{n/2-1}$ in the Friedmann equation,

$$(\log(a))' = \sqrt{\frac{8\pi\lambda}{3n}} m_P(\phi/m_P)^{n/2}\,, \qquad \text{we obtain}$$

$$\frac{d\log(a)}{d\phi} = -\frac{8\pi}{n}\frac{\phi}{m_P^2}\,,$$

with solution

$$a(\tau) = a_i \exp\left(\frac{4\pi}{nm_P^2}(\phi_i^2 - \phi^2)\right)\,. \tag{1.159}$$

This case is illustrated in Fig. 1.11.

(ii) If the potential is more complicated and has a very flat regime in the vicinity of its maximum $\phi = \sigma \ll m_P$, like, e.g., the Coleman–Weinberg potential (Kolb & Turner, 1990),

$$W(\phi) = \frac{1}{2}\sigma^4 + \phi^4\left[\ln\left(\frac{\phi^2}{\sigma^2}\right) - \frac{1}{2}\right]\,,$$

we speak of **small-field inflation**. This potential passes through 0 at $\phi = \sigma$. In this case, the slow roll conditions are satisfied for field values $|\phi| \lesssim \sigma$, which are much smaller than the Planck mass.

During a potential dominated phase where $\rho \sim -P \sim W \sim$ constant, the solutions of the Friedmann equations are

$$a = a_0 \exp(\tau H) = \frac{1}{H|t|} \quad (-\infty < t < 0, \quad -\infty < \tau < \infty)\,, \tag{1.160}$$

$$H^2 = \frac{8\pi G}{3} W = \text{constant}\,, \tag{1.161}$$

$$\mathcal{H} = aH = \frac{1}{|t|}\,. \tag{1.162}$$

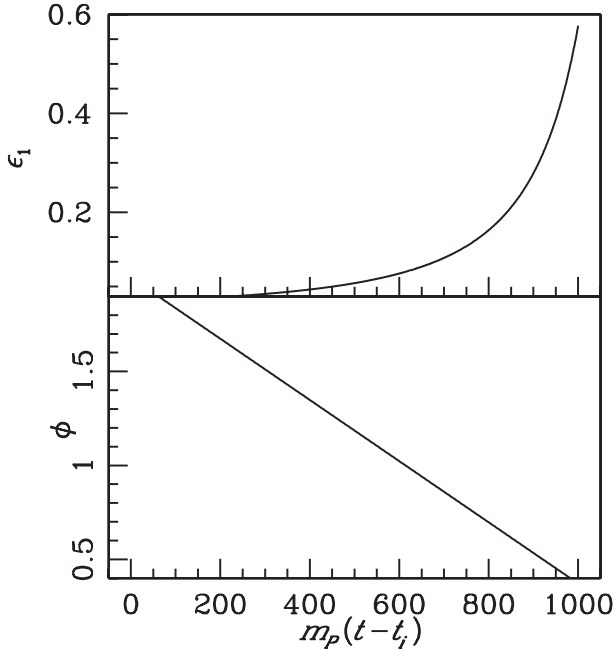


Fig. 1.11. Large-field inflation for $W = \lambda m_P^2 \phi^2/2$. The bottom panel shows the inflaton ϕ in units of m_P rolling linearly in time. In the upper panel the evolution of the slow roll parameter, $\epsilon_1(t)$ is indicated. As long as $\phi > m_P$, $\epsilon_1 = -2\epsilon_2$ stays small. At $\phi \sim m_P$, ϵ_1 starts to grow and inflation stops.

The limit $\tau \to \infty$ corresponds to $t \to 0$. The above solution is a portion of de Sitter spacetime.[5]

Denoting by indices i and f the beginning and the end of inflation, the number of e-foldings of expansion during inflation is given by

$$N(\phi_f, \phi_i) = \ln\left(\frac{a(\tau_f)}{a(\tau_i)}\right).$$

Using

$$N(\phi_f, \phi_i) = \ln a_f - \ln a_i = \int_{a_i}^{a_f} \frac{da}{a},$$

we obtain

$$N(\phi_f, \phi_i) = \int_{a_i}^{a_f} \frac{1}{a}\, da = \int_{\tau_i}^{\tau_f} \frac{a'}{a}\, d\tau = \int_{\tau_i}^{\tau_f} H\, d\tau. \tag{1.163}$$

[5] de Sitter spacetime is the solution to the Einstein equation $G_{\mu\nu} = \Lambda g_{\mu\nu}$ with $\Lambda > 0$. The solution with $\Lambda < 0$ is called anti-de Sitter, see Hawking & Ellis (1973).

With Eq. (1.149) we can write

$$H\,d\tau = H\,\frac{d\tau}{d\phi}\,d\phi = H\frac{d\phi}{\phi'} = -\frac{3H^2\,d\phi}{W_{,\phi}}\,.$$

The number of e-foldings is hence given by

$$N(\phi_f,\phi_i) = -3\int_{\phi_i}^{\phi_f}\frac{H^2}{W_{,\phi}}\,d\phi \simeq -\frac{8\pi}{m_P^2}\int_{\phi_i}^{\phi_f}\frac{W}{W_{,\phi}}\,d\phi = -2\sqrt{\pi}\int_{\phi_i}^{\phi_f}\frac{1}{\sqrt{\epsilon_1}}\frac{d\phi}{m_P}$$
$$\sim \frac{8\pi}{n}\frac{\phi_i^2}{m_P^2}\,. \tag{1.164}$$

The last $\sim$ sign is valid only for large-field inflation, where $W \propto \phi^n$ and we suppose

$$\phi_f \sim m_P \ll \phi_i\ .$$

The slow roll conditions imply

$$N_{\rm tot} = N(\phi_f,\phi_i) \gg 1. \tag{1.165}$$

For $w = P/\rho =$ constant we have

$$|\Omega(\tau) - 1| = \frac{3|K|}{8\pi G a^2\rho} \propto a^{1+3w}\,.$$

During an inflationary phase, $w = -1$, $|\Omega(\tau) - 1|$ decreases like $1/a^2$. To reduce it from a value of order unity down to $\sim 10^{-60}$ we therefore need about $30\ln(10) \sim 70$ e-foldings of inflation.

1.5.3 Pre-heating and reheating

When inflation ends, ϕ decays rapidly and starts oscillating about its minimum. The details of this process depend on the couplings of the inflaton to other degrees of freedom, which eventually decay into the degrees of freedom of the standard model. For this discussion we consider a simple toy model with $\mathcal{L}_\phi = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m_\phi^2\phi^2$. At the end of inflation the inflaton oscillates as

$$\phi = \phi_0(\tau)\cos(m_\phi\tau)$$

with a slowly varying amplitude $\phi_0(\tau) \simeq m_P$. The inflatons have vanishing momentum and their number density is

$$n_\phi = \frac{\rho_\phi}{m_\phi} = \frac{1}{2m_\phi}((\phi')^2 + m_\phi^2\phi^2) \simeq m_\phi m_P^2\,. \tag{1.166}$$

For example for $m_\phi = 10^{15}$ GeV this amounts to the huge number density of $n_\phi \sim 10^{95}\ \mathrm{cm}^{-3}$.

Independent of the detailed form of the potential, to lowest order, ϕ is a harmonic oscillator with frequency $m_\phi = \sqrt{W_{,\phi\phi}(\sigma)}$ (as long as the quadratic term in the potential does not vanish). For a harmonic oscillator, when averaging over one period we have

$$\langle W \rangle = \frac{1}{2a^2} \langle \dot{\phi}^2 \rangle \,,$$

so that

$$\langle p_\phi \rangle = \left\langle \frac{1}{2a^2} \dot{\phi}^2 - W \right\rangle = 0, \quad \text{and hence } \langle \rho_\phi \rangle \propto a^{-3} \,.$$

We assume that during these oscillations, the coupling of ϕ to other degrees of freedom becomes relevant and the inflaton finally decays into a mix of elementary particles. In a first approximation we can describe the coupling with the other degrees of freedom by means of a term of dissipation of the form $\Gamma\dot{\phi}$ in the equation of motion for ϕ,

$$\phi'' + 3H\phi' + \Gamma\phi' = -W_{,\phi}(\phi) \,. \tag{1.167}$$

As long as $H \gg \Gamma$ (during inflation), particle production is negligible. When $H \simeq \Gamma$, reheating takes place and the inflaton energy is rapidly dissipated into other particles which couple to the inflaton.

In order to discuss the decay of the inflaton in somewhat more detail, we consider a toy model where the interaction is dominated by the coupling of ϕ to a scalar field χ with Lagrangian

$$\mathcal{L}_\chi = -\frac{1}{2}\, \partial_\mu \chi \, \partial^\mu \chi - \frac{1}{2} m_\chi^2 \chi^2 \,. \tag{1.168}$$

The interaction between the inflaton ϕ and the matter field χ is supposed to be of the form

$$\mathcal{L}_{\rm int} = -\frac{1}{2} g \phi \chi^2 \,, \tag{1.169}$$

where g is a coupling constant with the dimension of mass. The full Lagrangian is then given by

$$\mathcal{L} = \mathcal{L}_\phi + \mathcal{L}_{\rm int} + \mathcal{L}_\chi \,. \tag{1.170}$$

The decay rate of the ϕ particles in Born approximation is

$$\Gamma_\phi \sim \frac{g^2}{m_\phi} \,.$$

However, inserting this into Eq. (1.167) is only a good approximation, when the mean number of χ particles already present in a given momentum mode k is small

so that we may neglect stimulated emission. The effective mass of χ-particles is $\sqrt{m_\chi^2 + g\phi(t)}$ so that their momentum is

$$k = \left(\frac{m_\phi^2}{4} - m_\chi^2 - g\phi(t) \right)^{1/2} .$$

Here we have taken into account that each inflaton decays into two χ-particles. Now, $\phi(t) \in [-m_P, m_P]$. Hence, if $m_\phi^2 \gg m_\chi^2 + g m_P$, the band of possible momenta is given by $k \in [k_0 - \Delta k, k_0 + \Delta k]$ with

$$k_0 = \sqrt{\frac{m_\phi^2}{4} - m_\chi^2} \simeq \frac{m_\phi}{2} \quad \text{and} \quad \Delta k \simeq \frac{g m_P}{m_\phi} \ll k_0 .$$

Because $\Delta k \ll k_0$ this situation is called 'narrow band pre-heating'. As we shall see below, this process leads to resonant amplification.

The number of χ-particles with momentum k is roughly given by the total number of χ-particles divided by the number of 'elemetary phase space volumes', $(2\pi)^3$, in the allowed volume of phase space, $4\pi k_0^2 (2\Delta k)$. This yields

$$N_k \simeq \frac{4\pi^2 n_\chi}{g m_\phi m_P} \simeq \frac{2\pi^2 m_P n_\chi}{g n_\phi} .$$

For the second $\simeq$ sign we made use of Eq. (1.166). This occupation number exceeds unity as soon as a fraction g/m_P of ϕ-particles is converted into χ-particles. After that moment, stimulated emission can no longer be neglected.

To calculate the evolution of χ-particles in more detail we vary the Lagrangian with respect to χ to obtain the χ-equation of motion,

$$\chi'' + 3H\chi' - a^{-2}\nabla^2\chi + (m_\chi^2 + g\phi_0 \cos(m_\phi \tau))\chi = 0 .$$

To study qualitatively the decay of the ϕ-particles into χ, we neglect expansion by setting $H = 0$, $a = 1$ and $\phi_0 =$ constant. Fourier transforming the above equation, we then obtain for the mode χ_k

$$\chi_k'' + \left[\omega_k^2 + 2\mu \cos(m_\phi \tau)\right]\chi_k = 0, \quad \mu = \frac{g\phi_0}{2} , \quad \omega_k^2 = k^2 + m_\chi^2 .$$

This equation is known as the *Mathieu equation*. Its solutions are characterized by resonance bands of widths $\Delta\omega_k^{(n)}$ centred at the frequencies

$$\omega_k^{(n)} = \frac{n}{2} m_\phi .$$

The widths are of the order of

$$\frac{\Delta\omega_k^{(n)}}{\omega_k^{(n)}} \simeq \left(\frac{2\mu}{\omega_k^{(n)2}}\right)^n = \frac{\Delta k^{(n)}}{k^{(n)}} \propto g^n \,.$$

For frequencies within these bands, χ_k is amplified exponentially fast (for more details see Arnol'd (1978)). Since the width of the nth resonance is proportional to g^n, it appears only in nth-order perturbation theory. For small couplings, g only, the first resonance $\omega_k^{(1)} = m_\phi/2$ with $\Delta\omega_k^{(1)} = \Delta k$ is relevant. When we take into account the expansion of the Universe, the frequency ω_k becomes time dependent. A given frequency therefore spends only a finite time in the resonance band and the energy transfer from ϕ into χ remains perfectly finite. Nevertheless, this parametric resonance is much more efficient than the decay obtained by some effective damping rate Γ.

After parametric resonance, χ is not yet in a thermal state. This period is therefore called '*pre-heating*'. After pre-heating, the coupling of χ to other degrees of freedom leads to thermalization; this process is called *reheating*. The importance of pre-heating lies in its efficiency in transferring energy. If the χ-field couples strongly to the standard-model particles, reheating and thermalization can proceed much faster over resonant decay than over the necessarily weak average coupling of the inflaton to other particles.

If the condition $m_\phi^2 > m_\chi^2 + g\phi(t)$ is not satisfied, $\Delta\omega_k^{(1)} = \Delta k$ is not small and we have 'broad-band' resonance. In this case, the mass of the χ-particles can be larger than the mass of the ϕ-particles and only the coherent decay of several inflatons can lead to χ-production. For a discussion of the main physical processes in this case see Mukhanov (2005). One of the most interesting consequences of broad-band resonance is that it can lead to the production of particles that are heavier than the inflaton.

The temperature at the end of reheating depends on the details of the model. It can go from 1 TeV $< T <$ 10^{13} GeV.

1.5.4 Resolution of the 'cosmological problems'

At the end of the reheating process, $\tau = \tau_{\rm rh}$, all the energy is supposed to be thermalized and the Universe is dominated by relativistic particles, satisfying $P = \rho/3$ such that

$$\rho \propto a^{-4} \,.$$

To determine the duration of inflation necessary in order to solve the horizon problem, we consider the entropy, S_H, contained in a volume which corresponds to

one Hubble scale, H_i^{-1}, at the beginning of inflation. Since expansion is adiabatic after inflation, the entropy inside a given physical volume remains constant. The requirement that the present Hubble scale, H_0^{-1}, be smaller than the size of the causal horizon is therefore equivalent to $S_H > S_{H_0}$, where S_{H_0} denotes the entropy inside the volume H_0^{-3}. The entropy inside a causal volume, $H_i^{-3}(a/a_i)^3$ is given by its value

$$S_H \simeq H_i^{-3} \left(\frac{a_{\rm rh}}{a_i}\right)^3 T_{\rm rh}^3 ,$$

after reheating. The Hubble parameter at the beginning of inflation is

$$H_i^2 \simeq \frac{8\pi}{3m_P^2} W(\phi_i) ,$$

so that

$$S_H \simeq H_i^{-3} \left(\frac{a_f}{a_i}\right)^3 \left(\frac{a_{\rm rh}}{a_f}\right)^3 T_{\rm rh}^3 \simeq \frac{m_P^3}{W_i^{3/2}} e^{3N_{\rm tot}} \frac{\rho_f}{T_{\rm rh}} .$$

For the last $\simeq$ sign we have assumed that the Universe was roughly matter dominated from the end of inflation until the end of reheating, $\rho \propto a^{-3}$ and $\rho_{\rm rh} \sim T_{\rm rh}^4$. With $\rho_f \sim W_f$, this yields

$$S_H \simeq \frac{m_P^3 W_f}{T_{\rm rh} W_i^{3/2}} e^{3N_{\rm tot}} .$$

In order to solve the entropy problem, we require that this entropy is at least as large at the entropy in the present Hubble horizon, $S_H > S_{H_0} \simeq T_0^3 H_0^{-3} \simeq 10^{88}$. This now results in

$$N_{\rm tot} \geq N_{\rm min} = \frac{88}{3} \ln(10) + \ln\left(\frac{T_{\rm rh}^{1/3} W_i^{1/2}}{m_P W_f^{1/3}}\right) . \tag{1.171}$$

For example, in a model with $W = \frac{1}{2} m_\phi^2 \phi^2$, we have large-field inflation which stops roughly when $\phi = \phi_f \simeq m_P$ so that $W_f = \frac{1}{2}(m_\phi m_P)^2$ and $W_i = \frac{1}{2}(m_\phi \phi_i)^2$. Hence

$$N_{\rm min} = \frac{88}{3} \ln(10) + \frac{1}{3} \ln\left(\frac{T_{\rm rh} m_\phi}{m_P^2}\right) + \ln\left(\frac{\phi_i}{m_P}\right) .$$

If $N_{\rm tot} \geq N_{\rm min}$ the horizon problem is also solved. Indeed, since the entropy inside a comoving volume is conserved after inflation, the present volume of radius H_0^{-1} has grown out of a radius which was smaller than H_i^{-1} at the beginning of inflation, and therefore was already in causal contact before the beginning of inflation.

To solve the flatness problem we must enlarge the curvature scale to $R_K(\tau_0) \geq H_0^{-1}$. This is equivalent to $S_K(\tau_0) \geq S_H(\tau_0) \simeq 10^{88}$. With

$$R_K^3(\tau_{\rm rh}) = R_K^3(\tau_i)\left(\frac{a_f}{a_i}\right)^3 = \frac{H_i^{-3}}{|\Omega_i - 1|^{3/2}}\left(\frac{a_f}{a_i}\right)^3 ,$$

this leads to

$$N_{\rm tot} \geq N_{\rm min} + \frac{1}{2}\log|\Omega_i - 1| . \tag{1.172}$$

Comparing $N_{\rm min}$ with Eq. (1.164), we find that successful inflation with a simple $\frac{1}{2}m_\phi^2\phi^2$ potential requires $\phi_i \gtrsim$ a few times m_P. After an inflationary period which is sufficiently long, so that the conditions (1.171) and (1.172) are satisfied, both, the horizon and flatness problems are resolved. During such an inflationary phase also all unwanted relics are diluted by a factor of $\exp(3N_{\rm tot})$.

Finally, it is important to note that we do not require a perfectly homogeneous and isotropic universe, or even thermal equilibrium prior to inflation. We just need a small 'patch' in an otherwise arbitrary, chaotic, universe, within which the gradient and kinetic energy are much smaller than the potential energy, so that the slow roll conditions are satisfied. This patch then inflates to encompass the entire present Hubble volume. This idea of 'chaotic inflation' goes back to Linde (1989) and it is of course much more satisfactory than a model where the Universe has to start out with homogeneous and isotropic spatial sections before inflation.

When discussing inflation, one of the most mysterious problems of gravity becomes apparent: while adding a constant to the potential W of the scalar field does not affect any of the other interactions, it severely alters gravity. It modifies cosmic expansion in the same way as adding a cosmological constant. What determines the correct level of a potential? This question is equivalent to the problem of the cosmological constant. Why is the present cosmological constant so small, $\Lambda/(8\pi G) \simeq (2 \times 10^{-3}\,{\rm eV})^4$, much smaller than any fundamental energy scale? The problem is even more serious when we remember that in quantum field theory we use the freedom to add or subtract a constant from the potential by absorbing the infinite zero-point energy into it. Furthermore, at each phase transition this zero-point energy changes by a finite, calculable amount. Before the discovery of the accelerated expansion of the Universe which is most simply interpreted as a cosmological constant, $\Lambda/(8\pi G) \simeq (2 \times 10^{-3}\,{\rm eV})^4 \neq 0$, it was justifiable to assume that the freedom of the cosmological constant has to be used in order to annihilate any vacuum energy contribution from quantum field theory, so that the effective cosmological constant would vanish, $\Lambda_{\rm eff} = \Lambda + 8\pi G W_0 = 0$. Present observations, however, indicate that this compensation takes place only approximately, leaving a small but non-vanishing effective cosmological constant, $\Lambda_{\rm eff} \neq 0$, which starts

to dominate the expansion of the Universe just at present time, when there are sufficiently developed intelligent beings in the Universe which wonder about it. In all the cosmic past, this cosmological constant was completely negligible, and in all the cosmic future, it will be the only relevant contribution to expansion. Only at present it is comparable with the mean mass density of the Universe. Apart from the bizarre value of $\Lambda_{\rm eff}$, we thus also have a strange coincidence problem.

This is presently one of the deepest problems of physics. Ordinary quantum field theory does not determine the vacuum energy of quantum fields, but only changes which may happen depending on the external conditions. We may hope that a quantum theory of gravity addresses the cosmological constant problem. The cosmological constant may even represent our first observational data related to quantum gravity.

Exercises

(The exercises marked with an asterisk are solved in Appendix A10.1.)

Ex. 1.1 Coordinates
Find the coordinate transformation leading from the coordinates used in Eq. (1.9) to those of Eq. (1.10) and finally of Eq. (1.8).

Ex. 1.2 FL universes are conformally flat
Show that FL universes are conformally flat (also when the curvature does not vanish) and find the coordinate transformation $(\tau, r) \to (\sigma, \rho)$ such that

$$-d\tau^2 + a^2(\sigma)\gamma_{ij}\, dx^i\, dx^j = A^2(\sigma, \rho)\eta_{\mu\nu}\, dX^\mu\, dX^\nu\,, \tag{1.173}$$

with $\sigma = X^0$ and $\rho^2 = \sum_{i=1}^3 (X^i)^2$.

Ex. 1.3 Matter and radiation mixture
Consider a FL universe containing a mixture of non-relativistic matter (dust) and radiation with vanishing curvature. The respective densities and pressures are ρ_m, ρ_r and $P_m = 0$, $P_r = \rho_r/3$. We denote the ratio of radiation to matter by $R = \rho_r/\rho_m$.
(a) Determine w and c_s^2 as functions of R. What is the time dependence of R?
(b) For a given redshift $z_{\rm eq} \gg 1$ of matter and radiation equality determine the scale factor as a function of conformal and of physical time; normalize the scale factor to 1 at equality, $a_{\rm eq} = 1$.
(c) Determine $t_{\rm eq}$ and $\tau_{\rm eq}$ as functions of $z_{\rm eq}$, and H_0.

Ex. 1.4 Cosmological constant*
Investigate the dynamics of a FL universe with matter ($P = 0$) and a cosmological constant Λ.
(i) Show for a sufficiently small cosmological constant and positive curvature that the Universe recollapses in a 'big crunch', while for a larger cosmological constant or non-positive curvature, the Universe expands forever.

(ii) Show furthermore that for an even higher cosmological constant there are solutions which have no big bang in the past, but issue from a previous contracting phase. The transition from the contracting to an expanding phase is called the 'bounce'.

(iii) Make a plot in the plane $(\Omega_m, \Omega_\Lambda)$ distinguishing the regimes determined above.

(iv) For case (ii), determine (numerically) the redshift of the bounce as a function of Ω_Λ for fixed $\Omega_m = 0.1$. Discuss.

Ex. 1.5 Distribution functions

Show that in the non-relativistic limit, $m \gg T$ both, the Fermi–Dirac and the Bose–Einstein distributions reduce to a Maxwell–Boltzmann distribution and the number and energy density are given by

$$n = \frac{2}{(2\pi)^3} \exp(-(m-\mu)/T)(2\pi m T)^{3/2}\,, \quad \rho = mn\,, \tag{1.174}$$

where μ is the chemical potential.

Ex. 1.6 Liouville equation

Using that, in a FL universe the distribution function f only depends on (conformal) time t and $p = \sqrt{\gamma_{ij} p^i p^j}$, derive equation (1.87).

Ex. 1.7 The neutrino background

Determine the neutrino cross section for the reaction $e^- + \bar{\nu} \to e^- + \bar{\nu}$ at energy $E_\nu = T_\nu(t_0)$. Compare it with the cross section of the neutrinos detected in the super-Kamiokande experiment. Keeping the efficiency of super-Kamiokande, how large a water tank would you need to detect neutrinos from the cosmic background?

Ex. 1.8 Angular diameter distance

Determine the angular diameter distance to the last scattering surface under the assumptions $K = \Lambda = 0$. Under which angle do we presently see the causal horizon of this time, $a(t_{\rm rec})t_{\rm rec}$? How does this result change if one admits a cosmological constant so that $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$?

2

Perturbation theory

2.1 Introduction

In this chapter we develop in detail the theory of linear perturbations of a Friedmann–Lemaître universe. This theory is of utmost importance, since we assume that the observed structure in the Universe (galaxies, cluster voids, etc.) have grown out of small initial fluctuations. Their entire evolution from the generation of the fluctuations until the time when they become of order unity can be studied within linear perturbation theory. This is especially relevant for the fluctuations in the CMB which are still very small today. It is also one of the main reasons why CMB anisotropies are so important for observational cosmology: they can be calculated to very good accuracy within linear perturbation theory, which is simple and lends itself to highly accurate and fast computations.

The idea that the large-scale structure of our Universe might have grown out of small initial fluctuations via gravitational instability goes back to Newton (letter to Bentley, 1692 (Newton, 1958)). The first relativistic treatment of linear perturbations in a Friedmann–Lemaître universe was given by Lifshitz (1946). There he found that the gravitational potential cannot grow within linear perturbation theory and he concluded that galaxies have not been formed by gravitational instability.

Today we know that in order to form structures it is sufficient that matter density fluctuations can grow. Nevertheless, considerable initial fluctuations with amplitudes of the order of 10^{-5} are needed in order to reproduce the cosmic structures observed today. These are much larger than typical statistical fluctuations on scales of galaxies and we have to suggest a mechanism to generate them. Furthermore, the measurements of anisotropies in the cosmic microwave background show that the amplitude of fluctuations in the gravitational potential is constant over a wide range of scales, i.e., the fluctuation spectrum is scale independent.

As we shall see in Chapter 3, inflation generically produces such a spectrum of nearly scale-invariant fluctuations.

We begin this chapter by defining gauge-invariant perturbation variables. Then we present the basic perturbation equations. As examples for the matter equations we shall consider perfect fluids and scalar fields. Then we discuss light-like geodesics, which present a good approximation for CMB anisotropies on sufficiently large scales and are important for discussing the effect of lensing. The final section is devoted to the definition of the power spectrum and an elementary discussion of statistical issues. Due to their complexity and importance for the goal of this book, we devote special chapters to the perturbed Boltzmann equation for CMB anisotropies, Chapter 4, and for polarization Chapter 5.

2.2 Gauge-invariant perturbation variables

The observed Universe is not perfectly homogeneous and isotropic. Matter is arranged in galaxies and clusters of galaxies and there are large voids in the distribution of galaxies. Let us assume, however, that these inhomogeneities grew out of small variations of the geometry and of the energy–momentum tensor which we shall treat in first-order perturbation theory. For this we define the perturbed geometry by

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \varepsilon a^2 h_{\mu\nu} \,, \tag{2.1}$$

$\bar{g}_{\mu\nu}$ being the unperturbed Friedmann metric defined in the previous chapter. We conventionally set (absorbing the 'smallness' parameter ε into $h_{\mu\nu}$)

$$\begin{aligned} &g_{\mu\nu} = \bar{g}_{\mu\nu} + a^2 h_{\mu\nu} \,, \quad && \bar{g}_{00} = -a^2 \,, \quad && \bar{g}_{ij} = a^2 \gamma_{ij} \,, \quad && |h_{\mu\nu}| \ll 1 \,, \\ &T^{\mu}_{\ \nu} = \overline{T}^{\mu}_{\ \nu} + \theta^{\mu}_{\ \nu} \,, && \overline{T}^{0}_{\ 0} = -\bar{\rho} \,, && \overline{T}^{i}_{\ j} = \bar{P}\delta^{i}_{\ j} \,, && |\theta^{\mu}_{\ \nu}|/\bar{\rho} \ll 1 \,. \end{aligned} \tag{2.2}$$

2.2.1 Gauge transformation, gauge invariance

The first fundamental problem we want to discuss is the choice of gauge in cosmological perturbation theory.

For linear perturbation theory to apply, the spacetime manifold $\mathcal{M}$ with metric g and the energy–momentum tensor T of the real, observable Universe must be in some sense close to a FL universe, i.e., the manifold $\mathcal{M}$ with a Robertson–Walker metric $\bar{g}$ and a homogeneous and isotropic energy–momentum tensor $\overline{T}$. It is an interesting, non-trivial unsolved problem how to construct 'the best' $\bar{g}$ and $\overline{T}$ from the physical fields g and T in practice. There are two main difficulties: first, spatial averaging procedures depend on the choice of a hypersurface of constant time and they do not commute with derivatives, so that averaged fields $\bar{g}$ and $\overline{T}$ will, in general, not satisfy Einstein's equations. Second, averaging is in practice impossible over super-horizon scales.

Even though we cannot give a constructive prescription of how to define the nearly homogeneous and isotropic spatial slices from the physical spacetime, or the spatially averaged metric and energy–momentum tensor, we now assume that there exists an averaging procedure which leads to a FL universe with spatially averaged tensor fields $\overline{S}$, such that the deviations are small,

$$\frac{|T_{\mu\nu} - \overline{T}_{\mu\nu}|}{\max_{\{\alpha\beta\}}\{|\overline{T}_{\alpha\beta}|\}} \ll 1 \quad \text{and} \quad \frac{|g_{\mu\nu} - \overline{g}_{\mu\nu}|}{\max_{\{\alpha\beta\}}\{\overline{g}_{\alpha\beta}\}} \ll 1,$$

and where $\bar{g}$ and $\overline{T}$ satisfy Friedmann's equations. The latter condition can be achieved, e.g., by defining $\overline{T}$ via the Friedmann equations. Let us call such an averaging procedure 'admissible'. There may be many different admissible averaging procedures (e.g., over different hypersurfaces) leading to slightly different FL backgrounds. But since $|g - \bar{g}|$ is small of order ε, the difference of the two FL backgrounds must also be small of order ε and we can interpret it as part of the perturbation.

We now consider a fixed admissible FL background $(\bar{g}, \overline{T})$ as chosen. Since the theory is invariant under diffeomorphisms (coordinate transformations), the perturbations are not unique. For an arbitrary diffeomorphism ϕ and its push forward ϕ_*, the two metrics g and $\phi_*(g)$ describe the same geometry. Since we have chosen the background metric $\bar{g}$ we only allow diffeomorphisms which leave $\bar{g}$ invariant i.e., which deviate only in first order from the identity. Such an 'infinitesimal' diffeomorphism can be represented as the infinitesimal flow of a vector field X, $\phi = \phi_\varepsilon^X$. Remember the definition of the flow: for the integral curve, $\gamma_x(s)$, of X with starting point x, i.e., $\gamma_x(s=0) = x$ we have $\phi_s^X(x) = \gamma_x(s)$. In terms of the vector field X, to first order in ε, its push forward is then of the form

$$\phi_* = id + \varepsilon L_X + \mathcal{O}(\varepsilon^2)\,,$$

where L_X denotes the Lie derivative in direction X (see Appendix A2.2). The transformation $g \to \phi_*(g)$ is equivalent to $\bar{g} + \varepsilon a^2 h \to \bar{g} + \varepsilon(a^2 h + L_X \bar{g}) + \mathcal{O}(\varepsilon^2)$. Under an 'infinitesimal coordinate transformation' the metric perturbation h therefore transforms as

$$h \to h + a^{-2} L_X \bar{g}\,. \tag{2.3}$$

In the context of cosmological perturbation theory, infinitesimal coordinate transformations are called 'gauge transformations'. The perturbation of an arbitrary tensor field $S = \bar{S} + \varepsilon S^{(1)}$ obeys the gauge transformation law

$$S^{(1)} \to S^{(1)} + L_X \bar{S}\,. \tag{2.4}$$

Since every vector field X generates a gauge transformation $\phi = \phi_\varepsilon^X$, we can conclude that only perturbations of tensor fields with $L_X \overline{S} = 0$ for all vector fields X, i.e., with vanishing (or constant) 'background contribution' are gauge invariant. This result is called the 'Stewart–Walker Lemma' (Stewart & Walker, 1974).

The gauge dependence of perturbations has caused many controversies in the literature, since it is often difficult to extract the physical meaning of gauge-dependent perturbations, especially on super-horizon scales. This problem is solved by gauge-invariant perturbation theory which we are going to use throughout this book. The advantage of the gauge-invariant formalism is that the variables used have simple geometric and physical meanings and are not plagued by gauge modes. Although the derivation requires somewhat more work, the final system of perturbation equations is usually simple and well suited for numerical treatment. We shall also see, that on subhorizon scales, the gauge-invariant matter perturbation variables approach the usual, gauge-dependent ones. Since one of the gauge-invariant geometrical perturbation variables corresponds to the Newtonian potential, the Newtonian limit can be performed easily.

First we note that all relativistic equations are covariant and can therefore be written in the form $S = 0$ for some tensor field S. The corresponding background equation is $\overline{S} = 0$, hence $S^{(1)}$ is gauge invariant. It is thus always possible to express the corresponding perturbation equations in terms of gauge-invariant variables.

The principal sources of this chapter are the following reviews on gauge-invariant cosmological perturbation theory (Bardeen, 1980; Kodama & Sasaki, 1984; Mukhanov *et al.*, 1992; Durrer, 1994).

2.2.2 Harmonic decomposition of perturbation variables

Since the $\{t = \text{constant}\}$ hypersurfaces are homogeneous and isotropic, it is reasonable to perform a harmonic analysis: a (spatial) tensor field on these hypersurfaces can be decomposed into components which transform irreducibly under translations and rotations. All such components evolve independently. Decomposition into irreducible components of the translation symmetry corresponds to a harmonic analysis, i.e., decomposition into eigenfunctions of the Laplacian. For a scalar quantity f in the case $K = 0$ this is nothing else than its Fourier decomposition:

$$f(\mathbf{x}, t) = \frac{1}{(2\pi)^3} \int d^3k \, f(\mathbf{k}, t) \, e^{-i\mathbf{kx}} \, . \tag{2.5}$$

(The exponentials $Q_\mathbf{k}(\mathbf{x}) = e^{i\mathbf{kx}}$ are the unitary irreducible representations of the Euclidean translation group.) For $K = 1$ such a decomposition also exists, but the values k are the discrete eigenvalues of the Laplacian on the 3-sphere, $k^2 = \ell(\ell + 2)$

and for $K = -1$, they are bounded from below, $k^2 > 1$. Of course, the functions $Q_{\mathbf{k}}$ depend on the curvature K.

They form the complete orthogonal set of eigenfunctions of the Laplacian,

$$\Delta Q^{(S)}_{\mathbf{k}} = -k^2 Q^{(S)}_{\mathbf{k}} \,. \tag{2.6}$$

In addition, a tensorial variable (at fixed position $\mathbf{x}$) can be decomposed into irreducible components under the rotation group $SO(3)$.

For a spatial vector field, this is its decomposition into a gradient and a curl,

$$V_i = \nabla_i \varphi + B_i \,, \tag{2.7}$$

where

$$B^i_{|i} = 0 \,, \tag{2.8}$$

where we used $X_{|i}$ to denote the three-dimensional covariant derivative of X. Here φ is the spin-0 and $\mathbf{B}$ is the spin-1 component of the vector field V.

For a spatial symmetric tensor field we have

$$H_{ij} = H_L \gamma_{ij} + \left(\nabla_i \nabla_j - \frac{1}{3}\Delta\gamma_{ij}\right) H_T + \frac{1}{2}\left(H^{(V)}_{i|j} + H^{(V)}_{j|i}\right) + H^{(T)}_{ij} \,, \tag{2.9}$$

where

$$H^{(V)|i}_i = H^{(T)^i}_i = H^{(T)^j}_{i|j} = 0 \,. \tag{2.10}$$

Here H_L and H_T are spin-0 components, $H^{(V)}_i$ is the spin-1 component and $H^{(T)}_{ij}$ is the spin-2 component of the tensor field H.

We shall not need higher tensors (or spinors). As a basis for vector and tensor modes we use the vector- and tensor-type eigenfunctions of the Laplacian,

$$\Delta Q^{(V)}_j = -k^2 Q^{(V)}_j \quad \text{and} \tag{2.11}$$

$$\Delta Q^{(T)}_{ji} = -k^2 Q^{(T)}_{ji} \,, \tag{2.12}$$

where $Q^{(V)}_j$ is a transverse vector, $Q^{(V)|j}_j = 0$ and $Q^{(T)}_{ji}$ is a symmetric transverse traceless tensor, $Q^{(T)j}_j = Q^{(T)|i}_{ji} = 0$. Both, $Q^{(V)}_j$ and $Q^{(T)}_{ji}$ have two degrees of freedom. In the case of vanishing curvature we can use an orthonormal basis $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$ in the plane normal to $\mathbf{k}$ and define helicity basis vectors,

$$\mathbf{e}^{\pm} = \frac{1}{\sqrt{2}}(\mathbf{e}^{(1)} \pm i\mathbf{e}^{(2)}) \,. \tag{2.13}$$

In curved spaces the definition of the helicity basis is analogous, but somewhat more involved. Since we shall never need the explicit form of this basis, we shall not enter into this. Vector perturbations can be expanded in terms of this basis,

while tensor perturbations are expanded either in terms of the standard tensor basis given by

$$e^d_{ij} = \frac{1}{2}\left[\mathbf{e}^{(1)}_i \mathbf{e}^{(1)}_j - \mathbf{e}^{(2)}_i \mathbf{e}^{(2)}_j\right] , \tag{2.14}$$

$$e^\times_{ij} = \frac{1}{2}\left[\mathbf{e}^{(1)}_i \mathbf{e}^{(2)}_j + \mathbf{e}^{(2)}_i \mathbf{e}^{(1)}_j\right] , \tag{2.15}$$

or also in terms of a helicity basis defined by

$$e^{(+2)}_{ij} = \mathbf{e}^+_i \mathbf{e}^+_j = e^d_{ij} + i e^\times_{ij} , \tag{2.16}$$

$$e^{(-2)}_{ij} = \mathbf{e}^-_i \mathbf{e}^-_j = e^d_{ij} - i e^\times_{ij} . \tag{2.17}$$

We can develop the vector and tensor basis functions as

$$Q^{(V)}_j = Q^{(1)}\mathbf{e}^{(1)}_j + Q^{(2)}\mathbf{e}^{(2)}_j , \tag{2.18}$$

$$= Q^{(+)}\mathbf{e}^{(+)}_j + Q^{(-)}\mathbf{e}^{(-)}_j , \tag{2.19}$$

$$Q^{(T)}_{ji} = Q^{(d)} e^{(d)}_{ij} + Q^{(\times)} e^{(\times)}_{ij} , \tag{2.20}$$

$$= Q^{(+2)} e^{(+2)}_{ij} + Q^{(-2)} e^{(-2)}_{ij} . \tag{2.21}$$

The components in the 'helicity basis', $\mathbf{e}^{(\pm)}$ and $e^{(\pm 2)}_{ij}$ simply transform with a phase $e^{\pm i\varphi}$, and $e^{\pm 2i\varphi}$ respectively under rotations around $\mathbf{k}$ with angle φ. Hence vector perturbations are spin-1 fields while tensor perturbations are spin-2 fields. The functions $Q^{(+)}$ and $Q^{(+2)}$ have spin up, $m = +1$ and $m = +2$ respectively while $Q^{(-)}$ and $Q^{(-2)}$ have spin down. Scalar perturbations of course have spin zero. We shall make use of this spin structure especially in Chapters 4 and 5.

As in Eqs. (2.7) and (2.9), we can construct scalar-type vectors and symmetric, traceless tensors and vector-type symmetric tensors. To this goal we define

$$Q^{(S)}_j \equiv -k^{-1} Q^{(S)}_{|j} , \tag{2.22}$$

$$Q^{(S)}_{ij} \equiv k^{-2} Q^{(S)}_{|ij} + \frac{1}{3}\gamma_{ij} Q^{(S)} \quad \text{and} \tag{2.23}$$

$$Q^{(V)}_{ij} \equiv -\frac{1}{2k}\left(Q^{(V)}_{i|j} + Q^{(V)}_{j|i}\right) . \tag{2.24}$$

In the following we shall extensively use this decomposition and write down the perturbation equations for a given mode k.

The decomposition of the k-mode of a vector field is then of the form

$$V_i = V Q^{(S)}_i + V^{(V)} Q^{(V)}_i . \tag{2.25}$$

The decomposition of a tensor field is given by (compare Eq. (2.9))

$$H_{ij} = H_L Q^{(S)} \gamma_{ij} + H_T Q^{(S)}_{ij} + H^{(V)} Q^{(V)}_{ij} + H^{(T)} Q^{(T)}_{ij} . \tag{2.26}$$

Here B, $B^{(V)}$, H_L, H_T, $H^{(V)}$ and $H^{(T)}$ are functions of t and $\mathbf{k}$.

This decomposition is very useful since scalar vector and tensor amplitudes of each mode $\mathbf{k}$ evolve independently obeying ordinary differential equations in time.

2.2.3 Metric perturbations

Perturbations of the metric are of the form

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + a^2 h_{\mu\nu} \,. \tag{2.27}$$

We parametrize them as

$$h_{\mu\nu}\, dx^\mu\, dx^\nu = -2A\, dt^2 - 2B_i\, dt\, dx^i + 2H_{ij}\, dx^i\, dx^j, \tag{2.28}$$

and we decompose the perturbation variables B_i and H_{ij} according to (2.25) and (2.26).

Let us consider the behaviour of $h_{\mu\nu}$ under gauge transformations. We set the vector field defining the gauge transformation to

$$X = T\partial_t + L^i \partial_i \,. \tag{2.29}$$

Using the definition of the Lie derivative, we obtain (for details see exercises)

$$\begin{aligned} L_{\mathbf{X}}\bar{g} = a^2 \big[&-2\left(\mathcal{H}T + \dot{T}\right) dt^2 + 2\left(\dot{L}_i - T_{,i}\right) dt\, dx^i \\ &+ \left(2\mathcal{H}T\gamma_{ij} + L_{i|j} + L_{j|i}\right) dx^i\, dx^j \big] . \end{aligned} \tag{2.30}$$

Comparing this with (2.28) and using (2.4), we obtain

$$\begin{aligned} A &\to A + \mathcal{H}T + \dot{T} \,, \\ B_i &\to B_i - \dot{L}_i + T_{,i} \,, \\ H_{ij} &\to H_{ij} + \frac{1}{2}\left(L_{i|j} + L_{j|i}\right) + \mathcal{H}T\gamma_{ij} \,. \end{aligned}$$

Using the decompositions (2.25) for B_i and (2.26) this implies the following behaviour of the perturbation variables under gauge transformations (we also decompose the vector $L_i = L Q_i^{(S)} + L^{(V)} Q_i^{(V)}$):

$$A \to A + \mathcal{H}T + \dot{T} \,, \tag{2.31}$$

$$B \to B - \dot{L} - kT \,, \tag{2.32}$$

$$B^{(V)} \to B^{(V)} - \dot{L}^{(V)} \,, \tag{2.33}$$

$$H_L \to H_L + \mathcal{H}T + \frac{k}{3}L \,, \tag{2.34}$$

$$H_T \to H_T - kL \,, \tag{2.35}$$

$$H^{(V)} \to H^{(V)} - kL^{(V)}\,, \tag{2.36}$$

$$H^{(T)} \to H^{(T)}\,. \tag{2.37}$$

Two scalar and one vector variable can be set to zero by gauge transformations. We shall use this below to choose the longitudinal gauge for scalar perturbations, $B = H_T = 0$.

An interesting variable is also the 'shear' on the $t =$ constant hypersurfaces. We first introduce the normal to this hypersurface which is given by $n_\mu\, dx^\mu = -a(1+A)dt$, so that to first order

$$(n^\mu) = (-g^{\mu\nu}a(1+A)\delta^0_\nu) = a^{-1}(1-A,\, B^i)\,. \tag{2.38}$$

This vector field is normalized, $n^\mu n_\mu = -1$ and its scalar product with any vector field tangent to the $t =$ constant hypersurfaces and hence of the form $X = X^i\partial_i$ vanishes. We now introduce its covariant derivative, setting

$$n_{\mu;\nu} = P_{\mu\nu}\theta + a_\mu n_\nu + \sigma_{\mu\nu} + \omega_{\mu\nu} \tag{2.39}$$

where

$$P_{\mu\nu} = n_\mu n_\nu + g_{\mu\nu} \tag{2.40}$$

is the projection tensor onto the subspace of tangent space normal to n, $\theta \equiv n^\mu_{;\mu}$ is called the 'expansion', $a_\mu = n^\nu n_{\mu;\nu}$ is the acceleration, and $\sigma_{\mu\nu}$ and $\omega_{\mu\nu}$ are the shear and vorticity of the vector field n respectively. They are defined as

$$\sigma_{\mu\nu} = \frac{1}{2}P^\lambda_\mu P^\rho_\nu(n_{\lambda;\rho} + n_{\rho;\lambda}) - P_{\mu\nu}\theta \quad \text{and} \tag{2.41}$$

$$\omega_{\mu\nu} = \frac{1}{2}P^\lambda_\mu P^\rho_\nu(n_{\lambda;\rho} - n_{\rho;\lambda})\,. \tag{2.42}$$

This split of the covariant derivative of a vector field onto expansion, acceleration, shear and vorticity is standard and sometimes very convenient. For example, Frobenius' theorem from differential geometry (see, e.g., Wald (1984), Appendix B) implies, that there exists a hypersurface which is normal to a given vector field if and only if its vorticity vanishes. The direction of the theorem which we use here, namely that the vorticity vanishes if n is (locally) orthogonal to a hypersurface follows from a simple calculation (see Ex. 2.2). The other direction is more involved. In three dimensions it boils down to the well known result that each vector field with vanishing curl can (locally) be written as the gradient of some function.

In the background FL universe, without perturbations, only the expansion, $\theta = 3\mathcal{H}/a = 3H$ is non-zero. In the presence of scalar perturbations we obtain

$$\theta = \frac{3}{a}\mathcal{H}[1 + \mathcal{K}Q] , \qquad \mathcal{K} = -A + \frac{1}{3}\mathcal{H}^{-1}kB + \mathcal{H}^{-1}\dot{H}_L , \tag{2.43}$$

$$\sigma_{00} = \sigma_{0i} = \sigma_{i0} = 0 , \tag{2.44}$$

$$\sigma_{ij} = ak\left(k^{-1}\dot{H}_T - B\right)Q_{ij} \; = \; ak\sigma\, Q_{ij} , \tag{2.45}$$

$$a_i = -kAQ_i , \qquad a_0 = 0 , \tag{2.46}$$

$$\omega_{\mu\nu} = 0 . \tag{2.47}$$

Another interesting variable is the spatial curvature on the hypersurface of constant time. It is easily calculated to first order and one finds

$$(\delta R)^s = 4\frac{k^2 - 3K}{a^2}\left(H_L + \frac{1}{3}H_T\right) = 4\frac{k^2 - 3K}{a^2}\mathcal{R} . \tag{2.48}$$

Since these variables depend only on the time coordinate, they transform only with T under coordinate transformation. Inserting the transformation laws found above, Eqs. (2.31)–(2.35), we obtain

$$\mathcal{K} \to \mathcal{K} - \left(\mathcal{H} - \frac{\dot{\mathcal{H}}}{\mathcal{H}} + \frac{k^2}{3\mathcal{H}}\right)T , \tag{2.49}$$

$$\sigma \to \sigma + kT , \tag{2.50}$$

$$\mathcal{R} \to \mathcal{R} + \mathcal{H}T . \tag{2.51}$$

For vector and tensor perturbations only the perturbation of the shear does not vanish and we have

$$\sigma^{(V)}_{ij} = ak\left(k^{-1}\dot{H}^{(V)} - B^{(V)}\right)Q^{(V)}_{ij} = ak\sigma^{(V)}Q^{(V)}_{ij} , \tag{2.52}$$

$$\sigma^{(T)}_{ij} = a\dot{H}^{(T)}Q^{(T)}_{ij} . \tag{2.53}$$

Since there are no vector-type gauge transformations of the constant time hypersurfaces and no tensor-type gauge transformations at all, the quantities $\sigma^{(V)}$ and $H^{(T)}$ are gauge invariant.

One often chooses the gauge transformation $kL = H_T$ and $kT = B - \dot{L}$, so that the transformed variables H_T and B vanish. In this gauge (longitudinal gauge), scalar perturbations of the metric are of the form ($H_T|_{\rm long} = B|_{\rm long} = 0$):

$$h^{(S)}_{\mu\nu} = -2\Psi\, dt^2 - 2\Phi\gamma_{ij}\, dx^i\, dx^j . \tag{2.54}$$

Ψ and Φ are the so-called *Bardeen* potentials. In a generic gauge the Bardeen potentials are given by

$$\Psi = A - \mathcal{H}k^{-1}\sigma - k^{-1}\dot{\sigma}\ , \tag{2.55}$$

$$\Phi = -H_L - \frac{1}{3}H_T + \mathcal{H}k^{-1}\sigma = -\mathcal{R} + \mathcal{H}k^{-1}\sigma\ , \tag{2.56}$$

where $\sigma = k^{-1}\dot{H}_T - B$, is the scalar potential for the shear of the hypersurface of constant time defined in Eq. (2.45). A short calculation using Eqs. (2.31), (2.50) and (2.51) shows that Ψ and Φ are indeed invariant under gauge transformations.

In a FL universe the Weyl tensor (see Appendix A2.1) vanishes. It therefore is a gauge-invariant perturbation. For scalar perturbations one finds

$$E_{ij} \equiv C^{\mu}{}_{ivj}u_\mu u^\nu = -C^0{}_{i0j} = -\frac{1}{2}\left[(\Psi+\Phi)_{|ij} - \frac{1}{3}\Delta(\Psi+\Phi)\gamma_{ij}\right]\ , \tag{2.57}$$

All other components are also given by E_{ij}, see Appendix 3.1.

For vector perturbations it is convenient to set $kL^{(V)} = H^{(V)}$ so that $H^{(V)}$ vanishes and we have

$$h^{(V)}_{\mu\nu}\,dx^\mu\,dx^\nu = 2\sigma^{(V)}Q^{(V)}_i\,dt\,dx^i. \tag{2.58}$$

We shall call this gauge the 'vector gauge'.

The Weyl tensor from vector perturbation is given by

$$E_{ij} = -C^0{}_{i0j} = \frac{-k}{2}\dot{\sigma}^{(V)}Q^{(V)}_{ij}\ , \tag{2.59}$$

$$\begin{aligned} B_{ij} &\equiv \frac{1}{2}\epsilon_{iv}{}^{\rho\sigma}C_{\rho\sigma}{}^{j\alpha}u_\nu u_\alpha = \epsilon_{ilm}C^0_{jlm}\ , \\ &= \frac{1-k}{2}\sigma^{(V)}\epsilon_{ilm}\left[Q^{(V)}_{l|jm} - Q^{(V)}_{m|jl} - \frac{k^2}{2}\gamma_{jl}Q^{(V)}_m + \frac{k^2}{2}\gamma_{jm}Q^{(V)}_l\right]. \end{aligned} \tag{2.60}$$

Note that from their definition E_{ij} and B_{ij} are symmetric and since $u = (u^0, \mathbf{0})$ to lowest order, only $C^0{}_{i0j}$ and C^0_{ilm} respectively contribute. The tensors E_{ij} and B_{ij}, constructed as given above from the Weyl tensor for an arbitrary 4-velocity field u^μ are normal to u^μ and they determine the Weyl tensor fully (see Appendix 3).

Clearly there are no tensorial (spin-2) gauge transformations and hence $H^{(T)}_{ij}$ is gauge invariant. The expression for the Weyl tensor from tensor perturbation is

$$B_{ij} = -\dot{H}^{(T)}\epsilon_{ilm}\left[Q^{(T)}_{jl|m} - Q^{(T)}_{jm|l}\right]. \tag{2.61}$$

2.2.4 Perturbations of the energy–momentum tensor

Let $T^\mu_\nu = \overline{T}^\mu_\nu + \theta^\mu_\nu$ be the full energy–momentum tensor. We define its energy density ρ and its energy flux 4-vector u as the time-like eigenvalue and eigenvector

of T^μ_ν:

$$T^\mu_\nu u^\nu = -\rho u^\mu, \quad u^2 = -1\,. \tag{2.62}$$

We then parametrize their perturbations by

$$\rho = \bar\rho\,(1+\delta)\,, \quad u = u^0\partial_t + u^i\partial_i\,. \tag{2.63}$$

The component u^0 is fixed by the normalization condition,

$$u^0 = \frac{1}{a}(1-A)\,. \tag{2.64}$$

We further set

$$u^i = \frac{1}{a}v^i = \frac{1}{a}\left(vQ^{(S)i} + v^{(V)}Q^{(V)i}\right). \tag{2.65}$$

$P^\mu_\nu \equiv u^\mu u_\nu + \delta^\mu_\nu$ is the projection tensor onto the subspace of tangent space normal to u. We define the stress tensor

$$\tau^{\mu\nu} = P^\mu_\alpha P^\nu_\beta T^{\alpha\beta}. \tag{2.66}$$

With this we can write

$$T^\mu_\nu = \rho u^\mu u_\nu + \tau^\mu_\nu\,. \tag{2.67}$$

In the unperturbed case we have $\tau^0_\mu = \tau^\mu_0 = 0$ and $\tau^i_j = \bar P\delta^i_j$. Including first-order perturbations, the components $\tau_{0\mu}$ are determined by the perturbation variables which we have already introduced. We obtain

$$\tau^0_0 = 0\,, \quad \text{and} \quad \tau^j_0 = -\bar P v^j\,, \qquad \tau^0_j = \bar P(v_j - B_j)\,. \tag{2.68}$$

But τ^i_j contains in general new perturbations. We define

$$\tau^i_j = \bar P\left[(1+\pi_L)\,\delta^i_j + \Pi^i_j\right], \quad \text{with} \quad \Pi^i_i = 0\,. \tag{2.69}$$

From our definitions we can determine the perturbations of the energy–momentum tensor. A short calculation gives

$$T^0_0 = -\bar\rho(1+\delta)\,, \tag{2.70}$$

$$T^0{}_j = (\bar\rho + \bar P)(v_j - B_j)\,, \tag{2.71}$$

$$T^j{}_0 = -(\bar\rho + \bar P)v^j\,, \tag{2.72}$$

$$T^i{}_j = \bar P\left[(1+\pi_L)\delta^i_j + \Pi^i{}_j\right]\,. \tag{2.73}$$

The traceless part of the stress tensor, Π^i_j, is called the anisotropic stress tensor. We decompose it as

$$\Pi^i_j = \Pi Q^{(S)i}_j + \Pi^{(V)}Q^{(V)i}_j + \Pi^{(T)}Q^{(T)i}_j. \tag{2.74}$$

We now study the gauge transformation properties of these perturbation variables. First we note that ρ is a scalar and $L_X\bar{\rho} = \dot{\bar{\rho}}T = -3(1+w)\mathcal{H}\bar{\rho}T$. Here we made use of Eq. (1.22). The same is true for $\bar{P}(1+\pi_L)$ which is 1/3 of the trace of $\tau^{\mu}{}_{\nu}$. With Eq. (1.29), we obtain $L_X\bar{P} = \dot{\bar{P}}T = -3\frac{c_s^2}{w}(1+w)\mathcal{H}\bar{P}T$. The background contribution to the anisotropic stress tensor, $\Pi^{\mu}_{\nu} = \tau^{\mu}_{\nu} - \frac{1}{3}\tau^{\alpha}_{\alpha}\delta^{\mu}_{\nu}$, vanishes, hence Π^{μ}_{ν} is gauge invariant (the Stewart–Walker lemma). For perfect fluids $\Pi^{\mu}_{\nu} = 0$. For the velocity we use $L_X\bar{u} = [X,\bar{u}] = (-T\dot{a}a^{-2} - a^{-1}\dot{T})\partial_t - a^{-1}\dot{L}^i\partial_i$. Inserting our decomposition into scalar, vector and tensor perturbation variables for a fixed mode $\mathbf{k}$, we obtain finally the following transformation behaviour

$$\delta \to \delta - 3(1+w)\mathcal{H}T\ , \tag{2.75}$$

$$\pi_L \to \pi_L - 3\frac{c_s^2}{w}(1+w)\mathcal{H}T\ , \tag{2.76}$$

$$v \to v - \dot{L}\ , \tag{2.77}$$

$$\Pi \to \Pi\ , \tag{2.78}$$

$$v^{(V)} \to v^{(V)} - \dot{L}^{(V)}\ , \tag{2.79}$$

$$\Pi^{(V)} \to \Pi^{(V)}\ , \tag{2.80}$$

$$\Pi^{(T)} \to \Pi^{(T)}\ . \tag{2.81}$$

Apart from the anisotropic stress perturbations, there is only one gauge-invariant variable which can be obtained from the energy–momentum tensor alone, namely

$$\Gamma = \pi_L - \frac{c_s^2}{w}\delta\ . \tag{2.82}$$

One can show (see Appendix 5) that Γ is proportional to the divergence of the entropy flux of the perturbations. Adiabatic perturbations are characterized by $\Gamma = 0$.

Gauge-invariant density and velocity perturbations can be found by combining δ, v and $v_i^{(V)}$ with metric perturbations. We shall use

$$V \equiv v - \frac{1}{k}\dot{H}_T = v^{\rm long}\ , \tag{2.83}$$

$$D_s \equiv \delta + 3(1+w)\mathcal{H}(k^{-2}\dot{H}_T - k^{-1}B) \equiv \delta^{\rm long}\ , \tag{2.84}$$

$$\begin{aligned} D &\equiv \delta^{\rm long} + 3(1+w)\frac{\mathcal{H}}{k}V = \delta + 3(1+w)\frac{\mathcal{H}}{k}(v-B) \\ &= D_s + 3(1+w)\frac{\mathcal{H}}{k}V\ , \end{aligned} \tag{2.85}$$

$$\begin{aligned} D_g &\equiv \delta + 3(1+w)\left(H_L + \frac{1}{3}H_T\right) = \delta^{\rm long} - 3(1+w)\Phi \\ &= D_s - 3(1+w)\Phi\ , \end{aligned} \tag{2.86}$$

$$V^{(V)} \equiv v^{(V)} - \frac{1}{k}\dot{H}^{(V)} = v^{(\mathrm{vec})} , \tag{2.87}$$

$$\Omega \equiv v^{(V)} - B^{(V)} = v^{(\mathrm{vec})} - B^{(V)} , \tag{2.88}$$

$$\Omega - V^{(V)} = \sigma^{(V)} . \tag{2.89}$$

Here v^{long}, δ^{long} and $v^{(\mathrm{vec})}$ are the velocity (and density) perturbations in the longitudinal and vector gauge respectively, and $\sigma^{(V)}$ is the metric perturbation in vector gauge and the shear of the $t =$ constant hypersurfaces (see Eqs. (2.52) and (2.58)).

These variables can be interpreted nicely in terms of gradients of the energy density and the shear and vorticity of the velocity field (Ellis & Bruni, 1989). Here we just calculate the covariant derivative of the velocity field u^μ and decompose it like the normal field n^μ. In a non-perturbed FL universe these two vector fields coincide. With our definition of variables, a short calculation using $u_{\mu;\nu} = u_{\mu,\nu} - \Gamma^\beta_{\mu\nu}u_\beta$ gives

$$u_{\mu;\nu} = P^{(f)}_{\mu\nu}\theta^{(f)} + a^{(f)}_\mu u_\nu + \sigma^{(f)}_{\mu\nu} + \omega^{(f)}_{\mu\nu} , \tag{2.90}$$

where the projection, $P^{(f)}$, expansion, $\theta^{(f)}$, acceleration, $a^{(f)}$, shear, $\sigma^{(f)}$ and vorticity, $\omega^{(f)}$ are defined as in Eqs. (2.40)–(2.42), just the normal field n^μ is replaced by u^μ, the energy flux of the fluid. We indicate this by the superscript $^{(f)}$. For scalar perturbations one finds

$$\theta^{(f)} = \frac{3}{a}\mathcal{H}[1 + \mathcal{K}^{(f)}Q] , \qquad \mathcal{K}^{(f)} = -A + \mathcal{H}^{-1}\left(\dot{H}_L + \frac{k}{3}v\right) , \tag{2.91}$$

$$\sigma^{(f)}_{00} = \sigma^{(f)}_{0i} = \sigma^{(f)}_{i0} = 0 , \tag{2.92}$$

$$\sigma^{(f)}_{ij} = ak(k^{-1}\dot{H}_T - v)Q_{ij} \;=\; ak\sigma^{(f)}Q_{ij} , \tag{2.93}$$

$$a^{(f)}_i = -A^{(f)}Q_i \;, A^{(f)} = kA - \mathcal{H}(v - B) + (\dot{v} - \dot{B}) \;, a_0 = 0 , \tag{2.94}$$

$$\omega_{\mu\nu} = 0 . \tag{2.95}$$

Contrary to n^μ, the vector field u^μ is defined independently of the coordinate system. Therefore, and since $a^{(f)}_\mu$ and $\sigma^{(f)}_{\mu\nu}$ vanish in the background FL universe, the variables $A^{(f)}$ and V are gauge invariant. For V we have already noticed this before. Furthermore, it is easy to check that

$$A^{(f)} = k\Psi - \mathcal{H}V + \dot{V} ,$$

which is a gauge-invariant variable called the 'peculiar acceleration'.

For vector perturbations we obtain

$$\sigma^{(f)}_{00} = \sigma^{(f)}_{0i} = \sigma^{(f)}_{i0} = 0 , \tag{2.96}$$

$$\sigma^{(f)}_{ij} = ak\left(k^{-1}\dot{H}^{(V)} - v^{(V)}\right)Q^{(V)}_{ij} \;=\; -akV^{(V)}Q_{ij} , \tag{2.97}$$

$$\omega_{i0}^{(f)} = \omega_{0i}^{(f)} = 0\,, \tag{2.98}$$

$$\omega_{ij}^{(f)} = a\left(v^{(V)} - B^{(V)}\right)\left[Q_{i|j}^{(V)} - Q_{j|i}^{(V)}\right] \;=\; a\Omega\left[Q_{i|j}^{(V)} - Q_{j|i}^{(V)}\right]\,, \tag{2.99}$$

$$a_i = \dot{V}^{(V)} Q_i^{(V)}\,. \tag{2.100}$$

It is interesting to note that the energy flux of scalar perturbations is hypersurface orthogonal, $\omega^{(S)} = 0$, while vector perturbations do have non-vanishing curl if $v^{(V)} \neq B^{(V)}$. A coordinate system with $v = B$ is called 'comoving'.

Tensor perturbations do not admit a perturbed energy flux so that for them the above perturbation variables vanish.

We now want to show that on scales much smaller than the Hubble scale, $k \gg \mathcal{H} \sim t^{-1}$, the metric perturbations are much smaller than δ and v and we can thus neglect the difference between different gauges and/or gauge-invariant variables. This is especially important when comparing experimental results with gauge-invariant calculations. Let us neglect spatial curvature in the following order of magnitude argument. Then, the perturbations of the Einstein tensor are a combination of the second derivatives of the metric perturbations, $\mathcal{H}$ times the first derivatives and $\mathcal{H}^2$ or $\dot{\mathcal{H}}$ times metric perturbations. The first-order perturbation of Einstein's equations therefore generically yield the following order of magnitude estimate $8\pi G\delta T_{\mu\nu} = \delta G_{\mu\nu}$:

$$\mathcal{O}\left(\frac{\delta T_{\mu\nu}}{\rho}\right) \underbrace{\mathcal{O}\left(8\pi G\rho\right)}_{\mathcal{O}(a'/a)^2 = \mathcal{O}\left(a^2/t^2\right)} = \mathcal{O}\left(\frac{1}{t^2}a^2h + \frac{k}{t}a^2h + k^2a^2h\right)\,, \tag{2.101}$$

$$\mathcal{O}\left(\frac{\delta T_{\mu\nu}}{\rho}\right) = \mathcal{O}\left(h + kth + (kt)^2h\right)\,. \tag{2.102}$$

For $kt \gg 1$ this gives $\mathcal{O}(\delta, v) = \mathcal{O}\left(\delta T_{\mu\nu}/\rho\right) \gg \mathcal{O}(h)$. Therefore, on subhorizon scales the differences between δ, $\delta^{\rm long}$, D_g and D are negligible as are the differences between v and V or $v^{(V)}$, $V^{(V)}$ and $\Omega^{(V)}$. Since measurements of density and velocity perturbations can only be made on subhorizon scales, we may therefore use any of the gauge-invariant perturbation variables to compare with measurements.

2.3 The perturbation equations

We do not derive the first-order perturbations of Einstein's equations. By elementary algebraic methods, this is quite lengthy and cumbersome. However, we recommend that the student simply determines $\delta G_{\mu\nu}$ in longitudinal (vector) gauge using some algebraic package like Maple or Mathematica and then writes down the resulting Einstein equations using gauge-invariant variables. Since we know that these variables do not depend on the coordinates chosen, the equations obtained in this way are valid in any gauge. Here, we just present the resulting equations in

gauge-invariant form. A rapid derivation by hand is possible using the $3+1$ formalism of general relativity and working with Cartan's formalism for the Riemann curvature (see Durrer & Straumann, 1988). In order to simplify the notation, we suppress the overbar on background quantities whenever this does not lead to confusion.

2.3.1 Einstein's equations

The constraints

The Einstein equations $G_{0i\mu} = 8\pi G T_{0i\mu}$ lead to two scalar and one vector constraint equations,

$$\left.\begin{aligned} 4\pi G a^2 \rho D &= -(k^2 - 3K)\Phi \quad (00) \\ 4\pi G a^2 (\rho + P) V &= k\left(\mathcal{H}\Psi + \dot{\Phi}\right) \quad (0i) \end{aligned}\right\} \quad \text{(scalar)}\,, \tag{2.103}$$

$$8\pi G a^2 (\rho + P)\Omega = \frac{1}{2}\left(2K - k^2\right)\sigma^{(V)} \quad (0i) \qquad \text{(vector)}\,. \tag{2.104}$$

The dynamical equations

The Einstein equations $G_{ij} = 8\pi G T_{ij}$ provide two scalar, one vector and one tensor perturbation equations,

scalar:

$$k^2(\Phi - \Psi) = 8\pi G a^2 P \Pi^{(S)} \qquad (i \neq j)\,, \tag{2.105}$$

$$\begin{aligned} &\ddot{\Phi} + 2\mathcal{H}\dot{\Phi} + \mathcal{H}\dot{\Psi} + \left[2\dot{\mathcal{H}} + \mathcal{H}^2 - \frac{k^2}{3}\right]\Psi \\ &\quad = 4\pi G a^2 \rho \left[\frac{1}{3}D + c_s^2 D_s + w\Gamma\right] \qquad (ii)\,, \end{aligned} \tag{2.106}$$

vector:

$$k\left(\dot{\sigma}^{(V)} + 2\mathcal{H}\sigma^{(V)}\right) = 8\pi G a^2 P \Pi^{(V)}\,, \tag{2.107}$$

tensor:

$$\ddot{H}^{(T)} + 2\mathcal{H}\dot{H}^{(T)} + \left(2K + k^2\right) H^{(T)} = 8\pi G a^2 P \Pi^{(T)}\,. \tag{2.108}$$

The second dynamical scalar equation is somewhat cumbersome and not often used, since we may use one of the conservation equations given below instead. For the derivation of the perturbed Einstein equation the following relations are useful. They can be derived from the Friedmann equations (1.20)–(1.22); a possible cosmological constant is included in ρ and P.

$$4\pi G a^2 \rho (1 + w) = \mathcal{H}^2 - \dot{\mathcal{H}} + K\,, \tag{2.109}$$

$$\dot{\mathcal{H}} = -\frac{1 + 3w}{2}\left(\mathcal{H}^2 + K\right)\,, \tag{2.110}$$

$$4\pi G a^2 \rho (1+w) 3 c_s^2 = \frac{\ddot{\mathcal{H}}}{\mathcal{H}} - \dot{\mathcal{H}} - \mathcal{H}^2 - K \,, \tag{2.111}$$

$$c_s^2 = \frac{\frac{\ddot{\mathcal{H}}}{\mathcal{H}} - \dot{\mathcal{H}} - \mathcal{H}^2 - K}{3[\mathcal{H}^2 - \dot{\mathcal{H}} + K]} \,. \tag{2.112}$$

For the calculations below we shall also make use of

$$\dot{w} = 3(w - c_s^2)(1+w)\mathcal{H} \,. \tag{2.113}$$

Note that for perfect fluids, where $\Pi^i_j \equiv 0$, we have $\Phi = \Psi$. As we shall see below, for perfect fluids with $\Gamma = \Pi = 0$, the behaviour of scalar perturbations is given by Ψ, which describes a damped wave propagating with speed c_s^2.

Tensor perturbations are given by $H^{(T)}$, which for perfect fluids also obeys a damped wave equation propagating with the speed of light. On small scales (over short time periods) when $t^{-2} \lesssim 2K + k^2$, the damping term can be neglected and H_{ij} represents propagating gravitational waves. For vanishing curvature or $k^2 \gg K$, small scales are just the sub-Hubble scales, $kt \gtrsim 1$. For $K < 0$, waves oscillate with a somewhat smaller frequency, $\omega = \sqrt{2K + k^2} < k$, while for $K > 0$ the frequency is somewhat higher than k.

Vector perturbations of a perfect fluid are determined by the $\sigma^{(V)}$ equation, Eq. (2.107), which implies $\sigma^{(V)} \propto 1/a^2$. Hence vector perturbations do not oscillate, they simply decay.

2.3.2 Energy–momentum conservation

The conservation equations, $T^{\mu\nu}_{\ ;\nu} = 0$ lead to the following perturbation equations:

$$\left.\begin{aligned} \dot{D}_g + 3\left(c_s^2 - w\right)\mathcal{H} D_g + (1+w)kV + 3w\mathcal{H}\Gamma = 0 \\ \dot{V} + \mathcal{H}\left(1 - 3c_s^2\right)V = k\left(\Psi + 3c_s^2\Phi\right) + \frac{c_s^2 k}{1+w} D_g \\ + \frac{wk}{1+w}\left[\Gamma - \frac{2}{3}\left(1 - \frac{3K}{k^2}\right)\Pi\right] \end{aligned}\right\} \text{ (scalar)} \,, \tag{2.114}$$

$$\dot{\Omega} + \left(1 - 3c_s^2\right)\mathcal{H}\Omega = -\frac{w}{2(1+w)}\left(k - \frac{2K}{k}\right)\Pi^{(V)} \qquad \text{(vector)} \,. \tag{2.115}$$

It is sometimes also useful to express the scalar conservation equations in terms of the variable pair (D, V). Using $D = D_g + 3(1+w)\left[\mathcal{H}k^{-1}V + \Phi\right]$ in (2.114) one obtains after some algebra and making use of the $(0i)$ constraint equation (2.103)

$$\dot{D} - 3w\mathcal{H}D = -\left(1 - \frac{3K}{k^2}\right)[(1+w)kV + 2\mathcal{H}w\Pi] \,, \tag{2.116}$$

$$\dot{V} + \mathcal{H}V = k\left[\Psi + \frac{c_s^2}{1+w}D + \frac{w}{1+w}\Gamma - \frac{2}{3}\left(1 - \frac{3K}{k^2}\right)\frac{w}{1+w}\Pi\right] . \tag{2.117}$$

Replacing Ψ in Eq. (2.117) via the (00) and (ij) Einstein equations, (2.103) and (2.105), and replacing V via Eq. (2.116) we can derive a second-order equation for D. A lengthy but straightforward calculation gives

$$\ddot{D} + (1 + 3c_s^2 - 6w)\mathcal{H}\dot{D} + \left[\left(\frac{9}{2}w^2 - 12w + 9c_s^2 - \frac{3}{2}\right)\mathcal{H}^2 + \frac{3}{2}(3w^2 - 1)K + (k^2 - 3K)c_s^2\right] D = -(k^2 - 3K)w\Gamma - 2\left(1 - \frac{3K}{k^2}\right)\mathcal{H}w\dot{\Pi} + 2\left[(3w^2 + 3c_s^2 - 2w)\mathcal{H}^2 + w(3w + 2)K + \frac{k^2 - 3K}{3}w\right]\left(1 - \frac{3K}{k^2}\right)\Pi\,. \tag{2.118}$$

The conservation equations can, of course, also be obtained from the Einstein equations since they are equivalent to the contracted Bianchi identities (see Appendix A2.1). For scalar perturbations we have four independent equations and six variables. For vector perturbations we have two equations and three variables, while for tensor perturbations we have one equation and two variables. To close the system we must add some matter equations. The simplest prescription is to set $\Gamma = \Pi_{ij} = 0$. This matter equation, which describes adiabatic perturbations of a perfect fluid gives us exactly two additional equations for scalar perturbations and one each for vector and tensor perturbations. In this simple case, the tensor equation simply describes free gravitational waves propagating in a FL background. If $c_s^2 \neq 0$ also the scalar equation (2.118) is a wave equation. It describes what we shall call 'acoustic oscillations' of the fluid where the fluid pressure counter-acts gravitational collapse. The vector perturbation equation, however, is of first order. $\Pi^{(V)} = 0$ implies $\sigma^{(V)} \propto 1/a^2$ and $\Omega \propto a^{-1+3c_s^2}$. Hence vector perturbations of the metric simply decay if there are no anisotropic stresses to source them.

Another simple example is a universe with matter content given by a scalar field. We shall discuss this case in the next section. More complicated are several interacting particle species of which some have to be described by a Boltzmann equation. This is the actual universe at late times, $z \lesssim 10^7$.

2.3.3 *Mixtures of several fluids*

Here we only consider fluid components that are non-interacting, so that their energy–momentum tensor is separately conserved, i.e., equations (2.114) and (2.115) hold for each α component separately. The Einstein equations, however

determine the metric perturbations induced by the full perturbations,

$$\rho D_g = \sum_\alpha \rho_\alpha D_{g\alpha} \,, \tag{2.119}$$

$$(\rho + P)V = \sum_\alpha (\rho_\alpha + P_\alpha) V_\alpha \,, \tag{2.120}$$

$$P\Pi = \sum_\alpha P_\alpha \Pi_\alpha \,, \tag{2.121}$$

$$\begin{aligned} P\Gamma = \pi_L - c_s^2 \delta\rho &= \sum_\alpha P_\alpha \Gamma_\alpha + \sum_\alpha (c_\alpha^2 - c_s^2)\delta\rho_\alpha \\ &= \sum_\alpha P_\alpha \Gamma_\alpha + P\Gamma_{\rm rel} \,. \end{aligned} \tag{2.122}$$

In order to see that $\Gamma_{\rm rel}$ is gauge invariant we use Eq. (1.29)

$$c_s^2 = \frac{\dot P}{\dot\rho} = \sum_\beta \frac{(1+w_\beta)\rho_\beta c_\beta^2}{(1+w)\rho} \,.$$

Also using $\sum_\beta (1+w_\beta)\rho_\beta = (1+w)\rho$ we find

$$\begin{aligned} P\Gamma_{\rm rel} &= \sum_\alpha (c_\alpha^2 - c_s^2)\delta\rho_\alpha = \sum_{\alpha\beta} \frac{(1+w_\beta)\rho_\beta (1+w_\alpha)\rho_\alpha}{\rho + P}(c_\alpha^2 - c_\beta^2)\frac{\delta_\alpha}{1+w_\alpha} \\ &= \frac{1}{2}\sum_{\alpha\beta} \frac{(1+w_\beta)(1+w_\alpha)\rho_\beta\rho_\alpha}{\rho+P}(c_\alpha^2 - c_\beta^2)\left[\frac{\delta_\alpha}{1+w_\alpha} - \frac{\delta_\beta}{1+w_\beta}\right] \\ &= \frac{1}{2}\sum_{\alpha\beta} \frac{(1+w_\beta)(1+w_\alpha)\rho_\beta\rho_\alpha}{\rho+P}(c_\alpha^2 - c_\beta^2)\left[\frac{D_{g\alpha}}{1+w_\alpha} - \frac{D_{g\beta}}{1+w_\beta}\right] \\ &= \frac{1}{2}\sum_{\alpha\beta} \frac{(1+w_\beta)\rho_\beta (1+w_\alpha)\rho_\alpha}{\rho+P}(c_\alpha^2 - c_\beta^2) S_{\alpha\beta}, \end{aligned} \tag{2.123}$$

where we define

$$S_{\alpha\beta} = \left[\frac{D_{g\alpha}}{1+w_\alpha} - \frac{D_{g\beta}}{1+w_\beta}\right] \,. \tag{2.124}$$

For the third equal sign above we have used the fact that the expression $[(1+w_\beta)(1+w_\alpha)\rho_\beta\rho_\alpha/(\rho+P)](c_\alpha^2 - c_\beta^2)$ is anti-symmetric in α and β and we therefore may also anti-symmetrize the remaining factor.

The individual components of the gauge-invariant velocity and density perturbations are defined via their energy–momentum tensors. Note that

$$V_\alpha = v_\alpha - k^{-1}\dot H_T \,, \qquad \text{and} \tag{2.125}$$

$$D_{g\alpha} = \delta_\alpha + 3(1+w_\alpha)\mathcal{R} \,, \tag{2.126}$$

$$D_\alpha = \delta_\alpha + 3(1+w_\alpha)\mathcal{H}k^{-1}(v_\alpha - B) \,, \tag{2.127}$$

$$= D_{g\alpha} + 3(1+w_\alpha)\left[\mathcal{H}k^{-1}V_\alpha + \Phi\right] \,. \tag{2.128}$$

It is easy to check that the conservation equations (2.114) are also valid for a mixture of conserved components, so that we have

$$\dot{D}_{g\alpha} + 3\left(c_\alpha^2 - w_\alpha\right)\mathcal{H}D_{g\alpha} = -(1+w_\alpha)kV_\alpha - 3w_\alpha\mathcal{H}\Gamma_\alpha \,, \tag{2.129}$$

$$\dot{V}_\alpha + \mathcal{H}\left(1 - 3c_\alpha^2\right)V_\alpha = k\left(\Psi + 3c_\alpha^2\Phi\right) + \frac{c_\alpha^2 k}{1+w_\alpha}D_{g\alpha} + \frac{w_\alpha k}{1+w_\alpha}\left[\Gamma_\alpha - \frac{2}{3}\left(1 - \frac{3K}{k^2}\right)\Pi_\alpha\right] . \tag{2.130}$$

However, if we rewrite the conservation equations in terms of the variables (D_α, V_α) new terms appear since we have to use the Einstein equations in the derivation. A somewhat tedious but straightforward calculation, replacing $D_{g\alpha}$ with the help of Eq. (2.128) and then eliminating $\dot{\Phi}$ with the Einstein equation $(0i)$ gives

$$\dot{D}_\alpha - 3w_\alpha\mathcal{H}D_\alpha = \frac{9}{2}(\mathcal{H}^2 + K)k^{-1}(1+w)(1+w_\alpha)[V - V_\alpha] - \left(1 - \frac{3K}{k^2}\right)[(1+w_\alpha)kV_\alpha + 2\mathcal{H}w\Pi_\alpha] \,, \tag{2.131}$$

$$\dot{V}_\alpha + \mathcal{H}V_\alpha = k\left[\Psi + \frac{c_\alpha^2}{1+w_\alpha}D_\alpha + \frac{w_\alpha}{1+w_\alpha}\Gamma_\alpha - \frac{2}{3}\left(1 - \frac{3K}{k^2}\right)\frac{w_\alpha}{1+w_\alpha}\Pi_\alpha\right] . \tag{2.132}$$

It is sometimes more useful to describe mixed systems in terms of variables related to differences of individual components. With $S_{\alpha\beta}$ given in Eq. (2.124) and defining

$$V_{\alpha\beta} = V_\alpha - V_\beta \,, \tag{2.133}$$

$$\Gamma_{\alpha\beta} = \frac{w_\alpha}{1+w_\alpha}\Gamma_\alpha - \frac{w_\beta}{1+w_\beta}\Gamma_\beta \,, \tag{2.134}$$

$$\Pi_{\alpha\beta} = \frac{w_\alpha}{1+w_\alpha}\Pi_\alpha - \frac{w_\beta}{1+w_\beta}\Pi_\beta \,, \tag{2.135}$$

one can derive the following system of equations from Eqs. (2.129) and (2.130)

$$\dot{S}_{\alpha\beta} = -kV_{\alpha\beta} - 3\mathcal{H}\Gamma_{\alpha\beta} \,, \tag{2.136}$$

$$\dot{V}_{\alpha\beta} + \mathcal{H}V_{\alpha\beta} - \frac{3}{2}\mathcal{H}(c_\alpha^2 + c_\beta^2)V_{\alpha\beta} - \frac{3}{2}\mathcal{H}(c_\alpha^2 - c_\beta^2)\sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P}\left(V_{\alpha\gamma} + V_{\beta\gamma}\right) = k\left[\frac{c_\alpha^2 - c_\beta^2}{1+w}D + \frac{c_\alpha^2 + c_\beta^2}{2}S_{\alpha\beta} + \frac{c_\alpha^2 - c_\beta^2}{2}\sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P}\left(S_{\alpha\gamma} + S_{\beta\gamma}\right) + \Gamma_{\alpha\beta} - \frac{3}{2}\left(1 - \frac{3K}{k^2}\right)\Pi_{\alpha\beta}\right] . \tag{2.137}$$

We present a detailed derivation of these equations in Appendix 6.

We shall use these equations when we discuss mixtures of cold dark matter and radiation. More details on mixed systems which also include interactions can be found in Kodama & Sasaki (1984). In Ex. 2.3, we discuss a simple example of a mixed system. Interacting mixed systems are not very relevant for us, since we shall describe them with a Boltzmann equation approach that we develop in Chapter 4.

2.3.4 The Bardeen equation

The systems of equations which we have presented here are, of course, not closed. To close them one needs to add evolution equations for the matter variables, such as $\Pi^{(T)}$ for tensor perturbations, a relation between $\Pi^{(V)}$ and $\Omega^{(V)}$ for vector perturbations, and expressions for Γ and $\Pi = \Pi^{(S)}$ for scalar perturbations.

For scalar perturbations we can actually derive an evolution equation for Ψ, where Γ and Π enter only as source terms. Replacing D and D_s in (2.106) by use of (2.85) and (2.103) and replacing Φ by Π and Ψ via Eq. (2.105) leads to

$$\ddot{\Phi} + 3\mathcal{H}(1+c_s^2)\dot{\Phi} + \left[3(c_s^2 - w)\mathcal{H}^2 - (2 + 3w + 3c_s^2)K + c_s^2 k^2\right]\Phi$$
$$= \frac{8\pi G a^2 P}{k^2}\left[\mathcal{H}\dot{\Pi} + [2\dot{\mathcal{H}} + 3\mathcal{H}^2(1 - c_s^2/w)]\Pi - \frac{1}{3}k^2\Pi + \frac{k^2}{2}\Gamma\right]. \qquad (2.138)$$

This is the Bardeen equation. To derive it we also made use of (2.110) to replace $\dot{\mathcal{H}}$.

This equation is especially useful in terms of another gauge-invariant variable which we now introduce: the scalar curvature on the comoving hypersurface. The comoving hypersurface is defined by having the normal n on the constant time hypersurface equal to the particle 4-velocity u. Using $(n^\nu) = a^{-1}(1 - A, B^j)$ and $(u^\nu) = a^{-1}(1 - A, v^j)$ this implies $v = B$. From the definitions of σ and V we thus have $V = -\sigma_{\rm co}$ in this coordinate system. In comoving gauge we therefore have (see Eqs. (2.84)–(2.86))

$$D_s = \delta_{\rm co} - 3(1+w)k^{-1}\mathcal{H}V\ , \qquad (2.139)$$
$$D = \delta_{\rm co}\ , \qquad (2.140)$$
$$D_g = \delta_{\rm co} + 3(1+w)\mathcal{R}_{\rm co} = D - 3(1+w)[k^{-1}\mathcal{H}V + \Phi]\ , \qquad (2.141)$$

so that

$$-\mathcal{R}_{\rm co} = \frac{1}{3(1+w)}[D - D_g] = k^{-1}\mathcal{H}V + \Phi\ . \qquad (2.142)$$

Here the index 'co' indicates comoving coordinates. Using the $(0i)$ Einstein equation, Eq. (2.103), we obtain

$$-\mathcal{R}_{\rm co} = \frac{2}{3(1+w)}\left[\Psi + \mathcal{H}^{-1}\dot{\Phi}\right] + \Phi \equiv \zeta \ . \tag{2.143}$$

We are especially interested in the evolution of the curvature perturbation variable ζ in situations where we can neglect anisotropic stresses. Then the right-hand side of Eq. (2.138) simply becomes $4\pi G a^2 P\Gamma$ and $\Phi = \Psi$. The definition (2.143) together with the Bardeen equation then yields in the spatially flat case, $k = 0$

$$\dot{\zeta} = \frac{\mathcal{H}}{\mathcal{H}^2 - \dot{\mathcal{H}}}\left[\frac{3w}{2}\mathcal{H}^2\Gamma - c_s^2 k^2\Psi\right] , \tag{2.144}$$

$$= \frac{w}{w+1}\mathcal{H}\Gamma - \frac{2}{3(w+1)}\mathcal{H}^{-1} - c_s^2 k^2\Psi \ . \tag{2.145}$$

For adiabatic perturbations, $\Gamma = 0$ the curvature perturbation ζ is therefore conserved on super-Hubble scales, $k/\mathcal{H} \ll 1$ at early times when curvature is certainly negligible. This will be very useful when we want to specify initial conditions in Chapter 3.

Also note that for constant w and constant Bardeen potential $\Psi = \Phi$, the curvature perturbation ζ differs from the Bardeen potential only by a multiplicative constant.

2.3.5 A special case

Here we want to discuss the scalar perturbation equations for a simple, but important, special case. We consider adiabatic perturbations of a perfect fluid. In this case there are no anisotropic stresses, $\Pi = 0$. Furthermore, the pressure fluctuation $\delta P = \pi_L P$ is related to the density fluctuation $\delta\rho$ by $\delta P = c_s^2\delta\rho$, hence $\Gamma = 0$. Eq. (2.138) then becomes simply a second-order equation for the Bardeen potential $\Psi = \Phi$, which is, in this case the only dynamical degree of freedom,

$$\ddot{\Psi} + 3\mathcal{H}(1+c_s^2)\dot{\Psi} + \left[(1+3c_s^2)(\mathcal{H}^2 - K) - (1+3w)(\mathcal{H}^2 + K) + c_s^2 k^2\right]\Psi = 0 \ . \tag{2.146}$$

This is a damped wave equation. When we may neglect curvature, and if $w =$ constant so that $c_s^2 = w$, the time-dependent mass term $m^2(t) = -(1+3c_s^2)(\mathcal{H}^2 - K) + (1+3w)(\mathcal{H}^2 + K)$ vanishes. Eq. (2.146) then reduces to

$$\ddot{\Psi} + 6\frac{1+w}{(1+3w)t}\dot{\Psi} + wk^2\Psi = 0 \ , \tag{2.147}$$

where we have used that

$$a \propto t^{2/(1+3w)} = t^q \quad \text{and} \quad \mathcal{H} = \frac{2}{1+3w}\frac{1}{t} = \frac{q}{t} \quad q = \frac{2}{1+3w} \,.$$

Eq. (2.147) has an exact solution of the form

$$\Psi = \frac{1}{a}\left(A j_q(\sqrt{w}kt) + B y_q(\sqrt{w}kt)\right) \,, \tag{2.148}$$

where j_q and y_q denote the spherical Bessel functions of order q. Using $j_q(x) \propto x^q$ and $y_q(x) \propto x^{-q-1}$ for $x \ll 1$, we find that the A-mode is constant while the B-mode decays like $1/(a^2 t)$ on super-Hubble scales. If both modes are generated with similar amplitudes, the B-mode is therefore negligible after a few expansion times. On sub-Hubble scales, $\sqrt{w}kt \gg 1$, the solution oscillates with frequency $\sqrt{w}k$ and decay like $1/(at)$. The only exception is the case of cosmic dust (CDM) with $w = 0$. In this case the oscillatory term drops and the solution is of the form

$$\Psi = A + \frac{B}{(kt)^5} \,. \tag{2.149}$$

For later use we collect the main results in the following equation: for power law expansion $a \propto t^q$ we find

$$\Psi = \begin{cases} \text{constant} & \text{for} \quad \sqrt{w}kt \ll 1 \\ \frac{A}{a\sqrt{w}kt}\sin(\sqrt{w}kt - \frac{q}{2}\pi) & \text{for} \quad \sqrt{w}kt \gg 1 \,, \end{cases} \quad w \neq 0 \,. \tag{2.150}$$

We now consider a universe which starts out in a radiation dominated era with a spectrum (see Section 2.6) $\langle|\Psi|^2\rangle k^3 = A_S(k/H_0)^{n-1}$ and which becomes matter dominated at some time $t_{\rm eq}$. Late in the matter dominated era the spectrum of Ψ is therefore approximately given by (see Fig. 2.1)

$$\langle|\Psi|^2\rangle k^3 = A_S(k/H_0)^{n-1}\begin{cases} 1 & \text{for} \quad kt_{\rm eq} < 1 \\ (kt_{\rm eq})^{-4}\cos^2(kt_{\rm eq}) & \text{for} \quad kt_{\rm eq} > 1 \,. \end{cases} \tag{2.151}$$

As we shall see in Chapter 3, inflation generically leads to a spectrum which is close to scale invariant[1] $n \simeq 1$. A formal definition of the spectrum, interpreting Ψ, or more precisely the amplitude A as a random variable is given in Section 2.6. Here we may just consider it as the square of the Fourier transform of Ψ and ignore the expectation value $\langle\cdots\rangle$.

Another interesting case (especially when discussing inflation) is the scalar field. There, as we shall see in Chapter 3, $\Pi = 0$, but in general $\Gamma \neq 0$ since $\delta p/\delta\rho \neq \dot{p}/\dot{\rho}$. Nevertheless, since this case again has only one dynamical degree of freedom, we

[1] The reason for the definition of n, such that $\langle|\Psi|^2\rangle k^3 \propto k^{n-1}$ is purely historical and not very logical, but as always, it is difficult to change conventions without leading to confusion. For compatibility with the literature we therefore keep this convention.

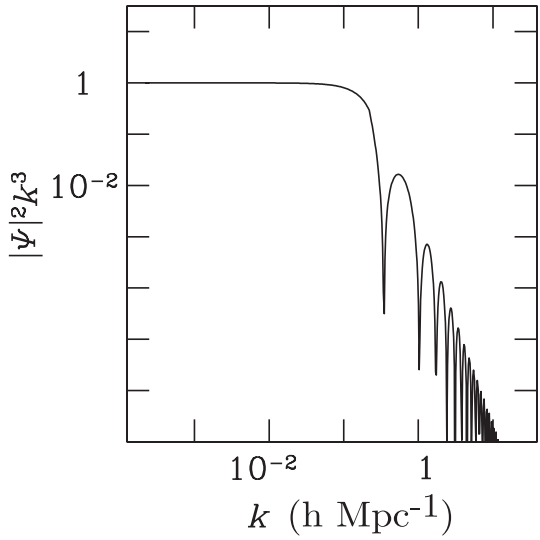


Fig. 2.1. The approximate form of the power spectrum $\langle|\Psi|^2\rangle k^3$ for a scale-invariant initial spectrum, $n = 1$ is plotted.

can express the perturbation equations in terms of one single second-order equation for Ψ. In Chapter 3 we shall find the following equation for a perturbed scalar field cosmology

$$\ddot{\Psi} + 3\mathcal{H}(1 + c_s^2)\dot{\Psi} + [(1 + 3c_s^2)(\mathcal{H}^2 - K) - (1 + 3w)(\mathcal{H}^2 + K) + k^2]\Psi = 0 \,. \tag{2.152}$$

The only difference between the perfect fluid and scalar field perturbation equation is that the latter is missing the factor c_s^2 in front of the oscillatory k^2 term. It is useful to define also the variable (Mukhanov *et al.*, 1992)

$$u = a[4\pi G(\mathcal{H}^2 - \dot{\mathcal{H}} + K)]^{-1/2}\Psi \,, \tag{2.153}$$

which satisfies the equation

$$\ddot{u} + (c_s^2 k^2 - \ddot{\theta}/\theta)u = 0 \,, \tag{2.154}$$

where

$$\theta = \frac{3\mathcal{H}}{2a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}} + K}} \,. \tag{2.155}$$

A second-order linear differential equation of the form (2.152) can always be transformed into one of the form of Eq. (2.154) by a suitable transformation of variables. We show this in Ex. 2.5.

In terms of the curvature variable

$$\zeta \equiv \frac{2(\mathcal{H}^{-1}\dot{\Psi} + \Psi)}{3(1+w)} + \Psi \,, \tag{2.156}$$

Eq. (2.145) is equivalent to

$$\dot{\zeta} = \frac{2}{3(1+w)\mathcal{H}} \left\{ \left[(2+3w+3c_s^2)K - c_s^2 k^2 \right] \Psi + \frac{1+3w}{2} \frac{K}{\mathcal{H}} \dot{\Psi} \right\} \,. \tag{2.157}$$

If K is negligible, this implies again that ζ is conserved on super-Hubble scales, $k/\mathcal{H} \ll 1$.

The evolution of ζ is closely related to the canonical variable v defined by

$$v = \frac{-a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}}}}{\sqrt{4\pi G} c_s \mathcal{H}} \zeta \,, \tag{2.158}$$

if $K = 0$. It satisfies the equation

$$\ddot{v} + (c_s^2 k^2 - \ddot{z}/z) v = 0 \,, \tag{2.159}$$

for

$$z = \frac{a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}}}}{c_s \mathcal{H}} \,. \tag{2.160}$$

The significance of the canonical variable v which has been introduced in Mukhanov *et al.* (1992) will be discussed in Chapter 3.

2.4 Simple examples

We first discuss two simple applications which are important to understand the CMB anisotropy spectrum.

2.4.1 The pure dust fluid for $K = 0, \Lambda = 0$

We assume the dust to have $w = c_s^2 = p = 0$ and $\Pi = \Gamma = 0$. Equation (2.146) then reduces to

$$\ddot{\Psi} + \frac{6}{t}\dot{\Psi} = 0 \,, \tag{2.161}$$

with the general solution,

$$\Psi = \Psi_0 + \Psi_1 \frac{1}{t^5} \,, \tag{2.162}$$

with arbitrary constants Ψ_0 and Ψ_1. Since the perturbations are supposed to be small initially, they cannot diverge for $t \to 0$, and we have therefore to choose the decaying mode, $\Psi_1 = 0$. Another way to argue is as follows: if the mode Ψ_1 has to be small

already at some early initial time $t_{\rm in}$, it will be much smaller later and may hence be neglected at late times. But also the Ψ_0 mode is only constant and not growing. This fact led Lifshitz, who was the first to analyse relativistic cosmological perturbations, to the conclusions that linear perturbations do not grow in a FL universe and cosmic structure cannot have evolved by gravitational instability (Lifshitz, 1946). However, the important point to note here is that, even if the gravitational potential remains constant, matter density fluctuations do grow on subhorizon scales and therefore inhomogeneities can evolve on scales that are smaller than the Hubble scale. To see this we consider the conservation equations (2.114), (2.105) and the Poisson equation (2.103). For the pure dust case, $w = c_s^2 = \Pi = \Gamma = 0$, they reduce to

$$\dot{D}_g = -kV \quad \text{(energy conservation)}\,, \tag{2.163}$$

$$\dot{V} + \mathcal{H}V = k\Psi \quad \text{(gravitational acceleration)}\,, \tag{2.164}$$

$$-\frac{2k^2}{3\mathcal{H}^2}\Psi = \left(D_g + 3\left(\Psi + \mathcal{H}k^{-1}V\right)\right) \quad \text{(Poisson)}\,, \tag{2.165}$$

where we have used the relation

$$D = D_g + 3(1+w)\left(\Phi + \mathcal{H}k^{-1}V\right)\,. \tag{2.166}$$

The Friedmann equation for dust gives $\mathcal{H} = 2/t$. Setting $kt = x$ and a prime $= d/dx$, the system (2.163)–(2.165) becomes

$$D_g' = -V\,, \tag{2.167}$$

$$V' + \frac{2}{x}V = \Psi\,, \tag{2.168}$$

$$\frac{6}{x^2}\left(D_g + 3\left(\Psi + \frac{2}{x}V\right)\right) = -\Psi\,. \tag{2.169}$$

We use (2.169) to eliminate Ψ and (2.167) to eliminate D_g, leading to

$$\left(18 + x^2\right)V'' + \left(\frac{72}{x} + 4x\right)V' - \left(\frac{72}{x^2} + 4\right)V = 0\,. \tag{2.170}$$

The general solution of Eq. (2.170) is

$$V = V_0 x + \frac{V_1}{x^4}\,. \tag{2.171}$$

The V_1 mode is the decaying mode (corresponding to Ψ_1) which we neglect. The perturbation variables are then given by

$$V = V_0 x\,, \tag{2.172}$$

$$D_g = -15V_0 - \frac{1}{2}V_0 x^2\,, \tag{2.173}$$

$$V_0 = \Psi_0/3\,. \tag{2.174}$$

We distinguish two regimes.
(*i*) Super-horizon, $x \ll 1$, where we have

$$V = \frac{1}{3}\Psi_0 x \,, \tag{2.175}$$

$$D_g = -5\Psi_0 \,, \tag{2.176}$$

$$\Psi = \Psi_0 \,. \tag{2.177}$$

Note that even though V is growing, it always remains much smaller than Ψ or D_g on super-horizon scales. Hence the largest fluctuations are of order Ψ, which is constant.
(*ii*) Subhorizon, $x \gg 1$, where the solution is dominated by the terms

$$V = \frac{1}{3}\Psi_0 x \,, \tag{2.178}$$

$$D_g = -\frac{1}{6}\Psi_0 x^2 \,, \tag{2.179}$$

$$\Psi = \Psi_0 = \text{constant} \,. \tag{2.180}$$

Note that for dust

$$D = D_g + 3\Psi + \frac{6}{x}V = -\frac{1}{6}\Psi_0 x^2 \,.$$

In the variable D the constant term has disappeared and we have $D \ll \Psi$ on super-horizon scales, $x \ll 1$.

On subhorizon scales, the density fluctuations grow like the scale factor $\propto x^2 \propto a$. Nevertheless, Lifshitz' conclusion (Lifshitz, 1946) that pure gravitational instability cannot be the cause of structure formation has some truth. If we start from tiny thermal fluctuations of the order of 10^{-35}, they can only grow to about 10^{-30} due to this mild, power law instability during the matter dominated regime. Or, to put it differently, if we want to form structure by gravitational instability, we need initial fluctuations of the order of at least 10^{-5}, much larger than thermal fluctuations. One possibility for creating such fluctuations is quantum particle production in the classical gravitational field during inflation. The rapid expansion of the Universe during inflation quickly expands microscopic scales at which quantum fluctuations are important to cosmological scales where these fluctuations are then 'frozen in' as classical perturbations in the energy density and the geometry. We will discuss the induced spectrum of fluctuations in Chapter 3.

2.4.2 The pure radiation fluid, $K = 0, \Lambda = 0$

In this limit we set $w = c_s^2 = \frac{1}{3}$ and $\Pi = \Gamma = 0$ so that $\Phi = \Psi$. We conclude from $\rho \propto a^{-4}$ that $a \propto t$. For radiation, the general solution (2.148) becomes

$$\Psi(x) = \frac{1}{x}\left[A j_1(x) + B y_1(x)\right] , \tag{2.181}$$

where we have set $x = kt/\sqrt{3} = c_s kt$ and used the fact that $a \propto x$. On super-horizon scales, $x \ll 1$, we have (see Appendix A4.3)

$$\Psi(x) \simeq \frac{A}{3} + \frac{B}{x^3} . \tag{2.182}$$

We assume that the perturbations have been initialized at some early time $x_{\rm in} \ll 1$ and that at this time the two modes have been comparable. If this is the case then $B \ll A$ and we may neglect the B-mode at later times, so that (see Abramowitz & Stegun, 1970)

$$\Psi(x) = \frac{A}{x} j_1(x) = A\left(\frac{\sin(x)}{x^3} - \frac{\cos(x)}{x^2}\right) . \tag{2.183}$$

To determine the density and velocity perturbations, we use the energy conservation and Poisson equations for radiation, with a prime denoting d/dx these become, for radiation,

$$D_g' = -\frac{4}{\sqrt{3}} V , \tag{2.184}$$

$$-2x^2\Psi = D_g + 4\Psi + \frac{4}{\sqrt{3}x} V . \tag{2.185}$$

Inserting the solution (2.183) for Ψ, we obtain

$$D_g = 2A\left[\cos(x) - \frac{2}{x}\sin(x)\right] , \tag{2.186}$$

$$V = -\frac{\sqrt{3}}{4} D_g' , \tag{2.187}$$

$$\Psi = -\frac{D_g + \frac{4}{\sqrt{3}x} V}{4 + 2x^2} . \tag{2.188}$$

In the **super-horizon regime**, $x \ll 1$, we obtain

$$\Psi = \frac{A}{3} , \quad D_g = -2A\left(1 + \frac{1}{6}x^2\right) , \quad V = \frac{A}{2\sqrt{3}} x . \tag{2.189}$$

On **subhorizon scales**, $x \gg 1$, we find oscillating solutions with constant amplitude and with frequency $k/\sqrt{3}$:

$$V = \frac{\sqrt{3}A}{2}\sin(x)\,, \tag{2.190}$$

$$D_g = 2A\cos(x)\,, \quad \Psi = -A\cos(x)/x^2\,. \tag{2.191}$$

The radiation fluid cannot simply 'collapse' under gravity. As in acoustic waves, the restoring force provided by the pressure leads to oscillations with constant amplitude. These are called the 'acoustic oscillations' of the radiation fluid. As we shall see in the next section, they are responsible for the acoustic peaks in the CMB fluctuation spectrum.

Also for radiation perturbations

$$D = -\frac{2A}{3}x^2 \ll \Psi$$

is small on super-horizon scales, $x \ll 1$.

The perturbation amplitude is given by the largest gauge-invariant perturbation variable. We conclude therefore that perturbations outside the Hubble horizon are frozen to first order. Once they enter the horizon they start to collapse, but pressure resists the gravitational force and the radiation fluid fluctuations oscillate at constant amplitude. The perturbations of the gravitational potential oscillate and decay like $1/a^2$ inside the horizon.

2.4.3 The mixed dust and radiation fluid for $K = 0, \Lambda = 0$

We now consider a mixed matter (also called 'dust' since we neglect its pressure) and radiation fluid with comparable perturbation amplitudes in the fluid variables. At early times we are in the radiation dominated era, and radiation perturbations will not be affected at all by the subdominant gravitational potential from matter fluctuations. As before, the radiation variables and the gravitational potential perform acoustic oscillations,

$$\Psi = \frac{A}{x}j_1(x) = A\left[\frac{\sin(x)}{x^3} - \frac{\cos(x)}{x^2}\right]\,, \tag{2.192}$$

$$D_{gr} = 2A\left[\cos(x) - \frac{2}{x}\sin(x)\right]\,, \tag{2.193}$$

$$V_r = \frac{\sqrt{3}A}{2}\left[\left(1 + \frac{2}{x^2}\right)\sin(x) - \frac{2}{x}\cos(x)\right]\,. \tag{2.194}$$

In the radiation era the matter equations become ($x = \frac{kt}{\sqrt{3}}$)

$$D'_{gm} + \sqrt{3}V_m = 0\,, \tag{2.195}$$

$$(aV_m)' = a\sqrt{3}\Psi = \frac{a\sqrt{3}A}{x}\, j_1(x)\,. \tag{2.196}$$

These equations can be solved simply by integration leading to

$$V_m = \frac{-\sqrt{3}A}{x}\, j_0(x) + V_1/x = -\sqrt{3}A\frac{\sin(x)}{x^2} + V_1/x\,, \tag{2.197}$$

$$D_{gm} = -3A\left[\frac{\sin(x)}{x} + \mathrm{Ci}(x) - \ln(x) + z_0\right] - \sqrt{3}V_1\ln(x)\,. \tag{2.198}$$

Here Ci is the integral cosine function defined by $\mathrm{Ci}(x) = \int_0^x \frac{1-\cos(z)}{z}\, dz$ (see Abramowitz & Stegun, 1970). The condition that V be small at very early times, $x \ll 1$ requires $V_1 = \sqrt{3}A$. The constant z_0 is an arbitrary integration constant. With this the above solutions become

$$V_m = \frac{\sqrt{3}A}{x}\left[1 - \frac{\sin(x)}{x}\right]\,, \tag{2.199}$$

$$D_{gm} = -3A\left[\frac{\sin(x)}{x} - \mathrm{Ci}(x) + z_0\right]\,. \tag{2.200}$$

On large scales, $x \ll 1$, we obtain the behaviour

$$\Psi = \frac{A}{3}\,, \tag{2.201}$$

$$D_{gr} = -2A\,, \tag{2.202}$$

$$V_r = \frac{A}{2\sqrt{3}}x\,, \tag{2.203}$$

$$V_m = \frac{A}{2\sqrt{3}}x\,, \tag{2.204}$$

$$D_{gm} = -3A(1 + z_0)\,. \tag{2.205}$$

The most natural condition to fix the constant z_0 is the requirement that at very early times perturbations are adiabatic, $\Gamma_{\rm tot} = \pi_L - (c_s^2/w)\delta = 0$. We use $\pi_L = \delta P_r/P_r = \delta\rho_r/\rho_r$ and

$$c_s^2/w = \frac{4}{R+3}\,, \qquad \text{where} \qquad R \equiv \frac{\rho_r}{\rho_m + \rho_r}\,.$$

Here we have used the fact that $P = P_r = \rho_r/3$ and $\rho_r \propto a^{-4}$, while $\rho_m \propto a^{-3}$. For the entropy production we then obtain

$$\Gamma_{\rm tot} = 4\frac{1-R}{R+3}\left(\frac{3}{4}\delta_r - \delta_m\right) , \tag{2.206}$$

so that $\Gamma_{\rm tot} = 0$ implies $\delta_m = \frac{3}{4}\delta_r$. According to the definition of D_g, Eq. (2.85) this is equivalent to $D_{gm} = (3/4)D_{gr}$. To achieve this we have to set $z_0 = -\frac{1}{2}$ so that

$$D_{gm} = -\frac{3}{2}A . \tag{2.207}$$

With this choice, perturbations are adiabatic on super-Hubble scales. But since D_{gm} and D_{gr} evolve differently on sub-Hubble scales, there clearly $\Gamma_{\rm tot} \neq 0$. We shall use the notion 'adiabatic' in the sense that the *initial conditions* are such that $\Gamma_{\rm tot}(t_{\rm in}) = 0$ for some early initial time $t_{\rm in}$ such that $kt_{\rm in} \ll 1$.

On sub-Hubble scales, $x \gg 1$, the radiation perturbations oscillate as in the ordinary radiation universe, but the matter perturbations grow logarithmically, $D_{gm} \simeq 3A\mathrm{Ci}(x) \simeq 3A\ln(x)$ for $x \gg 1$. This severe suppression of growth of matter perturbations during the radiation dominated era is called the 'Mészáros effect' (Mészáros, 1974). Physically, the reason for this suppression is that matter self-gravity $\propto 4\pi G\rho_m$ is too weak during the radiation dominated regime to overcome damping which (in the same units) is $\propto \mathcal{H}^2 \propto G\rho_r$. Neglecting self-gravity in the matter equation would yield $D_{gm} =$ constant, which is nearly correct.

We now go over to the matter dominated regime. There, the matter perturbations are not affected by radiation and behave as given in Eqs. (2.172)–(2.174),

$$\Psi = \Psi_0 , \tag{2.208}$$

$$V_m = \frac{1}{\sqrt{3}}\Psi_0 x , \tag{2.209}$$

$$D_{gm} = -5\Psi_0\left(1 + \frac{1}{10}x^2\right) . \tag{2.210}$$

Keeping in mind that $x = kt/\sqrt{3}$, these solutions correspond exactly to Eqs. (2.172)–(2.174). The radiation perturbation equations reduce to

$$D'_{gr} = -\frac{4}{\sqrt{3}}V_r , \tag{2.211}$$

$$D''_{gr} + D_{gr} = -8\Psi_0 , \tag{2.212}$$

with the general solution

$$D_{gr} = B\sin(x) + C\cos(x) - 8\Psi_0 ,$$

$$V_r = -\frac{\sqrt{3}}{4}\left(B\cos(x) - C\sin(x)\right) .$$

Requiring that these solutions be connected smoothly to the radiation dominated solutions fixes the constants B and C. Therefore, on large scales, $x \ll 1$, V has to grow like x which implies $B \equiv 0$. The constant C is then determined by the condition that the perturbations be adiabatic for $x \ll 1$. This implies

$$D_{gr} \simeq C - 8\Psi_0 = \frac{4}{3} D_{gm} = -\frac{20}{3}\Psi_0 \quad \text{so that} \quad C = \frac{4}{3}\Psi_0 \,. \tag{2.213}$$

This leads to the following solution for the radiation perturbations in the matter dominated era

$$D_{gr} = 4\Psi_0 \left(\frac{1}{3}\cos(x) - 2 \right) , \tag{2.214}$$

$$V_r = \frac{1}{\sqrt{3}}\Psi_0 \sin(x) \,. \tag{2.215}$$

These are the exact solutions for decoupled but adiabatic matter and radiation fluctuations in the matter dominated era. To connect them to the solutions in the radiation dominated era, we require that D_{gm} be continuous at the transition, $x = x_{\rm eq} = k t_{\rm eq}/\sqrt{3}$. This implies

$$\Psi_0 = A \begin{cases} \frac{3}{10} & \text{for} \quad x_{\rm eq} \ll 1 \\ 6\frac{\ln(x_{\rm eq})}{x_{\rm eq}^2} & \text{for} \quad x_{\rm eq} \gg 1 \,. \end{cases} \tag{2.216}$$

This approximation is of course relatively crude, since the radiation to matter transition is very gradual and not as abrupt as it is implemented here. It is also easy to see that we would not obtain exactly the same condition when requiring Ψ to be continuous at the transition. The main difference is that we do not obtain the logarithmic growth of the potential in the radiation dominated era from the continuity of Ψ. But this is clearly a failure since the log growth of D_{gm} leads to a larger gravitational potential in the matter era. For $x_{\rm eq} \simeq 1$ both approximations are bad and should be taken simply as order of magnitude estimates. More details on the coupled matter radiation system are found in Section 3.3. In Fig. 2.2 the exact solutions are plotted.

Instead of requiring adiabatic initial conditions one sometimes also requires $\Psi = \Phi = 0$ on super-Hubble scales. This is the so-called *iso-curvature* initial condition. We shall discuss it in Section 3.3.

2.5 Light-like geodesics and CMB anisotropies

After decoupling, $t > t_{\rm dec}$, photons follow to a good approximation light-like geodesics. The temperature shift of a Planck distribution of photons is equal to the energy shift of any given photon. The relative energy shift, red- or blue shift, is independent of the photon energy (gravity is 'achromatic').

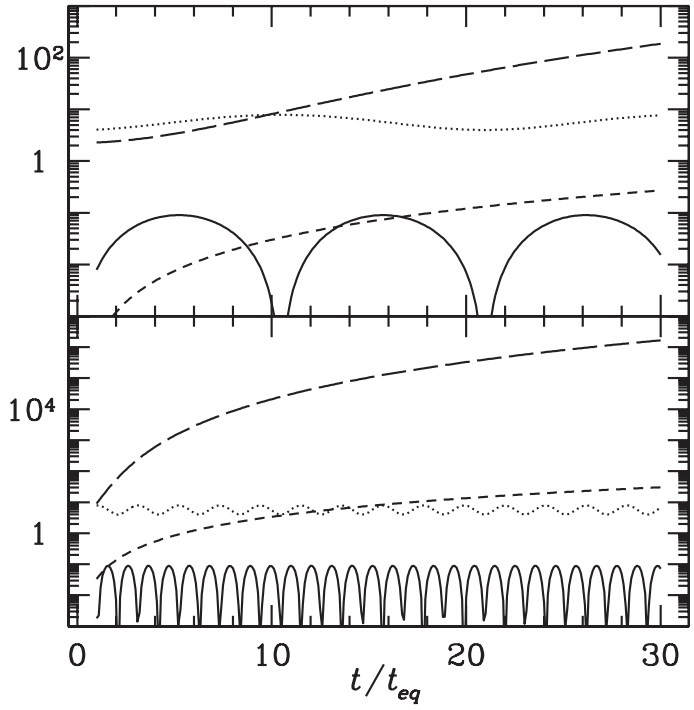


Fig. 2.2. The time evolution for $|D_{gm}|^2$ (long-dashed), $|D_{gr}|^2$ (dotted), $|V_m|^2$ (dashed) and $|V_r|^2$ (solid) is indicated as a function of $t/t_{\rm eq}$. The wave number in the top panel is $k_1 \ll 1/t_{\rm eq}$, while in the bottom panel $k_2 \gg 1/t_{\rm eq}$. Note that for a large wave number D_{gm} immediately starts growing and rapidly becomes much larger than D_{gr}, while for the small wave number (top panel) D_{gm} stays of the same order as D_{gr} until horizon entry, which is roughly at $t/t_{\rm eq} \sim 10$. After horizon entry D_{gm} starts growing while D_{gr} starts oscillating.

The unperturbed photon trajectory follows

$$(x^\mu(t)) \equiv \left(t, \int_t^{t_0} \mathbf{n}(t')\, dt' + \mathbf{x}_0\right),$$

where $\mathbf{x}_0$ is the photon position at time t_0 and $\mathbf{n}$ is the (parallel transported) photon direction. We determine the components of the photon momentum with respect to a geodesic basis $(\mathbf{e}_i)_{i=1}^3$ on the constant time hypersurfaces. We choose

$$\mathbf{e}_i = \begin{cases} \frac{\partial}{\partial x^i}\,, & \text{if } K = 0\,, \\ \epsilon_i\,, & \text{with } \gamma(\epsilon_i, \epsilon_j) = \delta_{ij} \text{ if } K \neq 0\,. \end{cases} \tag{2.217}$$

In other words, the vector fields ϵ_i form an orthonormal basis for the spatial metric γ_{ij}.

Our metric is of the form

$$d\tilde{s}^2 = a^2 ds^2 \,, \quad \text{with} \tag{2.218}$$

$$ds^2 = \left(\gamma_{\mu\nu} + h_{\mu\nu}\right) dx^\mu \, dx^\nu, \quad \gamma_{00} = -1, \; \gamma_{i0} = 0, \; \gamma_{ij} = \gamma_{ji} \,. \tag{2.219}$$

as before.

We make use of the fact that light-like geodesics are conformally invariant. More precisely, ds^2 and $d\tilde{s}^2$ have the same light-like geodesics, only the corresponding affine parameters are different. Let us denote the two affine parameters by λ and $\tilde{\lambda}$ respectively, and the tangent vectors to the unperturbed geodesic by

$$n = \frac{dx}{d\lambda} \,, \quad \tilde{n} = \frac{dx}{d\tilde{\lambda}} \;\; , \quad n^2 = \tilde{n}^2 = 0 \,, \;\; n^0 = 1 \,, \;\; \mathbf{n}^2 = 1 \,. \tag{2.220}$$

The photon 4-momentum p^μ is then given by $p^\mu = \omega n$, where ω is the constant energy of the photon moving in the metric ds^2. We have seen that in expanding space the photon momentum is redshifted. Actually, the components behave like $\tilde{n}^i \propto 1/a^2$ so that $\tilde{\mathbf{n}}^2 = a^2 \sum_i (\tilde{n}^i)^2 \propto 1/a^2$, hence we have to choose $\tilde{\lambda} = a^2\lambda$. As always for light-like geodesics, $\tilde{\lambda}$ and λ are only determined up to a multiplicative constant which we have fixed by the conditions $\mathbf{n}^2 = 1$ and $\tilde{\lambda} = a^2\lambda$.

Let us now introduce perturbations. We set $n^0 = 1 + \delta n^0$. The geodesic equation for the perturbed metric

$$ds^2 = (\gamma_{\mu\nu} + h_{\mu\nu}) \, dx^\mu \, dx^\nu \,, \tag{2.221}$$

yields, to first order,

$$\frac{d}{d\lambda} \delta n^\mu = -\delta\Gamma^\mu_{\alpha\beta} n^\alpha n^\beta \,. \tag{2.222}$$

For the energy shift, we have to determine δn^0. Since $g^{0\mu} = -\delta_{0\mu} + \text{first order}$, we obtain $\delta\Gamma^0_{\alpha\beta} = -\frac{1}{2}(h_{\alpha 0|\beta} + h_{\beta 0|\alpha} - \dot{h}_{\alpha\beta})$, so that

$$\frac{d}{d\lambda} \delta n^0 = h_{\alpha 0|\beta} n^\beta n^\alpha - \frac{1}{2} \dot{h}_{\alpha\beta} n^\alpha n^\beta \,. \tag{2.223}$$

Integrating this equation we use $h_{\alpha 0|\beta} n^\beta n^\alpha = \frac{d}{d\lambda}(h_{\alpha 0} n^\alpha)$, so that the change of n^0 between some initial time t_i and some final time t_f is given by

$$\delta n^0|_i^f = \left[h_{00} + h_{0j} n^j\right]_i^f - \frac{1}{2} \int_i^f \dot{h}_{\mu\nu} n^\mu n^\nu d\lambda \,. \tag{2.224}$$

The energy of a photon with 4-momentum $\tilde{p}^\mu$ as seen by an observer moving with 4-velocity $\tilde{u}$ is given by $E = -(\tilde{u} \cdot \tilde{p})$. Hence, the ratio of the energy of a photon

measured by some observer at t_f to the energy emitted at t_i is

$$\frac{E_f}{E_i} = \frac{(\tilde{n}\tilde{\cdot}\tilde{u})_f}{(\tilde{n}\tilde{\cdot}\tilde{u})_i} = \frac{a_i}{a_f}\frac{(n\cdot u)_f}{(n\cdot u)_i} , \tag{2.225}$$

where here $\tilde{\cdot}$ denotes the scalar product in an expanding universe, containing the factor a^2 and $\tilde{u}$ is the emitter and receiver 4-velocity in an expanding universe, $\tilde{u} = a^{-1}u$, while u_f and u_i are the 4-velocities of the observer and emitter respectively in the non-expanding conformally related geometry given by

$$u = (1 - A)\partial_t + v^i \mathbf{e}_i = a\tilde{u} . \tag{2.226}$$

Together with $\tilde{n} = a^{-2}n$ this implies the result (2.225). The ratio $a_i/a_f = T_i/T_f$ is the usual (unperturbed) redshift which relates n and $\tilde{n}$. An observer measuring a temperature T_0 receives photons that were emitted at the time $t_{\rm dec}$ of decoupling of matter and radiation, at the fixed temperature $T_{\rm dec}$. In first-order perturbation theory, we find the following relation between the unperturbed temperatures T_f, T_i, the true temperatures $T_0 = T_f + \delta T_f$, $T_{\rm dec} = T_i + \delta T_i$, and the photon density perturbation:

$$\frac{a_i}{a_f} = \frac{T_f}{T_i} = \frac{T_0}{T_{\rm dec}}\left(1 - \frac{\delta T_f}{T_f} + \frac{\delta T_i}{T_i}\right) = \frac{T_0}{T_{\rm dec}}\left(1 - \frac{1}{4}\delta_r|_i^f\right) , \tag{2.227}$$

where δ_r is the intrinsic density perturbation in the radiation and we have used $\rho_r \propto T^4$ in the last equality. Inserting the above equation and Eq. (2.224) into Eq. (2.225), and using Eq. (2.28) for the definition of $h_{\mu\nu}$, as well as Eqs. (2.55), (2.56), (2.86) and (2.83) one finds, after integration by parts, the following result for scalar perturbations:

$$\frac{E_f}{E_i} = \frac{T_0}{T_{\rm dec}}\left\{1 - \left[\frac{1}{4}D_g^{(r)} + V_j^{(b)}n^j + \Psi + \Phi\right]_i^f + \int_i^f (\dot{\Psi} + \dot{\Phi})\, d\lambda\right\} . \tag{2.228}$$

Here $D_g^{(r)}$ denotes the density perturbation in the radiation fluid, and $V^{(b)}$ is the peculiar velocity of the baryonic matter component (the emitter and observer of radiation).

Evaluating Eq. (2.228) at final time t_0 (today) and initial time $t_{\rm dec}$, we obtain the temperature difference of photons coming from different directions $\mathbf{n}_1$ and $\mathbf{n}_2$

$$\frac{\Delta T}{T} \equiv \frac{\Delta T(\mathbf{n}_1)}{T} - \frac{\Delta T(\mathbf{n}_2)}{T} \equiv \frac{E_f}{E_i}(\mathbf{n}_1) - \frac{E_f}{E_i}(\mathbf{n}_2) . \tag{2.229}$$

Direction-independent contributions to E_f/E_i do not enter in this difference.

The largest contribution to $\Delta T/T$ is the dipole term, $V_j^{(b)}(t_0)n^j$ which simply describes our motion with respect to the emission surface. Its amplitude is about

1.2×10^{-3} and it has been measured so accurately that even the yearly variation due to the motion of the Earth around the sun has been detected.

For the higher multipoles (polynomials in n^j of degree 2 and higher) we can set

$$\frac{\Delta T(\mathbf{n})}{T} = \left[\frac{1}{4}D_g^{(r)} + V_j^{(b)}n^j + \Psi + \Phi\right](t_{\rm dec}, \mathbf{x}_{\rm dec}) + \int_{t_{\rm dec}}^{t_0}(\dot{\Psi} + \dot{\Phi})(t, \mathbf{x}(t))\, dt \ , \tag{2.230}$$

where $\mathbf{x}(t)$ is the unperturbed photon position at time t for an observer at $\mathbf{x}_0$, and $\mathbf{x}_{\rm dec} = \mathbf{x}(t_{\rm dec})$ (if $K = 0$ we simply have $\mathbf{x}(t) = \mathbf{x}_0 - (t_0 - t)\mathbf{n}$). The first term in Eq. (2.230) is the one we have discussed in the previous section. It describes the intrinsic inhomogeneities of the radiation density on the surface of last scattering, due to acoustic oscillations prior to decoupling, see Eq. (2.193). Depending on the initial conditions, it can also contribute significantly on super-horizon scales. This is especially important in the case of adiabatic initial conditions. As we have seen in Eq. (2.213), in a dust + radiation universe with $\Omega = 1$, adiabatic initial conditions imply $D_g^{(r)}(k, t) = -\frac{20}{3}\Psi(k, t)$ and $V^{(b)} = V^{(r)} \ll D_g^{(r)}$ when $kt \ll 1$. With $\Phi = \Psi$ the square bracket of Eq. (2.230) therefore gives for adiabatic perturbations

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{\rm (OSW)}_{\rm adiabatic} = \frac{1}{3}\Psi(t_{\rm dec}, \mathbf{x}_{\rm dec}) \ ,$$

on super-horizon scales. The contribution to $\Delta T/T$ from the last scattering surface on very large scales is called the 'ordinary Sachs–Wolfe effect' (OSW). It was derived for the first time by Sachs and Wolfe (1967). For iso-curvature perturbations, the initial conditions require $D_g^{(r)}(k, t) \to 0$ for $t \to 0$ so that the contribution of $D_g^{(r)}$ to the ordinary Sachs–Wolfe effect can be neglected,

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{\rm (OSW)}_{\rm iso\text{-}curvature} = 2\Psi(t_{\rm dec}, \mathbf{x}_{\rm dec}) \ .$$

The second term in (2.230) describes the relative motion of emitter and observer. This is the Doppler contribution to the CMB anisotropies. It appears on the same angular scales as the acoustic term; we call the sum of the acoustic and Doppler contributions 'acoustic peaks'.

The integral in Eq. (2.230) accounts for the red- or blue shifts caused by the time dependence of the gravitational potential along the path of the photon, and represents the so-called integrated Sachs–Wolfe (ISW) effect. In a $\Omega = 1$, pure dust universe, as we have seen, the Bardeen potentials are constant and there is no integrated Sachs–Wolfe effect; the blue shift which the photons acquire by falling into a gravitational potential is exactly cancelled by the redshift induced by climbing

out of it. This is no longer true in a universe with substantial radiation contribution, curvature, or a cosmological constant. The sum of the ordinary Sachs–Wolfe term and the integral is the full Sachs–Wolfe contribution.

For **vector** perturbations $\delta^{(r)}$ and A vanish and Eq. (2.225) leads to

$$\left(\frac{E_f}{E_i}\right)^{(V)} = \frac{a_i}{a_f}\left[1 - V_j^{(b)} n^j\Big|_i^f + \int_i^f \sigma_j n^j \, d\lambda\right] . \tag{2.231}$$

We obtain a Doppler term and a gravitational contribution. For **tensor** perturbations, i.e., gravitational waves, only the gravitational part remains:

$$\left(\frac{E_f}{E_i}\right)^{(T)} = \frac{a_i}{a_f}\left[1 - \int_i^f \dot{H}_{lj} n^l n^j d\lambda\right] . \tag{2.232}$$

Equations (2.228), (2.231) and (2.232) are the manifestly gauge-invariant results for the energy shift of photons due to scalar, vector and tensor perturbations. Disregarding again the dipole contribution due to our proper motion, Eqs. (2.231) and (2.232) imply the vector and tensor temperature fluctuations

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{(V)} = V_j^{(b)}(t_{\rm dec}, \mathbf{x}_{\rm dec}) n^j + \int_i^f \sigma_j(t, \mathbf{x}(t)) n^j \, d\lambda \, , \tag{2.233}$$

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{(T)} = -\int_i^f \dot{H}_{lj}(t, \mathbf{x}(t)) n^l n^j \, d\lambda \, . \tag{2.234}$$

Note that for models where initial fluctuations have been laid down in the very early universe, vector perturbations are irrelevant as we have already pointed out. In this sense Eq. (2.233) is here mainly for completeness. However, in models where perturbations are sourced by some inherently inhomogeneous component (e.g. topological defects, see Durrer *et al.* (2002)) vector perturbations can be important.

2.6 Power spectra

2.6.1 Generics

The quantities that we can determine from a given model are usually not the precise values of perturbation variables as $\Psi(\mathbf{k}, t)$, but only expectation values like $\langle|\Psi(\mathbf{k}, t) \cdot \Psi^*(\mathbf{k}', t)\rangle$. In different realizations, e.g., of the same inflationary model, the 'phases' $\alpha(\mathbf{k}, t)$ given by $\Psi(\mathbf{k}, t) = \exp(i\alpha(\mathbf{k}))|\Psi(k, t)|$ are different. They are random variables. If we assume that the random process which generates the fluctuations Ψ is stochastically homogeneous and isotropic, these phases have a vanishing 2-point correlator for different values of $\mathbf{k}$. However, the quantity which we can

calculate for a given model and which then has to be compared with observations is the power spectrum, defined below. Power spectra are the 'harmonic transforms' of the 2-point correlation functions.[2] If the perturbations of the model under consideration are Gaussian, a relatively generic prediction from inflationary models, then the 2-point functions and therefore the power spectra contain the full statistical information of the model.

There is one additional problem to consider: one can never 'measure' expectation values. We have only one Universe, i.e., one realization of the stochastic process which generates the fluctuations at our disposal for observations. The best we can do when we want to determine the mean square fluctuation on a given scale λ is to average over many disjoint patches of size λ, assuming that this spatial averaging corresponds to an ensemble averaging; a type of 'ergodic hypothesis'. This works well as long as the scale λ is much smaller than the Hubble horizon, the size of the observable Universe. For $\lambda \sim \mathcal{O}(H_0^{-1})$ we can no longer average over many independent volumes and the value measured could be quite far from the ensemble average. This problem is known under the name 'cosmic variance' and we shall come back to it in Chapter 6. More details about the formal aspects of power spectra can be found in Appendix 7.

For an arbitrary scalar variable X in position space, we define the power spectrum in Fourier space by

$$\left\langle X\left(\mathbf{k}, t_0\right) X^*\left(\mathbf{k}', t_0\right)\right\rangle = (2\pi)^3 \delta(\mathbf{k}-\mathbf{k}') P_X(k)\,. \tag{2.235}$$

In flat space, $K = 0$, the function $X(\mathbf{k})$ is the ordinary Fourier transform of $X(\mathbf{x})$. If $K \neq 0$ the situation is more complicated. Then $X(\mathbf{k})$ represents an expansion of $X(\mathbf{x})$ in terms of eigenfunctions of the Laplacian and in the case $K > 0$ the Dirac δ-function has to be replaced by a discrete Kronecker δ.

The $\langle\ \rangle$ indicates a statistical average, ensemble average, over 'random initial conditions' in a given model. We assume that no point in space is preferred, in other words that $X(\mathbf{x})$ and any other stochastic field which we consider has the same distribution in every point $\mathbf{x}$. Such random fields are called 'statistically homogeneous' (or stationary). We further assume that the distribution of $X(\mathbf{x})$ has no preferred direction. This means that the random field X is statistically isotropic. These properties imply that the Fourier transform of the 2-point function is diagonal, i.e., they explain the factor $\delta(\mathbf{k}-\mathbf{k}')$ in Eq. (2.235) (see Ex. 2.4).

[2] The 'harmonic transform' in usual flat space is simply the Fourier transform. In curved space it is the expansion in terms of eigenfunctions of the Laplacian on that space, e.g., on the sphere it corresponds to the expansion in terms of spherical harmonics.

2.6.2 The matter power spectrum

Let us first consider the power spectrum of dark matter, $P_D(k)$, which is defined by

$$\left\langle D_{gm}\left(\mathbf{k}, t_0\right) D^*_{gm}\left(\mathbf{k}', t_0\right)\right\rangle = P_D(k)(2\pi)^3\delta(\mathbf{k}-\mathbf{k}') \,. \tag{2.236}$$

$P_D(k)$ is usually compared with the observed power spectrum of the galaxy distribution. This is clearly problematic since it is by no means evident what the relation between these two spectra should be. This problem is known under the name of 'biasing' and it is very often simply assumed that the dark matter and galaxy power spectra differ only by a constant factor. The hope is also that on sufficiently large scales, since the evolution of both, galaxies and dark matter is governed by gravity, their power spectra should not differ much. This hope seems to be reasonably justified. In Tegmark *et al.* (2004) it is found that the observed galaxy power spectrum and the matter power spectrum inferred from the observation of CMB anisotropies differ only by about 10% on large scales.

The power spectrum of velocity perturbations satisfies the relation

$$\left\langle V_j\left(\mathbf{k}, t_0\right) V_i^*\left(\mathbf{k}', t_0\right)\right\rangle = Q_j^{(S)}(\mathbf{k}) Q_i^{(S)*}(\mathbf{k}') P_V(k)(2\pi)^3\delta(\mathbf{k}-\mathbf{k}') \,, \tag{2.237}$$

$$P_V(k) \simeq H_0^2 \Omega_m^{1.2} P_D(k) k^{-2} \,. \tag{2.238}$$

For $\simeq$ we have used that $|kV(t_0)| = \dot{D}_g^{(m)}(t_0) \sim H_0 \Omega_m^{0.6} D_g$ on subhorizon scales (see e.g., Peebles, 1993).

2.6.3 The CMB power spectrum

Definition

The spectrum that we are most interested in and which can be both, measured and calculated to the best accuracy is the CMB anisotropy power spectrum. It is defined as follows: $\Delta T/T$ is a function of position $\mathbf{x}$, time t and photon direction $\mathbf{n}$. Here, $\mathbf{x} = \mathbf{x}_0$ and now, $t = t_0$, $\Delta T/T$ is a function on the sphere, $\mathbf{n} \in \mathbb{S}^2$. We develop it in terms of spherical harmonics, $Y_{\ell m}$s. We will often suppress the arguments t_0 and $\mathbf{x}_0$ in the following calculations. Since our fields are statistically homogeneous, averages over an ensemble of realizations (expectation values) are independent of position. Furthermore, we assume that the process generating the initial perturbations is statistically isotropic. This means that the distribution of $\Delta T/T(\mathbf{n})$ is the same for all directions $\mathbf{n}$. As for the Fourier transforms of random fields in space, this implies

that the harmonic transform of $\Delta T/T$ is diagonal. In other words, the off-diagonal correlators of the expansion coefficients $a_{\ell m}$ vanish and we have

$$\frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, t_0) = \sum_{\ell,m} a_{\ell m}(\mathbf{x}_0) Y_{\ell m}(\mathbf{n}), \quad \left\langle a_{\ell m} \cdot a^*_{\ell' m'} \right\rangle = \delta_{\ell\ell'}\delta_{mm'} C_\ell \,. \tag{2.239}$$

The C_ℓs are the CMB power spectrum.

The 2-point correlation function is related to the C_ℓs by

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{n}) \frac{\Delta T}{T}(\mathbf{n}') \right\rangle_{\mathbf{n}\cdot\mathbf{n}'=\mu} \\
&\quad = \sum_{\ell,\ell',m,m'} \left\langle a_{\ell m} \cdot a^*_{\ell' m'} \right\rangle Y_{\ell m}(\mathbf{n}) Y^*_{\ell' m'}(\mathbf{n}') \\
&\quad = \sum_\ell C_\ell \underbrace{\sum_{m=-\ell}^{\ell} Y_{\ell m}(\mathbf{n}) Y^*_{\ell m}(\mathbf{n}')}_{\frac{2\ell+1}{4\pi} P_\ell(\mathbf{n}\cdot\mathbf{n}')} \\
&\quad = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\mu) \,,
\end{aligned} \tag{2.240}$$

where we have used the addition theorem of spherical harmonics for the last equality; the P_ℓs are the Legendre polynomials (see Appendices A4.2.3 and A4.1).

Clearly the a_{lm}s from scalar, vector and tensor perturbations are uncorrelated,

$$\left\langle a^{(S)}_{\ell m} a^{(V)}_{\ell' m'} \right\rangle = \left\langle a^{(S)}_{\ell m} a^{(T)}_{\ell' m'} \right\rangle = \left\langle a^{(V)}_{\ell m} a^{(T)}_{\ell' m'} \right\rangle = 0 \,. \tag{2.241}$$

Since vector perturbations decay, their contributions, the $C^{(V)}_\ell$, are negligible in models where initial perturbations have been laid down very early, e.g., after an inflationary period. Tensor perturbations are constant on super-horizon scales and perform damped oscillations once they enter the horizon.

Scalar perturbations

Let us first discuss in somewhat more detail scalar perturbations. We specialize to the case $K = 0$ for simplicity. We suppose the initial perturbations to be given by a spectrum of the form

$$\left\langle \Psi(\mathbf{k}) \Psi^*(\mathbf{k}') \right\rangle k^3 = (2\pi)^3 k^3 P_\Psi(k) \delta(\mathbf{k} - \mathbf{k}') = (2\pi)^3 A_S (k t_0)^{n-1} \delta(\mathbf{k} - \mathbf{k}') \,. \tag{2.242}$$

We multiply by the constant t_0^{n-1}, the actual comoving size of the horizon, in order to keep A_S dimensionless for all values of n. The number n is called the spectral index. A_S then represents the amplitude of metric perturbations at horizon scale today, $k = 1/t_0$.

As we have seen in the previous section, the dominant contribution on *super-horizon scales* (neglecting the integrated Sachs–Wolfe effect $\int \dot{\Phi} + \dot{\Psi}$) is the ordinary Sachs–Wolfe effect, OSW, which for adiabatic perturbations is given by

$$\frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, t_0) \simeq \frac{1}{3}\Psi(x_{\rm dec}, t_{\rm dec}) \,. \tag{2.243}$$

Since $\mathbf{x}_{\rm dec} = \mathbf{x}_0 + \mathbf{n}(t_0 - t_{\rm dec})$, the Fourier transform of (2.243) gives

$$\frac{\Delta T}{T}(\mathbf{k}, \mathbf{n}, t_0) = \frac{1}{3}\Psi(\mathbf{k}, t_{\rm dec}) \cdot e^{i\mathbf{k}\mathbf{n}(t_0 - t_{\rm dec})} \,. \tag{2.244}$$

Using the decomposition (see Appendix A4.3)

$$e^{i\mathbf{k}\mathbf{n}(t_0 - t_{\rm dec})} = \sum_{\ell=0}^{\infty}(2\ell+1)i^\ell j_\ell(k(t_0 - t_{\rm dec}))P_\ell(\widehat{\mathbf{k}}\cdot\mathbf{n}) \,,$$

where j_ℓ are the spherical Bessel functions, we obtain ($k = |\mathbf{k}|$, $\hat{\mathbf{k}} = \mathbf{k}/k$)

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, t_0)\frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}', t_0)\right\rangle \qquad (2.245)\\
&\quad = \frac{1}{(2\pi)^6}\int d^3k\, d^3k'\, e^{i\mathbf{x}_0\cdot(\mathbf{k}-\mathbf{k}')}\left\langle \frac{\Delta T}{T}(\mathbf{k}, \mathbf{n}, t_0)\left(\frac{\Delta T}{T}\right)^*(\mathbf{k}', \mathbf{n}', t_0)\right\rangle\\
&\quad \simeq \frac{1}{(2\pi)^6 9}\int d^3k d^3k' e^{i\mathbf{x}_0\cdot(\mathbf{k}-\mathbf{k}')}\left\langle \Psi(\mathbf{k})\Psi^*(\mathbf{k}')\right\rangle \sum_{\ell,\ell'=0}^{\infty}(2\ell+1)(2\ell'+1)i^{\ell-\ell'}\\
&\qquad \cdot j_\ell(k(t_0 - t_{\rm dec}))j_{\ell'}(k'(t_0 - t_{\rm dec}))P_\ell(\hat{\mathbf{k}}\cdot\mathbf{n})\cdot P_{\ell'}(\hat{\mathbf{k}}'\cdot\mathbf{n}')\\
&\quad = \frac{1}{(2\pi)^3 9}\int d^3k\, P_\Psi(k)\sum_{\ell,\ell'=0}^{\infty}(2\ell+1)(2\ell'+1)i^{\ell-\ell'}\\
&\qquad \cdot\, j_\ell(k(t_0 - t_{\rm dec}))j_{\ell'}(k'(t_0 - t_{\rm dec}))P_\ell(\hat{\mathbf{k}}\cdot\mathbf{n})\cdot P_{\ell'}(\hat{\mathbf{k}}\cdot\mathbf{n}') \,. \qquad (2.246)
\end{aligned}$$

In the first equals sign we have used the unitarity of the Fourier transformation. Inserting $P_\ell(\hat{\mathbf{k}}\mathbf{n}) = \frac{4\pi}{2\ell+1}\sum_m Y^*_{\ell m}(\hat{\mathbf{k}})Y_{\ell m}(\mathbf{n})$ and $P_{\ell'}(\hat{\mathbf{k}}\mathbf{n}') = \frac{4\pi}{2\ell'+1}$ $\sum_{m'} Y^*_{\ell' m'}(\hat{\mathbf{k}})Y_{\ell' m'}(\mathbf{n}')$, integration over the directions $d\Omega_{\hat{k}}$ gives $\delta_{\ell\ell'}\delta_{mm'}\sum_m Y^*_{\ell m}(\mathbf{n})Y_{\ell m}(\mathbf{n}')$.

Also using $\sum_m Y^*_{\ell m}(\mathbf{n})Y_{\ell m}(\mathbf{n}') = \frac{2\ell+1}{4\pi}P_\ell(\mu)$, where $\mu = \mathbf{n}\cdot\mathbf{n}'$, we find

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, t_0)\frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}', t_0)\right\rangle_{\mathbf{n}\mathbf{n}'=\mu}\\
&\quad \simeq \sum_\ell \frac{2\ell+1}{4\pi}P_\ell(\mu)\frac{2}{\pi}\int \frac{dk}{k}\frac{1}{9}P_\Psi(k)k^3 j_\ell^2(k(t_0 - t_{\rm dec})) \,. \qquad (2.247)
\end{aligned}$$

Comparing this equation with Eq. (2.240) we obtain for *adiabatic perturbations* on scales $2 \leq \ell \ll \chi(t_0 - t_{\rm dec})/t_{\rm dec} \sim 100$:

$$C_\ell^{\rm (SW)} \simeq C_\ell^{\rm (OSW)} \simeq \frac{2}{9\pi} \int_0^\infty \frac{dk}{k} P_\Psi(k) k^3 j_\ell^2 \left(k\,(t_0 - t_{\rm dec})\right) \,. \tag{2.248}$$

The function $j_\ell^2\,(k(t_0 - t_{\rm dec}))$ peaks roughly at $k\,(t_0 - t_{\rm dec}) \simeq kt_0 \simeq \ell$. If Ψ is a pure power law on large scales, $kt_{\rm dec} \lesssim 1$ as in Eq. (2.242) and we set $k(t_0 - t_{\rm dec}) \sim kt_0$, the integral (2.248) can be performed analytically. For the ansatz (2.242), using the integral (A4.102) one finds

$$C_\ell^{\rm (SW)} = \frac{A_S}{9} \frac{\Gamma(3-n)\Gamma(\ell - \frac{1}{2} + \frac{n}{2})}{2^{3-n}\Gamma^2(2 - \frac{n}{2})\Gamma(\ell + \frac{5}{2} - \frac{n}{2})} \quad \text{for } -3 < n < 3 \,. \tag{2.249}$$

Of special interest is the *scale-invariant* or Harrison–Zel'dovich (HZ) spectrum, $n = 1$ (see Chapter 3). It leads to

$$\ell(\ell+1) C_\ell^{\rm (SW)} = \frac{A_S}{9\pi} \simeq \left\langle \left(\frac{\Delta T}{T}(\vartheta_\ell) \right)^2 \right\rangle \,, \quad \vartheta_\ell \equiv \pi/\ell \,. \tag{2.250}$$

This is precisely (within the accuracy of the experiment) the behaviour observed by the DMR (differential microwave radiometer) experiment aboard the satellite COBE (Smoot *et al.*, 1992) and more precisely with the WMAP (Wilkinson microwave anisotropy probe) satellite (Bennett *et al.*, 2003), $n = 0.95 \pm 0.02$ (see Table 6.1) .

As we shall see in Chapter 3, inflationary models predict very generically a HZ spectrum (up to small corrections). The DMR discovery has therefore been regarded as a great success, if not a proof, of inflation. There are, however, other models such as topological defects (see Section 6.8, or for more details Durrer *et al.*, 2002), or certain string cosmology models (Durrer *et al.*, 1999) which also predict scale-invariant, i.e., Harrison–Zel'dovich spectra of fluctuations. These models are outside the class investigated here, since in them perturbations are induced by seeds which evolve non-linearly in time. They are not simply layed down as initial conditions for the fluid perturbations but typically affect the perturbations of a given wavelength until it crosses the Hubble scale. This generically leads to iso-curvature perturbations which are not favoured by present data. We therefore investigate such models only briefly in Section 6.8.

For iso-curvature perturbations, the main contribution on large scales comes from the integrated Sachs–Wolfe effect and (2.248) is replaced by

$$C_\ell^{\rm (ISW)} \simeq \frac{8}{\pi} \int \frac{dk}{k} k^3 \left\langle \left| \int_{t_{\rm dec}}^{t_0} \dot{\Psi}(k,t) j_\ell^2(k(t_0 - t))\, dt \right|^2 \right\rangle . \tag{2.251}$$

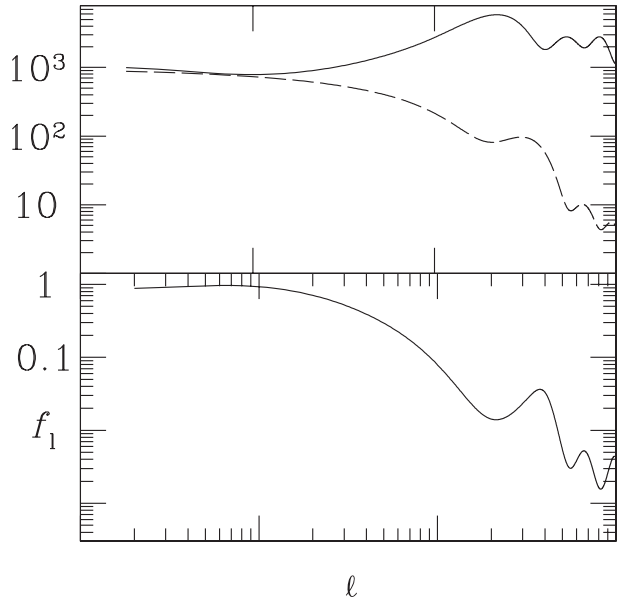


Fig. 2.3. Examples of COBE normalized adiabatic (solid line) and iso-curvature (dashed line) CMB anisotropy spectra, $\ell(\ell+1)C_\ell/(2\pi)$ in units of $(\mu K)^2$ are shown on the top panel. In the bottom panel the ratio of the iso-curvature to adiabatic temperature fluctuations is plotted.

Inside the horizon Ψ is roughly constant (matter dominated). Using the ansatz (2.242) for Ψ inside the horizon and setting the integral in (2.251) $\sim 2\Psi(k, t = 1/k) j_\ell^2(kt_0)$, we obtain again (2.249), but with $A_S^2/9$ replaced by $4A_S^2$. For a fixed amplitude A_S of perturbations, the Sachs–Wolfe temperature anisotropies coming from iso-curvature perturbations are therefore about six times larger than those coming from adiabatic perturbations (see Fig. 2.3).

On smaller scales, $\ell \gtrsim 100$, the contribution to $\Delta T/T$ is dominated by acoustic oscillations, the first two terms in Eq. (2.230). Instead of (2.251) we then obtain

$$C_\ell^{(AC)} \simeq \frac{2}{\pi}\int_0^\infty \frac{dk}{k} k^3 \left\langle \left| \frac{1}{4} D_r(k, t_{\rm dec}) j_\ell(kt_0) + V^{(r)}(k, t_{\rm dec}) j_\ell'(kt_0) \right|^2 \right\rangle . \tag{2.252}$$

To remove the SW contribution from $D_g^{(r)}$ we have simply replaced it by D_r which is much smaller than Ψ on super-horizon scales and therefore does not contribute to the SW terms. On subhorizon scales $D_r \simeq D_g^{(r)}$ and V_r are oscillating like sine or cosine waves depending on the initial conditions. Correspondingly the $C_\ell^{(AC)}$ will show peaks and minima. For adiabatic initial conditions $D_g^{(r)}$ and therefore D_r also

oscillates like a cosine. Its minima and maxima are at $k_n t_{\rm dec}/\sqrt{3} = n\pi$. Odd values of n correspond to maxima, 'contraction peaks', while even numbers are minima, 'expansion peaks'.

These are the 'acoustic peaks' of the CMB anisotropies. Sometimes they are misleadingly called 'Doppler peaks' referring to an old misconception that the peaks are due to the velocity term in the above formula. Actually the contrary is true. At maxima and minima of the density contrast, the velocity (being proportional to the derivative of the density) nearly vanishes. We shall therefore consistently call the CMB peak structure 'acoustic peaks'.

The angle θ_n, which subtends the scale $\lambda_n = \pi/k_n$ at the last scattering surface, is determined by the angular diameter distance to the last scattering surface, $d_A(t_{\rm dec})$ via the relation $\theta_n = \lambda_n/d_A(t_{\rm dec})$. Expanding the temperature anisotropies in spherical harmonics, the angular scale θ_n corresponds (roughly) to the harmonic number

$$\ell_n \simeq \pi/\theta_n = \pi d_A(t_{\rm dec})/\lambda_n = d_A(t_{\rm dec})k_n = n\sqrt{3}\pi d_A(t_{\rm dec})/t_{\rm dec}\ . \tag{2.253}$$

For a flat matter dominated universe $d_A(t_{\rm dec}) \simeq t_0$ leading to $\ell_n \simeq 180n$ (see Ex. 2.7). This crude approximation deviates by about 15% from the precise numerical value, which not only depends, with d_A, strongly on curvature but also on the Hubble parameter and on the cosmological constant. Furthermore, the peak positions depend on the sound speed of the radiation–baryon plasma which we have simply set to $c_s = 1/\sqrt{3}$ in this approximation. We shall discuss this parameter dependence of the peak positions in detail in Chapter 6. Note, however, that the position of the first peak differs significantly for the iso-curvature mode, for which $D_g^{(r)}$ oscillates like a sine. For generic initial conditions, we would expect a mixture of the sine and cosine modes which leads to a displacement of the first peak. The observed CMB anisotropies are consistent with a purely adiabatic mode and require, at least, that the adiabatic mode dominates (Bucher *et al.*, 2001; Trotta, 2006).

For a flat universe, $\Omega = 1$, the nth peak therefore is placed at

$$\ell_n \simeq k_n t_0 \cong n\pi\sqrt{3}\frac{t_0}{t_{\rm dec}}\ . \tag{2.254}$$

For a flat matter dominated universe we have $\frac{t_0}{t_{\rm dec}} \sim \sqrt{z_{\rm dec}} \sim 33.2$ which yields $\ell_1 \sim 180$. Here we have used $z_{\rm dec} \sim 1100$ (see Section 1.3). This approximation is not very good since the Universe is not very well matter dominated at $t_{\rm dec}$. A somewhat more accurate estimate (Ex. 2.7) gives $\ell_1 \sim 220$, in good agreement with the numerical value. Subsequent peaks are then given by $\ell_n = n\ell_1$.

Our discussion is only valid in flat space. In curved space the exponentials $\exp(ik(t_0 - t_{\rm dec}))$ have to be replaced with the harmonics of the curved spaces. For the positions of the peaks, this corresponds to replacing $k_n t_0$ by $k_n \chi(t_0)$, hence

replacing t_0 by the comoving angular diameter distance to the last scattering surface. Instead of Eq. (2.254) we then obtain the following approximate relation for the peak positions,

$$\ell_n \sim n\pi\sqrt{3}\frac{\chi(t_0)}{t_{\rm dec}} \ . \tag{2.255}$$

For values of Ω close to unity this scales like $1/\sqrt{\Omega}$ (see Section 1.2).

On very small scales the acoustic peaks are damped by the photon diffusion which takes place during the recombination process. This effect will be discussed with the Boltzmann equation approach in Chapter 4.

Tensor perturbations

For gravitational waves (which are tensor fluctuations), a formula analogous to (2.249) can be derived (see Appendix 8),

$$C_\ell^{(T)} = \frac{2}{\pi}\int dk\, k^2 \left|\int_{t_{\rm dec}}^{t_0} dt\ \dot{H}(t,k)\frac{j_\ell(k(t_0-t))}{(k(t_0-t))^2}\right|^2 \frac{(\ell+2)!}{(\ell-2)!} \ . \tag{2.256}$$

To a very crude approximation we may assume $\dot{H}^{(T)} = 0$ on super-horizon scales and $\int dt\ \dot{H}^{(T)} j_\ell(k(t_0-t)) \sim H^{(T)}(t=1/k)\, j_\ell(kt_0)$. For a pure power law,

$$k^3\,|H(k,t=1/k)|^2 = A_T(kt_0)^{n_T} \ , \tag{2.257}$$

one obtains

$$\begin{aligned} C_\ell^{(T)} &\simeq \frac{2}{\pi}\frac{(\ell+2)!}{(\ell-2)!}A_T\int\frac{dx}{x}x^{n_T}\frac{j_\ell^2(x)}{x^4} \\ &= \frac{(\ell+2)!}{(\ell-2)!}A_T\frac{\Gamma(6-n_T)\Gamma(\ell-2+\frac{n_T}{2})}{2^{6-n_T}\Gamma^2(\frac{7}{2}-n_T)\Gamma(\ell+4-\frac{n_T}{2})} \ . \end{aligned} \tag{2.258}$$

For a scale-invariant spectrum ($n_T = 0$) this results in

$$\ell(\ell+1)C_\ell^{(T)} \simeq \frac{8}{15\pi}\frac{\ell(\ell+1)}{(\ell+3)(\ell-2)}A_T \ . \tag{2.259}$$

The singularity at $\ell = 2$ in this crude approximation is not real, but there is some enhancement of $\ell(\ell+1)C_\ell^{(T)}$ at $\ell \sim 2$ (see Fig. 2.4).

Since tensor perturbations decay on subhorizon scales, $\ell \gtrsim 60$, they are not very sensitive to cosmological parameters.

Again, inflationary models (and topological defects) predict a scale-invariant spectrum of tensor fluctuations ($n_T \sim 0$).

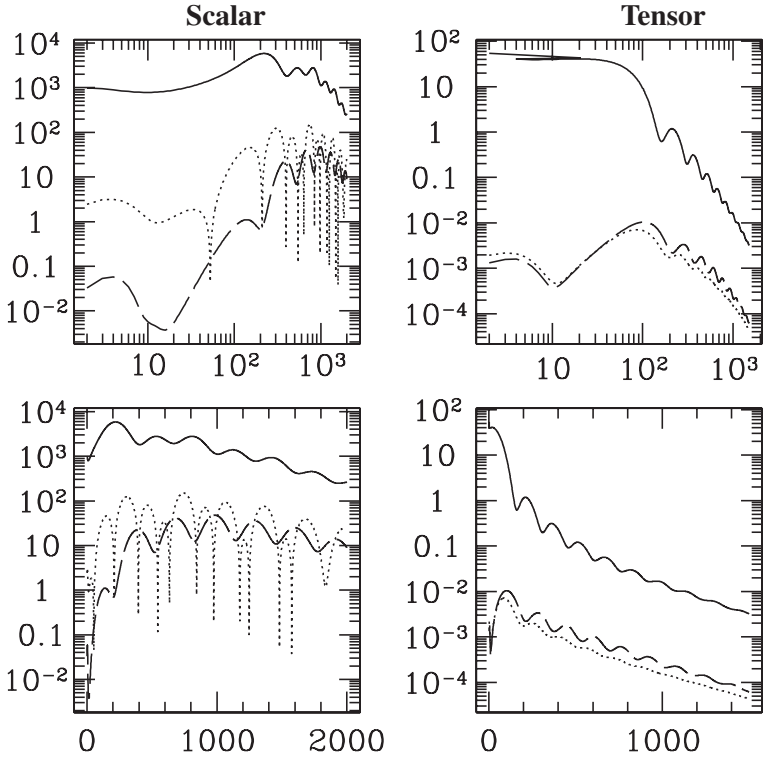


Fig. 2.4. Adiabatic scalar and tensor CMB anisotropy spectra are plotted, $\ell(\ell+1)C_\ell/(2\pi)$ in units of $(\mu K)^2$ in log-scale (top panels), where the Sachs–Wolfe plateau is clearly visible and in linear scale (bottom panels) which shows the equal spacing of the acoustic peaks. The solid line shows the temperature spectrum, the dashed line is the polarization and the dotted line shows the temperature–polarization cross correlation (see Chapter 5). The latter can become negative, the deep spikes in the dotted curves in the left-hand panels are actually sign changes. The left-hand side shows scalar fluctuation spectra, while the right-hand side shows tensor spectra. The observational data are well fitted by a purely scalar spectrum. Comparison of data and a model scalar spectrum are shown in Figs. 6.5–6.7.

Comparing the tensor and scalar result for scale-invariant perturbations we obtain for large scales, $\ell < 50$

$$\frac{C_\ell^{(T)}}{C_\ell^{(S)}} \simeq \frac{72}{15}\frac{A_T}{A_S} \equiv r \ . \tag{2.260}$$

Present CMB anisotropy data favour a roughly scale-invariant spectrum with amplitude

$$\ell(\ell+1)C_\ell \simeq 6\times 10^{-10} \ \text{ for } \ \ell \lesssim 50 \ .$$

If the perturbations are purely scalar, this requires $A_S \simeq 1.7 \times 10^{-8}$, if they were purely tensorial (which we know they are not), we would need $A_T \simeq 3.5 \times 10^{-9}$. In general observations require

$$\frac{A_S}{9\pi}(1+r) \simeq 6 \times 10^{-10} \,. \tag{2.261}$$

On very small angular scales, $\ell \gtrsim 800$, fluctuations are damped by collisional damping (Silk damping). This effect has to be discussed with the Boltzmann equation for photons which is presented in detail in Chapter 4.

2.7 Final remarks

In this chapter we have developed the basics of cosmological perturbation theory. Perturbation theory is an important tool especially to calculate CMB anisotropies and polarization since these are very small and can be determined reliably within linear cosmological perturbation theory. Here we have discussed the Einstein equations and the propagation of light-like geodesics. Linear perturbations of the somewhat more involved Boltzmann equation which is more adequate to study CMB anisotropies and polarization will be developed in Chapters 4 and 5. To determine the evolution of the matter density fluctuations at late times, linear perturbation theory has to be complemented with the theory of weakly non-linear Newtonian gravity and with N-body simulations. To finally understand the formation of galaxies non-gravitational highly non-linear physics, like heating and cooling mechanisms, dissipation and nuclear reactions have to be taken into account. These topics go beyond the scope of this book.

Exercises

(The exercises marked with an asterisk are solved in Appendix A10.2.)

Ex. 2.1 Gauge transformations*

Using the general formulae of Appendix A2.2 derive Eq. (2.30),

$$\begin{aligned} L_{\mathbf{X}}\bar{g} = a^2[&-2(\mathcal{H}T + \dot{T})dt^2 + 2(\dot{L}_i - T_{,i})dt\, dx^i \\ &+ \left(2\mathcal{H}T\gamma_{ij} + L_{i|j} + L_{j|i}\right) dx^i dx^j]. \end{aligned} \tag{2.262}$$

Ex. 2.2 Normals to hypersurfaces

We consider the normal n^μ to some hypersurface of the form $t =$ constant for a suitably chosen time coordinate t. Then the corresponding 1-form, is of the form $n_\mu\, dx^\mu = f\, dt$. The function $f = \pm\sqrt{-g^{00}}$ is determined by the normalization

condition. Show that in this case the vorticity

$$\omega_{\mu\nu} = \frac{1}{2} P_\mu^\lambda P_\nu^\rho (n_{\lambda;\rho} - n_{\rho;\lambda}) \,, \tag{2.263}$$

vanishes. Here P_ν^ρ is the projection tensor defined in Eq. (2.40).

Hint: If you are familiar with forms, you find from Wald (1984) that a vector field is hypersurface orthogonal if and only if the corresponding 1-form $\alpha = n_\mu \, dx^\mu$ satisfies $\alpha \wedge d\alpha = 0$. Show that for a form given by $\alpha = f(x)\, dt$ this is always the case. Here f is an arbitrary function on the spacetime manifold.

Ex. 2.3 Adiabaticity*

Consider a mixture of two non-interacting fluids with sound speeds c_1, c_2 and enthalpies w_1, w_2. Determine Γ from the single fluid perturbation variables Γ_α, $D_{g\alpha}$, V_α, Π_α, $\alpha = 1, 2$. If the intrinsic perturbations of each fluid are adiabatic, $\Gamma_\alpha = 0$, what is the condition that the total perturbation be adiabatic, $\Gamma = 0$? Derive an evolution equation for Γ under the condition $\Gamma_\alpha = 0$.

Ex. 2.4 Power spectrum

For a spatial, statistically homogeneous and isotropic random variable $X(\mathbf{x})$ with vanishing mean we define the 2-point function

$$\xi(r) = \langle X(\mathbf{x}) X(\mathbf{x} + r\mathbf{n}) \rangle \,.$$

Homogeneity requires that ξ does not depend on the position of the first point, $\mathbf{x}$ and isotropy means that ξ is independent of the direction $\mathbf{n}$ in which the second point $\mathbf{y} = \mathbf{x} + r\mathbf{n}$ lies. Hence ξ depends only on the distance r. Show that the power spectrum defined in Eq. (2.235) is the Fourier transform of the correlation function ξ.

Ex. 2.5 Variable transformation

We consider an ordinary linear second-order differential equation for ϕ, which is of the form

$$\ddot{\phi} + f(t)\dot{\phi} + \omega^2(t)\phi(t) = 0 \,. \tag{2.264}$$

Show that the variable $\psi \equiv h(t)\phi$ with $h(t) = \exp\left(\frac{1}{2}\int^t f(t')\, dt'\right)$, satisfies the equation

$$\ddot{\psi} + \tilde{\omega}^2(t)\psi(t) = 0 \,, \tag{2.265}$$

where

$$\tilde{\omega}^2 = \omega^2 + \ddot{h}/h \,.$$

Use these findings to derive Eq. (2.154).

Hint: Use

$$\frac{d}{dt}[\rho(1+w)]^{-1/2} = \frac{3}{2}\mathcal{H}(1 + c_s^2)[\rho(1+w)]^{-1/2} \,.$$

To derive this equation make use of Eq. (2.113) and the energy conservation equation. Finally, use Eq. (2.109).

This shows that a linear second-order differential equation can always be brought into the form of the equation for a harmonic oscillator with a time-dependent frequency. This is useful to know, not only because we can quantize this system easily, but also since we know that an instability sets in, when $\tilde{\omega}^2 < 0$. By studying the time dependence of the frequency, we can therefore infer the qualitative behaviour of the solution.

Ex. 2.6 Perturbations in universes with non-flat spatial section

Consider a universe filled with dust, $c_s^2 = w = 0$. In this case, the Bardeen equation reduces to

$$\ddot{\Psi} + 3\mathcal{H}\dot{\Psi} - 2K\Psi = 0\,. \tag{2.266}$$

Solve this equation for both cases, $K > 0$ and $K < 0$ and discuss the results. What happens for $K < 0$ at late times? How do perturbations evolve during a collapsing universe?

Ex. 2.7 Acoustic peaks

Consider a universe with matter and radiation but no curvature or cosmological constant. Solve the Friedmann equation exactly and determine $t_0/t_{\rm dec}$ given that $z_{\rm dec} \simeq 1100$. Insert a realistic value for Ω_r and keep h in the expression. Use the result to approximate the positions of the acoustic peaks in this Universe. Discuss qualitatively the change in the peak position in the following cases:

(a) addition of a cosmological constant (at fixed Ω_r and for vanishing curvature).

(b) addition of curvature (at fixed Ω_r and $\Lambda = 0$).

3

Initial conditions

So far we have only studied the evolution of perturbations assuming that the initial conditions are fixed and given once and for all. Now we want to study how classical perturbations are generated out of quantum fluctuations during a simple inflationary phase. The fact that inflation generates a nearly scale-invariant spectrum of scalar perturbations in good agreement with the observations of the cosmic microwave background is to be considered as its greatest success. The solution of the flatness and entropy problems with an inflationary phase are actually 'post-dictions' while the scale-invariant spectrum of scalar perturbations was first predicted in Mukhanov & Chibishov (1982) and Mukhanov *et al.* (1992), long before its discovery by the COBE satellite by Smoot *et al.* (1992). It represents therefore a real prediction of inflation. There are also other models for structure formation which predict a scale-invariant spectrum of fluctuations but which disagree with the detailed spectrum of CMB fluctuations such as topological defects (Durrer *et al.*, 2002).

In this chapter we first study perturbations in a FL universe filled with a scalar field. Next we discuss the generation of fluctuations during inflation. We especially determine the spectral index of scalar and tensor perturbations and the ratio of their amplitudes in the slow roll approximation. This will lead us to the well known consistency relation for slow roll inflation. We study in detail the simple case of one scalar field, the 'inflaton'.

Finally, we discuss more general initial conditions which are relevant if more than one scalar field plays a role during inflation, so-called mixed adiabatic and iso-curvature fluctuations.

As we discussed in Chapter 1, an inflationary phase supresses curvature and in order for curvature to be of order unity or smaller today, it must have been very small in the early Universe. In this chapter which deals mainly with the early Universe, we therefore neglect curvature, $K = 0$.

3.1 Scalar field perturbations

We consider here the special case of a FL universe filled with self-interacting scalar field matter. The action is given by

$$S = \frac{-1}{16\pi G}\int d^4x\,\sqrt{|g|}R - \int d^4x\,\sqrt{|g|}\left(\frac{1}{2}\partial_\mu\varphi\,\partial^\mu\varphi + W(\varphi)\right), \tag{3.1}$$

where φ denotes the scalar field and W is the potential. The energy–momentum tensor is obtained by varying the action w.r.t. the metric $g^{\mu\nu}$,

$$T_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - \left[\frac{1}{2}\partial_\lambda\varphi\,\partial^\lambda\varphi + W\right]g_{\mu\nu}\,. \tag{3.2}$$

The energy density ρ and the energy flux u are defined by

$$T^\mu_\nu u^\nu = -\rho u^\mu\,. \tag{3.3}$$

For the homogeneous and isotropic FL background we obtain (see also Chapter 1)

$$\rho = \frac{1}{2a^2}\dot\varphi^2 + W\,, \qquad (u^\mu) = \frac{1}{a}(1,\vec{0})\,. \tag{3.4}$$

The pressure is

$$T^i_j = P\delta^i_j\,, \qquad P = \frac{1}{2a^2}\dot\varphi^2 - W\,. \tag{3.5}$$

We want to derive the linear perturbation equations for the evolution of scalar field and metric perturbations. We define the scalar field perturbation modes,

$$\varphi = \bar\varphi + \delta\varphi Q^{(S)}\,. \tag{3.6}$$

Clearly, the scalar field only generates scalar perturbations in the energy momentum tensor (to first order). With the definition (3.2) we obtain

$$\begin{aligned}\delta T_{\mu\nu} = {}& \partial_\mu\bar\varphi\,\partial_\nu\,\delta\varphi + \partial_\nu\,\bar\varphi\partial_\mu\,\delta\varphi + a^{-2}\bar{\dot\varphi}\,\delta\dot\varphi\bar g_{\mu\nu}\\ &+\left[\frac{1}{2a^2}(\dot{\bar\varphi})^2 - \bar W\right]\delta g_{\mu\nu} - \bar W'\delta\varphi\bar g_{\mu\nu}\,.\end{aligned} \tag{3.7}$$

Inserting Eq. (3.7) in the definition (3.3) of the energy density and energy–flux, and setting $\rho = \bar\rho + \delta\rho Q^{(S)}$ and

$$(u^\mu) = \frac{1}{a}\left(1 - AQ^{(S)}, v\,Q^{(S)}_i\right)\,, \tag{3.8}$$

we find

$$\delta\rho = \frac{1}{a^2}\dot{\bar\varphi}\,\delta\dot\varphi - \frac{1}{a^2}\dot{\bar\varphi}^2 A + W_{,\varphi}\,\delta\varphi\,, \tag{3.9}$$

and

$$v = \frac{k}{\dot{\bar{\varphi}}} \left(\delta\varphi + \dot{\bar{\varphi}} k^{-1} B \right) . \tag{3.10}$$

The stress tensor, $T_{ij} = \varphi_{,i}\, \varphi_{,j} - \left[\frac{1}{2} \partial_\lambda \varphi \partial^\lambda \varphi + W \right] g_{ij}$ yields

$$P\pi_L = \frac{1}{a^2} \dot{\bar{\varphi}} \delta\dot{\varphi} - \frac{1}{a^2} \dot{\bar{\varphi}}^2 A - W_{,\varphi}\, \delta\varphi \quad \text{and} \quad \Pi = 0 . \tag{3.11}$$

We now define a gauge-invariant scalar field perturbation which corresponds to the value of $\delta\varphi$ in longitudinal gauge.

$$\delta\varphi^{(gi)} = \delta\varphi + \dot{\bar{\varphi}} k^{-1} (B - k^{-1} \dot{H}_T) = \delta\varphi - \dot{\bar{\varphi}} k^{-1} \sigma = \delta\varphi^{(\mathrm{long})} . \tag{3.12}$$

The second and third expressions give $\delta\varphi^{(gi)}$ in a generic gauge. Under a gauge transformation the scalar field perturbation simply changes by $\delta\varphi \to \delta\varphi + \dot{\bar{\varphi}} T$. Since $\sigma \to \sigma + kT$, it is clear that the combination $\delta\varphi^{(gi)}$ is gauge-invariant. On the other hand, in longitudinal gauge $B = H_T = 0$, so that $\delta\varphi^{(gi)} = \delta\varphi^{(\mathrm{long})}$. This variable is very simply related to the other gauge-invariant scalar variables. Short calculations give

$$V = k \delta\varphi^{(gi)} / \dot{\bar{\varphi}} , \tag{3.13}$$

$$D_g = -(1+w) \left[4\Psi + 2\mathcal{H} k^{-1} V - k^{-1} \dot{V} \right] , \tag{3.14}$$

$$D_s = -(1+w) \left[\Psi + 2\mathcal{H} k^{-1} V - k^{-1} \dot{V} \right] , \tag{3.15}$$

$$D = -(1+w) \left[\Psi - \mathcal{H} k^{-1} V - k^{-1} \dot{V} \right] , \tag{3.16}$$

$$\Gamma = \frac{2 W_{,\varphi}\, \dot{\bar{\varphi}}}{P \dot{\rho}} \left[\rho D_s - \dot{\rho} k^{-1} V \right] , \tag{3.17}$$

$$\Pi = 0 . \tag{3.18}$$

The last equation implies that the two Bardeen potentials are equal for scalar field perturbations, $\Phi = \Psi$. Using this we can write the perturbed Einstein equations fully in terms of the Bardeen potentials Ψ and V. We actually only need (2.104). Since we need them mainly to discuss inflation where curvature is negligible, we write them down here only for the case $K = 0$:

$$k^2 \Psi = 4\pi G \dot{\varphi}^2 \left[\Psi - \mathcal{H} k^{-1} V - k^{-1} \dot{V} \right] , \tag{3.19}$$

$$\dot{\Psi} + \mathcal{H} \Psi = 4\pi G \dot{\varphi}^2 k^{-1} V , \tag{3.20}$$

where we have used $a^2 \rho (1+w) = \dot{\varphi}^2$. To simplify the notation, we have dropped the overbar on the background quantities. With the help of Eqs. (2.104) and (2.106) one can easily generalize these equations to the case with curvature. Using (3.20) to eliminate V and $\dot{V}$ from Eq. (3.19), leads to the following second-order equation

for the Bardeen potential:

$$\ddot{\Psi} + 2(\mathcal{H} - \ddot{\varphi}/\dot{\varphi})\dot{\Psi} + (2\dot{\mathcal{H}} - 2\mathcal{H}\ddot{\varphi}/\dot{\varphi} + k^2)\Psi = 0\,. \tag{3.21}$$

Here we have also used the fact that $4\pi G\dot{\varphi}^2 = 4\pi G a^2 \rho(1+w) = \mathcal{H} - \dot{\mathcal{H}}$. Inserting the definition $c_s^2 = \dot{P}/\dot{\rho} = -\frac{1}{3\mathcal{H}}(2\frac{\ddot{\varphi}}{\dot{\varphi}} + \mathcal{H})$, we can also write (3.21) in the form

$$\ddot{\Psi} + 3\mathcal{H}(1+c_s^2)\dot{\Psi} + (2\dot{\mathcal{H}} + (1+3c_s^2)\mathcal{H}^2 + k^2)\Psi = 0\,. \tag{3.22}$$

This equation differs from the Ψ equation for a perfect fluid only in the term proportional to k^2 which is not multiplied with the adiabatic sound speed c_s^2. Indeed the scalar field is not in a thermal state with fixed entropy, $\Gamma \neq 0$. It is in a fully coherent state so that field fluctuations propagate with the speed of light and not with some adiabatic sound speed. On large scales, $kt \ll 1$ this difference is not relevant, but on sub-Hubble scales it does play a certain role.[1]

During slow roll inflation we can express the background variables in terms of $\mathcal{H}$ and the slow roll parameters ϵ_1 and ϵ_2 defined in Chapter 1. With the definition (1.153) and the expression for $\varphi''/(3H\varphi')$ above, we obtain

$$\frac{\ddot{\varphi}}{\mathcal{H}\dot{\varphi}} = \frac{H(d\varphi/d\tau) + d^2\varphi/d\tau^2}{H(d\varphi/d\tau)} = 1 + 3\epsilon_2 + \epsilon_1\,, \tag{3.23}$$

so that

$$2(\dot{\mathcal{H}} - \mathcal{H}\ddot{\varphi}/\dot{\varphi}) = -2\mathcal{H}^2(3\epsilon_2 + 2\epsilon_1)\,. \tag{3.24}$$

Inserting these results in Eq. (3.21) we find

$$\ddot{\Psi} - 2(3\epsilon_2 + \epsilon_1)\mathcal{H}\dot{\Psi} - \left[2\mathcal{H}^2(3\epsilon_2 + 2\epsilon_1) - k^2\right]\Psi = 0\,. \tag{3.25}$$

Hence on small scales, $(3\epsilon_2 + 2\epsilon_1)\mathcal{H}^2 \ll k^2$, Ψ oscillates, while on super-Hubble scales, $k/\mathcal{H} \ll 1$, it varies slowly as long as the slow roll parameters are small. During the transition from inflation to the radiation dominated era, where the slow roll parameters reach order unity, the Bardeen potential can however vary substantially. It is thus not very well suited to determine the amplitude of perturbations which have been induced during inflation in the radiation dominated era. We now show, that, on super-Hubble scales, the curvature variable ζ remains constant during

[1] Often, the terms 'Hubble scale' and 'horizon scale' are used interchangeably. For inflation, however they can differ by many orders of magnitude. During inflation the (comoving) Hubble scale, $\mathcal{H}^{-1} \simeq |t| = -t$ is decreasing and much smaller than the comoving horizon scale, $\int_{t_i}^{t} dt \simeq -t_i \gg \mathcal{H} \simeq -t$. The scale relevant for the behaviour of perturbations is, however always the Hubble scale, since $\mathcal{H}$ enters into the perturbation equations. The horizon is a global quantity, an integral, it does not determine whether perturbations are oscillating or whether they behave like a power law. It just happens that in a decelerating universe the two scales are often of the same order. In this chapter we shall be careful not to mix them up. We shall use the terms 'Hubble scale' for the Hubble scale and 'Hubble exit' for a scale growing larger than the Hubble scale during inflation.

the transition from inflation to the radiation dominated era. To study the evolution of super-Hubble perturbations from inflation into the radiation era, we shall therefore use the variable ζ.

As for the case of fluids, see Eqs. (2.154)–(2.156), we introduce the variable u

$$u = a[4\pi G(\mathcal{H}^2 - \dot{\mathcal{H}})]^{-1/2}\Psi \,, \tag{3.26}$$

which now satisfies the equation

$$\ddot{u} + (k^2 - \ddot{\theta}/\theta)u = 0 \,, \tag{3.27}$$

where

$$\theta = \frac{3\mathcal{H}}{2a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}}}} \,. \tag{3.28}$$

The difference to the fluid equations is just the factor c_s^2 in front of k^2 which for the scalar fields is replaced by 1. Scalar field fluctuations propagate with the speed of light. The curvature variable ζ in a scalar field background is given by

$$\zeta \equiv \frac{2(\mathcal{H}^{-1}\dot{\Psi} + \Psi)}{3(1+w)} + \Psi \,. \tag{3.29}$$

Note that we need $w > -1$ so that ζ is well defined. In a pure de Sitter space, we cannot work with the ζ-variable. This becomes even more evident when expressing the above relation in terms of the slow roll parameter ϵ_1. From (1.152) and

$$(1+w) = \frac{\dot{\varphi}^2}{\frac{1}{2}\dot{\varphi}^2 + a^2 W} = \frac{\frac{2}{3}\epsilon_1}{\frac{1}{3}\epsilon_1 + 1} \simeq \frac{2}{3}\epsilon_1 \,, \tag{3.30}$$

we obtain

$$\zeta \simeq \frac{\mathcal{H}^{-1}\dot{\Psi} + \Psi}{\epsilon_1} + \Psi \,. \tag{3.31}$$

For ζ to be well defined we therefore need $\epsilon_1 \neq 0$ (or the perturbations have to decay, $\dot{\Psi} = -\mathcal{H}\Psi$). From Eq. (3.25), using Eq. (1.153), one finds

$$\dot{\zeta} = -\frac{2k^2}{3(1+w)\mathcal{H}}\Psi = -\frac{(1+\frac{1}{3}\epsilon_1)k^2}{\epsilon_1\mathcal{H}}\Psi \simeq \frac{-k^2}{\epsilon_1\mathcal{H}}\Psi \,. \tag{3.32}$$

As in the case of fluids, this implies that the curvature perturbation ζ is conserved on super-Hubble scales, $k/\mathcal{H} \ll 1$. Using Eqs. (3.20) and (3.13) and $\dot{\phi}^2 = a^2\rho(1+w)$, we can express ζ also as

$$\zeta = \frac{\mathcal{H}}{\dot{\varphi}}\delta\varphi^{(gi)} + \Psi \,. \tag{3.33}$$

As we have seen in Chapter 2, Eqs. (2.159)–(2.161), the evolution of ζ is closely related to the one of v defined by

$$v = \frac{a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}}}}{\sqrt{4\pi G}\mathcal{H}}\zeta = a\,\delta\varphi^{(gi)} + \frac{a\dot{\varphi}}{\mathcal{H}}\Psi \,. \tag{3.34}$$

The variable v satisfies the equation of motion

$$\ddot{v} + (k^2 - \ddot{z}/z)v = 0 \,, \tag{3.35}$$

with

$$z = \frac{a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}}}}{\sqrt{4\pi G}\mathcal{H}} = a\sqrt{\frac{3(1+w)}{8\pi G}} = \frac{a\sqrt{a^2(\rho + P)}}{\mathcal{H}} = \frac{a\dot{\varphi}}{\mathcal{H}} \,. \tag{3.36}$$

Hence

$$v = z\zeta \,.$$

Also note that z is related to the slow roll parameter ϵ_1 by

$$\epsilon_1 = -\frac{dH/d\tau}{H^2} = \frac{\mathcal{H}^2 - \dot{\mathcal{H}}}{\mathcal{H}^2} = 4\pi G\frac{z^2}{a^2} \,. \tag{3.37}$$

The equation of motion (3.35) can be obtained from the Fourier decomposition of the action

$$S = -\frac{1}{2}\int d^4x\sqrt{|g|}\left(\partial_\mu v\,\partial^\mu v + m^2(t)v^2\right) \,, \tag{3.38}$$

where $m^2 = -(\ddot{z}/z)a^{-2}$. This is the action of a simple free field with time-dependent mass term. For a constant or slowly varying w we have $z \propto a$ and during inflation $\ddot{z}/z > 0$, hence $m^2(t) < 0$, which represents an instability and leads to the amplification of vacuum fluctuations (or particle creation). During ordinary expansion, $\ddot{z}/z < 0$ and the vacuum state is stable.

In the next section we want to study quantum fluctuations of the variable v.

Remark: Note that also Eq. (3.27) can be written as a Euler–Lagrange equation for a canonical scalar field Lagrangian with time-dependent mass term $m_\theta^2 = -(\ddot{\theta}/\theta)a^{-2}$ for the variable u defined in Eq. (3.26). The problem there is however, that we cannot 'switch off' gravity for this variable, which diverges in the limit $\mathcal{H}, \dot{\mathcal{H}} \to 0$. Hence u does not have well defined initial conditions when $k^2 \gg |\ddot{\theta}/\theta|$. Even though the perturbation equations take the form of canonical equations in this variable, it is therefore not the correct variable to quantize.

The above action (3.38) can also be obtained by perturbing the original action of the system to second order,

$$S = \int dx^4\sqrt{|g|}\left(\frac{-R}{16\pi G} - \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - W\right) \,.$$

A long and cumbersome calculation, removing several total derivatives (Mukhanov *et al.*, 1992) shows that to second order, the perturbation of this action is given by (3.38). A much more elegant recent derivation of this result is given in Maldacena (2003).

3.2 Generation of perturbations during inflation

So far we have simply assumed some initial fluctuation amplitude A, without investigating where it came from or what the k-dependence of A might be. In this section we discuss the most common idea about the generation of cosmological perturbations, namely their production from the quantum vacuum fluctuations during an inflationary phase.

The basic idea is simple: a time-dependent gravitational field generically leads to particle production, analogously to the electron–positron production in a classical, time-dependent electromagnetic field.

3.2.1 Scalar perturbations

The main result is the following: during inflation, the produced particles induce a perturbed gravitational potential with a (nearly) scale-invariant spectrum,

$$k^3|\zeta(k,t)|^2 = k^{n-1} \times \text{constant} \quad \text{with} \quad n \simeq 1\,. \tag{3.39}$$

The quantity $k^3|\zeta(k,t)|^2$ is the squared amplitude of the curvature perturbation at comoving scale $\lambda = \pi/k$. To make sure that this quantity is small on a broad range of scales, so that neither black holes are formed on small scales nor large deviations from homogeneity and isotropy on large scales appear, we must require $n \simeq 1$. These arguments were put forward for the first time by Harrison and Zel'dovich (Harrison, 1970; Zel'dovich, 1972) (still before the advent of inflation), leading to the name 'Harrison–Zel'dovich spectrum' for a scale-invariant perturbation spectrum as discussed in Chapter 2.

To derive the above result we consider a scalar field background with energy density which is dominated by the potential term, hence the slow roll parameters, ϵ_1 and ϵ_2 are small. Over a reasonably short period of time we can approximate them as constants leading to nearly power law expansion

$$a \propto |t|^q \quad \text{with} \quad q = -1 - \epsilon_1 + \mathcal{O}(\epsilon_1^2)\,. \tag{3.40}$$

We want to determine the Bardeen potential during the subsequent radiation and matter dominated era. For this we use the fact that the curvature perturbation ζ remains constant on super-Hubble scales. Hence if we calculate its amplitude at

Hubble crossing, $k/\mathcal{H} = 1$, during inflation it will remain constant until it re-enters the Hubble-horizon in the radiation or matter dominated era.

Quantization

To determine the initial conditions and the evolution of v we now quantize the variable v in the action

$$\delta S = -\frac{1}{2}\int dx^4 \left[(\partial_\mu v)^2 + m^2(t)v\right] = -\int dx^4\, \mathcal{L}\,, \tag{3.41}$$

with canonical momentum $\pi = \partial\mathcal{L}/\partial\dot{v} = \dot{v}$. We interpret $\hat{v}$ and $\hat{\pi}$ as operators on a Hilbert space which satisfy the standard canonical commutation relations given by

$$[\hat{v}(t,\mathbf{x}), \hat{v}(t,\mathbf{x}')] = [\hat{\pi}(t,\mathbf{x}), \hat{\pi}(t,\mathbf{x}')] = 0, \qquad \text{and} \tag{3.42}$$

$$[\hat{v}(t,\mathbf{x}), \hat{\pi}(t,\mathbf{x}')] = i\delta^3(\mathbf{x}-\mathbf{x}') \qquad (\hbar = 1)\,. \tag{3.43}$$

We now expand the operator $\hat{v}$ in Fourier modes,

$$\hat{v}(t,\mathbf{x}) = \frac{1}{(2\pi)^{3/2}}\int d^3k [v_k(t)\hat{a}_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}} + \bar{v}_k(t)\hat{a}^*_\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{x}}]\,. \tag{3.44}$$

The operators $\hat{a}_\mathbf{k}$ and their hermitean conjugates $\hat{a}^*_\mathbf{k}$ are the annihilation and creation operators. Since $\hat{v}$ describes a real field, $\bar{v}_k$ is the complex conjugate of v_k. Choosing the time-independent normalization

$$\bar{v}_k\dot{v}_k - v_k\dot{\bar{v}}_k = +i\,, \tag{3.45}$$

Eqs. (3.42) and (3.43) require that the operators $a_\mathbf{k}$ satisfy the usual commutation relations

$$[\hat{a}_\mathbf{k}, \hat{a}_{\mathbf{k}'}] = [\hat{a}^*_\mathbf{k}, \hat{a}^*_{\mathbf{k}'}] = 0 \quad \text{and} \quad [\hat{a}_\mathbf{k}, \hat{a}^*_{\mathbf{k}'}] = \delta^3(\mathbf{k}-\mathbf{k}')\,. \tag{3.46}$$

The time-dependent mode functions v_k obey the classical equation of motion (3.35).

At very early times, $k \gg \mathcal{H}$ we can neglect the mass term, $\ddot{z}/z$, in Eq. (3.35) and v_k is the mode function of a free massless scalar field. We assume that initially $\hat{v}$ is in the vacuum state so that

$$v_k(t) = \frac{1}{\sqrt{2k}}\exp(-ikt) \quad \text{for} \quad k \gg \mathcal{H}\,. \tag{3.47}$$

Note that we work (as is usual in quantum field theory) in the Heisenberg picture. The state, which we assume to be the vacuum state at very early times and which we denote by $|0\rangle$ does not evolve but the field operator does.

At late times, $k \ll \mathcal{H}$, we may neglect k^2 in Eq. (3.35) and the growing mode solution behaves like $v_k \propto z$, so that $\zeta_k = v_k/z$ remains constant as expected.

We want to calculate the power spectrum of

$$\hat{\zeta}(t,\mathbf{x}) = \frac{1}{(2\pi)^{3/2}}\int d^3k[\zeta_k(t)\hat{a}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} + \bar{\zeta}_k(t)\hat{a}^*_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{x}}]$$
$$= \frac{1}{(2\pi)^{3/2}}\int d^3k[\hat{\zeta}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} + \hat{\zeta}^*_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{x}}]\,, \tag{3.48}$$

defined by

$$\langle 0|\,\hat{\zeta}_{\mathbf{k}}\hat{\zeta}^*_{\mathbf{k}'}\,|0\rangle \equiv P_\zeta(k)\delta^3(\mathbf{k}-\mathbf{k}')\,. \tag{3.49}$$

Inserting Eq. (3.48) with $\zeta_k = v_k/z$ and using the properties of the vacuum, $\langle X|\,a_k\,|0\rangle = \langle 0|\,a^*_k\,|X\rangle = 0$, for an arbitrary state $|X\rangle$ as well as the commutation relations (3.46) one obtains (see exercises)

$$P_\zeta(k) = \frac{|\,v_k(t)\,|^2}{z^2}\,. \tag{3.50}$$

Perturbation spectrum from power law inflation

As a first simple example we consider power law inflation, $a \propto |t|^q$, $q \lesssim -1$. In this case $\dot{\mathcal{H}} \propto \mathcal{H}^2$ so that $z \propto a$ The evolution equation (3.35) for v then reduces to (we suppress the index k again)

$$\ddot{v} + \left(k^2 - \frac{q(q-1)}{t^2}\right)v = 0\,. \tag{3.51}$$

The solutions to this equation are of the form $(k|t|)^{1/2}H^{(i)}_\mu(kt)$, where $\mu = \frac{1}{2} - q$ and $H^{(i)}_\mu$ is the Hankel function of the ith kind ($i = 1$ or 2) of order μ. The initial condition (3.47) requires that only $H^{(2)}_\mu$ appears. Fixing the constants we obtain

$$v = -\frac{i\sqrt{\pi}\,\exp(iq\pi/2)}{2\sqrt{k}}(k|t|)^{1/2}H^{(2)}_\mu(kt)\,.$$

At late times, $k/\mathcal{H} \sim k|t| \ll 1$, we have $H^{(2)}_\mu(kt) \simeq (i/\pi)\Gamma(\mu)(kt/2)^{-\mu}$. Inserting this above we obtain

$$v \simeq C(\mu)e^{i\alpha}(k|t|)^{1/2-\mu}k^{-1/2}, \quad k|t| \ll 1, \tag{3.52}$$

with $C(\mu) = (2^{\mu-1}/\pi^{1/2})\Gamma(\mu)$. The phase $e^{i\alpha}$ is uninteresting, it disappears in the power spectrum. The power spectrum of $\zeta = v/z$ is thus given by

$$P_\zeta(k,t) = \left|\frac{v}{z}\right|^2 = C(\mu)^2\frac{(k|t|)^{1-2\mu}}{z^2k} \simeq \frac{4\pi C(\mu)^2}{\epsilon_1 m_P^2}\frac{(k|t|)^{1-2\mu}}{a^2k}\,, \quad k|t| \ll 1\,, \tag{3.53}$$

where we have used Eq. (3.37) in the last equals sign. Recalling that $1 - 2\mu = 2q$, we see that P_ζ is time independent on super-Hubble scales, as expected. We now

replace $|t|$ by $\mathcal{H} = -q/|t|$ and multiply the spectrum by k^3, which yields[2]

$$k^3 P_\zeta(k) = \frac{2\pi H^2}{\epsilon_1 m_P^2} \left(\frac{k}{\mathcal{H}}\right)^{3-2\mu} , \tag{3.54}$$

where we have used $H = \mathcal{H}/a$ and $C(\mu)^2 \simeq 1/2$. The latter approximation is obtained by setting $q = -1$ and $\mu = 3/2$ in the expression for $C(\mu)$ and $\mathcal{H}$ above. The amplitude of a given mode k at Hubble exit is given by

$$k^3 P_\zeta(k)\big|_{k=\mathcal{H}} = \frac{2\pi H^2}{\epsilon_1 m_P^2} . \tag{3.55}$$

The scalar spectral index n is defined by

$$n - 1 = \frac{d \log\left(k^3 P_\zeta\right)}{d \log(k)} . \tag{3.56}$$

From Eq. (3.54), using $\epsilon_1 = 1 - \dot{\mathcal{H}}/\mathcal{H}^2$, see Eq. (1.152), we obtain (see also Eq. 3.40)

$$n - 1 = 3 - 2\mu = 2 + 2q = -\frac{2\epsilon_1}{1 - \epsilon_1} \simeq -2\epsilon_1 . \tag{3.57}$$

While the equals signs are exact for power law inflation, the last approximate sign is valid only to first order in ϵ_1.

In Ex. 3.1 you show that power law inflation with a scalar field requires an exponential potential $V = V_0 \exp(-\alpha\varphi/m_P)$. The slow roll parameters are readily calculated, $\epsilon_1 = \alpha^2/16\pi = -(3/2)\epsilon_2$. We therefore can also write with the same accuracy as above

$$n - 1 = \frac{-6(\epsilon_1 + \epsilon_2)}{1 - \epsilon_1} \simeq -6(\epsilon_1 + \epsilon_2) . \tag{3.58}$$

We shall see in the next paragraph that this is the general result for slow roll inflation.

Slow roll inflation

We now want to derive equations (3.58) and (3.54) when expansion no longer follows a power law exactly, but the slow roll parameters ϵ_1 and ϵ_2 are small. Let us first calculate the mass term, $\ddot{z}/z$ in this case. From Eq. (3.36) we have $z = \dot{\varphi} a/\mathcal{H}$. Taking the derivative of this using Eq. (3.24) we obtain

$$\frac{\dot{z}}{z} = (1 + 2\epsilon_1 + 3\epsilon_2)\mathcal{H} , \tag{3.59}$$

[2] Remember that the conformal time is negative during inflation, $t < 0$.

leading to

$$\frac{\ddot{z}}{z} = \left(\frac{\dot{z}}{z}\right)^2 + \left(\frac{\dot{z}}{z}\right)^{\bullet}$$
$$= (1 + 2\epsilon_1 + 3\epsilon_2)\dot{\mathcal{H}} + (2\dot{\epsilon}_1 + 3\dot{\epsilon}_2)\mathcal{H} + (1 + 2\epsilon_1 + 3\epsilon_2)^2\mathcal{H}^2 \,. \quad (3.60)$$

As we saw in Eq. (1.154), the derivatives $\dot{\epsilon}_1$ and $\dot{\epsilon}_2$ are second order and can be neglected. Neglecting also all the other second-order terms we obtain

$$\frac{\ddot{z}}{z} = (2 + 9\epsilon_1 + 9\epsilon_2)\,\frac{1}{t^2} \,. \quad (3.61)$$

Here we have used $\mathcal{H}^2 = q^2/t^2 = (1+\epsilon_1)^2/t^2 \simeq (1+2\epsilon_1)/t^2$ and $\dot{\mathcal{H}} = -q/t^2 \simeq (1+\epsilon_1)/t^2$. Neglecting the time dependence of ϵ_1 and ϵ_2 (which is second order), the v equation (3.35) is therefore again a Bessel equation with $\mu^2 - 1/4 = 2 + 9\epsilon_1 + 9\epsilon_2$, hence

$$\mu = \frac{3}{2}\,(1 + 2\epsilon_1 + 2\epsilon_2) \,. \quad (3.62)$$

Inserting this in the power spectrum for ζ given in Eq. (3.54) we find

$$k^3 P_\zeta(k) = \frac{2\pi H^2}{\epsilon_1 m_P^2}\left(\frac{k}{\mathcal{H}}\right)^{-6(\epsilon_1+\epsilon_2)} \,, \quad k/\mathcal{H} \ll 1 \,. \quad (3.63)$$

The spectral index is therefore

$$n - 1 = -6(\epsilon_1 + \epsilon_2) \,, \quad (3.64)$$

which corresponds exactly to the expression (3.58) for power law inflation.

From the curvature spectrum it is now easy to determine the spectrum for the Bardeen potential Ψ in the matter dominated era, e.g. at recombination. As we have seen in Eq. (2.151), during power law expansion the growing mode of the Bardeen potential is constant on super-Hubble scales. In a matter dominated universe, $w = 0$, the Bardeen potential is even constant on all scales. The relation (3.29) then yields

$$k^3 P_\Psi(k) = \frac{9}{25} k^3 P_\zeta(k) = \frac{18\pi H^2}{25\epsilon_1 m_P^2}\left(\frac{k}{\mathcal{H}}\right)^{-6(\epsilon_1+\epsilon_2)} \,, \quad k/\mathcal{H} \ll 1 \,. \quad (3.65)$$

The amplitude A_S and the spectral index n of the Sachs–Wolfe contribution to the CMB power spectrum given in Eq. (2.250) are thus determined by the energy scale of inflation, H, and the slow roll parameters ϵ_1 and ϵ_2.

It is possible to develop the slow roll approximation further, to second and third order, which has been done in the literature (Hoffman & Turner, 2001; Schwarz

et al., 2001; Martin & Schwarz, 2003). But at first order it is already possible to constrain inflationary models by using the data, see Chapter 6.

3.2.2 Vector perturbations

In the simplest models of inflation where the only degrees of freedom are the scalar field and the metric, no vector perturbations are generated. But even if they are, subsequent evolution after inflation will lead to their decay. Indeed, in a perfect fluid background the anisotropic stress vanishes, $\Pi_{ij} = 0$. The evolution of vector perturbations is then given by Eq. (2.116) which implies for the fluid vorticity Ω

$$\Omega \propto a^{3c_s^2-1} . \tag{3.66}$$

For a radiation–matter fluid, $\dot{p}/\dot{\rho} = c_s^2 \leq 1/3$, this leads to a non-growing vorticity. The dynamical Einstein equation (2.108) gives

$$\sigma^{(V)} \propto a^{-2} , \tag{3.67}$$

and the constraint (2.105) reads (at early times, so that we can neglect curvature)

$$\Omega \sim (kt)^2 \sigma^{(V)} . \tag{3.68}$$

Therefore, even if they are created in the very early universe on super-Hubble scales during an inflationary period, vector perturbations of the metric will decay and become soon entirely negligible. Even if the vorticity remains constant in a radiation dominated universe, it will be so small on relevant scales at formation ($kt_{\rm in} \ll 1$) that we may safely neglect it.

Vector perturbations are irrelevant if perturbations have been created at some early time, e.g., during inflation. This result changes completely when considering 'active perturbations' such as topological defects where vector perturbations contribute significantly to the CMB anisotropies on large scales, see Durrer *et al.* (2002). It is interesting to note that, in a background without anisotropic stresses, vector perturbations do not satisfy a wave equation and therefore will not oscillate. Vorticity simply decays with time.

3.2.3 Tensor perturbations

The situation is different for tensor perturbations. Again we consider the perfect fluid case, $\Pi_{ij}^{(T)} = 0$. Equation (2.109) implies, if K is negligible,

$$\ddot{H}_{ij} + 2\mathcal{H}\dot{H}_{ij} + k^2 H_{ij} = 0 . \tag{3.69}$$

If the background has a power law evolution or is slowly rolling, $a \propto t^q$ with $q = -1 - \epsilon_1$, $\mathcal{H} = q/t$, this equation can be solved in terms of Bessel functions (see Abramowitz & Stegun 1970, Eq. 9.1.52). The less decaying mode solution to Eq. (3.69) is $H_{ij} = e_{ij} x^{1/2-q} Y_{a/2-9}(x)$, where Y_ν denotes the Bessel function of order ν, $x = kt$ and e_{ij} is a transverse traceless polarization tensor. This leads to

$$H_{ij} = \text{constant} \quad \text{for } x \ll 1 \,, \tag{3.70}$$

$$H_{ij} = \frac{1}{a} \quad \text{for } x \gtrsim 1 \,. \tag{3.71}$$

One may also quantize the tensor fluctuations which represent gravitons. Doing this, one obtains (up to small corrections) a scale-invariant spectrum of tensor fluctuations from inflation. For tensor perturbations the canonical variable is simply given by

$$h_{ij} = e_{ij} h = \frac{m_P a}{\sqrt{8\pi}} H_{ij} \,. \tag{3.72}$$

Here e_{ij} is a normalized transverse traceless polarization tensor, $e^i_i = k^i e_{ij} = 0$ and $e_{ij} e^{ij} = 2$. The evolution equation for h is obtained by inserting the ansatz (3.72) in Eq. (3.69),

$$\ddot{h} + (k^2 + m^2(t))h = 0 \,, \tag{3.73}$$

$$\text{with} \quad m^2(t) = -\frac{\ddot{a}}{a} = -(\dot{\mathcal{H}} + \mathcal{H}^2) = -(2 - \epsilon_1)\mathcal{H}^2$$

$$= -(2 - \epsilon_1)\left(\frac{1 + 2\epsilon_1}{t}\right)^2 \simeq -\frac{2 + 3\epsilon_1}{t^2} \,. \tag{3.74}$$

The variable h is canonically normalized and can therefore be quantized with the usual commutation relation. This is best seen by determining the perturbed action to second order,

$$S + \delta S = \frac{m_P^2}{16\pi} \int d^4x \sqrt{-(g + \delta g)}(R + \delta R) \,.$$

The derivation of δS to second order in h_{ij} is not very long for pure tensor perturbations (see Ex. 3.2).

During inflation the mass term $m^2(t)$ in Eq. (3.73) is negative, leading to particle creation. As for scalar perturbations, the vacuum initial conditions are given on scales which are inside the Hubble scale, $k^2 \gg |m^2|$, where expansion can be neglected and we may set

$$h_{\rm in} = \frac{1}{\sqrt{2k}} \exp(-ikt) \quad \text{for } k|t| \gg 1 \,.$$

Solving Eq. (3.73) with this initial condition, gives (up to an uninteresting phase)

$$h = \frac{\sqrt{\pi}}{2\sqrt{k}}(k|t|)^{1/2} H_\nu^{(2)}(kt) \,,$$

where $\nu^2 - 1/4 = 2 + 3\epsilon_1$ so that $\nu = 3/2 + \epsilon_1$. On super-Hubble scales, $H_\nu^{(2)}(kt) \propto (k|t|)^{-\nu}$ this results in

$$k^3 P_H = k^3 |H_{ij} H^{ij}| = 2\frac{8\pi k^3 |h|^2}{a^2 m_P^2} \simeq 8\pi \frac{H^2}{m_P^2}\left(\frac{k}{\mathcal{H}}\right)^{-2\epsilon_1} \,, \quad k/\mathcal{H} \ll 1 \,. \tag{3.75}$$

Note the factor 2 due to the two tensor modes. The pre-factor is obtained by setting $\nu = 3/2$ and $q = -1$. In the exponent, however we keep the slow roll parameter $\epsilon_1 \neq 0$.

From Eq. (3.75) we derive the tensor spectral index n_T defined by

$$n_T = \frac{d\log\left(k^3 P_H\right)}{d\log(k)} = -2\epsilon_1 \,. \tag{3.76}$$

After inflation H_{ij} is constant on super-Hubble scales. The gravity wave spectrum is therefore determined by the amplitude of the fluctuations at Hubble crossing,

$$k^3 P_H = A_T \,(kt_0)^{n_T} \,, \quad k/\mathcal{H} \ll 1 \,, \quad \text{with} \tag{3.77}$$

$$A_T = 8\pi \frac{H^2}{m_P^2}\bigg|_{k=\mathcal{H}_0} \simeq \frac{64\pi^2}{3m_P^4} W\bigg|_{k=\mathcal{H}_0} \,. \tag{3.78}$$

Note that here, as in Eq. (3.55), $H|_{k=\mathcal{H}} = k/a$ indicates the value of the Hubble parameter H at 'Hubble exit', i.e., during inflation when $1/k$ becomes larger than the comoving Hubble scale $1/\mathcal{H}$, corresponding to $a = a_1$ in Fig. 3.1. This is much larger than the value of $H = \mathcal{H}/a$ at re-entry, long after inflation, when again $k = \mathcal{H}$. At this second Hubble crossing time the Hubble parameter $H = k/a_2$ is much smaller since $a_2 \gg a_1$, is much larger, see Fig. 3.1.

The parameter A_T introduced here is the amplitude of the tensor spectrum at the present Hubble scale $t_0 \simeq 1/\mathcal{H}_0 = 1/H_0$, if we normalize the scale factor such that $a_0 = 1$. We could have chosen some arbitrary other reference scale k_0. But for definiteness and because of its relevance for CMB anisotropies we choose the Hubble scale. Equation (3.77) relates the tensor amplitude to the value of the inflaton potential at the moment when the comoving Hubble scale during inflation equals the present Hubble scale:

$$W_* \equiv W|_{\mathcal{H}=H_0} = \frac{3m_P^4}{64\pi^2} A_T \,. \tag{3.79}$$

Measuring the amplitude of tensor perturbations therefore allows us to determine the energy scale of inflation. If we simply require that tensor fluctuations do not generate

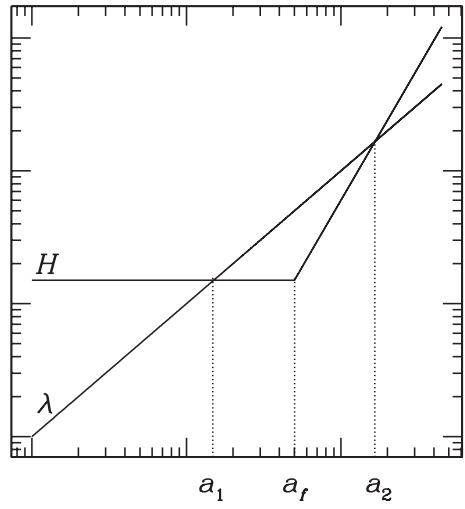


Fig. 3.1. We sketch the behaviour of the Hubble scale $H^{-1} = a\mathcal{H}^{-1}$ and some wavelength $\lambda = a/k$ during and after inflation as functions of the scale factor. At $a = a_1$ during inflation, the scale λ exits the Hubble scale and after inflation, at $a = a_2$, it re-enters. At $a = a_f$ inflation ends.

more than a third of the observed CMB anisotropies, according to Eq. (2.262), present observational limits require $A_T \lesssim 10^{-9}$, so that

$$W_*^{1/4} \lesssim 0.0015 \times m_P \simeq 2 \times 10^{16}\,\mathrm{GeV}. \tag{3.80}$$

On the other hand, this energy scale must be larger than the reheating temperature after inflation. This limits the number of e-folds N_* of inflationary expansion after Hubble exit of the present Hubble scale H_0^{-1}. This is roughly the same number of e-folds it takes after inflation until the present.

Let us briefly derive this statement. We denote the Hubble parameter during inflation, which is nearly constant, by H_i and neglect its slight time dependence. We consider a comoving wave number k which exits the Hubble scale at the value a_1 of the scale factor during inflation, when its wavelength is $\lambda = \lambda_1 = a_1/k = H_i^{-1}$. It re-enters after inflation at $a = a_2$, when its wavelength is $\lambda = \lambda_2 = a_2/k = H_2^{-1}$. The value of the scale factor at the end of inflation is denoted by a_f, see Fig. 3.1. The number of e-folds of inflation after a_1 is $N_1 = \ln(a_f/a_1)$, while the number of e-folds of expansion after inflation until λ re-enters the Hubble scale is $N_2 = \ln(a_2/a_f)$. In the radiation era after inflation $H \propto 1/\tau \propto 1/a^2$. If the scale re-enters during the radiation era we therefore obtain $(a_2/a_f)^2 = H_i/H_2 = H_i a_2/k$. On the other hand, $H_i = 1/\lambda_1 = k/a_1$, so that $H_i a_2/k = a_2/a_1$. Inserting this above yields $a_2^2/a_f^2 = a_2/a_1$ or, equivalently, $a_2/a_f = a_f/a_1$, which implies $N_1 = N_2$.

For a scale which enters only during the matter era we have to correct this result, since after matter and radiation equality the scale factor behaves as $a \propto \tau^{2/3}$ so that $H \propto 1/\tau \propto 1/a^{3/2}$. Denoting the redshift of matter–radiation equality by $z_{\rm eq}$, this leads to a correction factor $\sqrt{z_{\rm eq}/z_2}$, where $z_2 < z_{\rm eq}$ is the redshift at re-entry. The number of e-folds of inflationary expansion after horizon exit, N_1 is therefore related to the number of e-folds of expansion from the end of inflation until re-entry, by

$$N_1 \simeq \begin{cases} N_2 & \text{if } k \text{ re-enters in the radiation era, } z_2 > z_{\rm eq} \\ N_2 - \frac{1}{2}\ln(z_{\rm eq}/z_2) & \text{if } k \text{ re-enters in the matter era, } z_2 < z_{\rm eq}\,. \end{cases}$$

Neglecting the correction term $\frac{1}{2}\ln(z_{\rm eq}/z_2)$ which is never more than a few and denoting the reheating temperature by T_R, we therefore obtain the following limit for the number of e-folds of inflation after exit of the present Hubble scale,

$$N_* = \ln\left(\frac{T_R}{T_0}\right) \leq \ln\left(\frac{W_*^{1/4}}{T_0}\right) \leq \ln\left(\frac{2\times 10^{16}}{2.4\times 10^{-13}}\right) \simeq 66\,. \tag{3.81}$$

Note that this is a conservative estimate and the reheating temperature is most probably significantly lower.

Consistency relation

We have obtained the following results for the scalar and tensor power spectra induced during slow roll inflation

$$n - 1 = -6(\epsilon_1 + \epsilon_2)\,, \tag{3.82}$$

$$n_T = -2\epsilon_1\,, \tag{3.83}$$

$$\frac{P_H}{P_\zeta} = 4\epsilon_1 = -2n_T\,, \quad k/\mathcal{H} \ll 1\,. \tag{3.84}$$

Using the relation (2.143) between ζ and $\Psi = \Phi$ in the radiation and in the matter dominated era, we find that on large scales, where $\dot\Psi = 0$, ζ and Ψ differ only by a constant factor,

$$P_\Psi = \frac{4}{9}P_\zeta\,, \quad k/\mathcal{H} \ll 1\,, \quad \text{(radiation dominated era)}\,, \tag{3.85}$$

$$P_\Psi = \frac{9}{25}P_\zeta\,, \quad k/\mathcal{H} \ll 1\,, \quad \text{(matter dominated era)}\,. \tag{3.86}$$

Writing $k^3 P_\Psi = A_S(k/\mathcal{H}_0)^{n-1}$ for scalar perturbations and $k^3 P_H = A_T(k/\mathcal{H}_0)^{n_T}$ for tensor perturbations on super-Hubble scales, the relation (3.84) implies

$$\frac{A_T}{A_S} = \frac{100}{9}\epsilon_1 = -\frac{50}{9}n_T\,. \tag{3.87}$$

Equations (3.84) or equivalently (3.87) are often also called the consistency relation of slow roll inflation. It is one of the major goals of forthcoming CMB observations to measure tensor perturbations in order to test this relation which holds for both, slow roll and power law inflation, but might be violated if inflation occurred in several stages or not at all.

3.3 Mixture of dust and radiation revisited

In this section we want to study the perturbation of a mixture of dust and radiation in more detail. We shall find that the system has two regular perturbation modes which we can identify with the adiabatic and an iso-curvature mode. We determine the solutions on super-Hubble scales for both modes explicitly and discuss the implications for CMB anisotropies, especially for the positions of the acoustic peaks.

After inflation and reheating the Universe is radiation dominated. Only very much later, at redshift $z < 10^4$ do dark matter and baryons start dominating. It may also be that a scalar field (called quintessence) plays a certain subdominant role, but we neglect this possibility here. Curvature and a cosmological constant are certainly negligible at early times and we thus consider a mixture of radiation and matter only. As in Section 2.4.3 we define

$$R = \frac{\rho_r}{\rho} = \frac{\rho_r}{\rho_r + \rho_m}\,, \quad a = \frac{\rho_m}{\rho_r} = \frac{1-R}{R}\,. \tag{3.88}$$

The scale factor is normalized to unity at equality, $\rho_m(t_{\rm eq}) = \rho_r(t_{\rm eq}) = \rho(t_{\rm eq})/2$. Also note that by definition $0 \le R \le 1$ and $R \simeq 1$ during the radiation era, while $R \simeq 0$ in the matter era. With $w_r = c_r^2 = \frac{1}{3}$ and $w_m = c_m^2 = 0$ we obtain the following useful relations (see also Section 2.4.3 and Ex. 1.3)

$$\frac{\rho_m}{\rho} = 1 - R\,, \tag{3.89}$$

$$w = \frac{\rho_r/3}{\rho} = \frac{1}{3}R\,, \tag{3.90}$$

$$c_s^2 = \frac{\dot\rho_r/3}{\dot\rho} = \frac{\frac{4}{3}R}{4R + 3(1-R)} = \frac{4R}{3(R+3)}\,. \tag{3.91}$$

Integrating the Friedmann equation,

$$\left(\frac{\dot a}{a}\right)^2 = \frac{4\pi G}{3}\rho_{\rm eq}\left(a^{-1} + a^{-2}\right)\,, \tag{3.92}$$

we obtain the scale factor

$$a(t) = \left(\frac{t}{t_1}\right)^2 + 2\left(\frac{t}{t_1}\right) \quad \text{where} \quad t_1 \equiv \sqrt{\frac{3}{\pi G \rho_{\rm eq}}}\,. \tag{3.93}$$

The normalization $a(t_{\rm eq}) = 1$ yields $t_{\rm eq} = (\sqrt{2} - 1)t_1$. In terms of t_1, Eq. (3.92) leads to the following useful expression for the Hubble parameter

$$\mathcal{H}^2 = \frac{4(1+a)}{t_1^2 a^2} \,. \tag{3.94}$$

The radiation/matter mixture has no anisotropic stresses and no intrinsic entropy perturbations. Equation (2.124) then leads to

$$\Gamma = \Gamma_{\rm rel} = \frac{(\rho_r + P_r)\rho_m}{3w(w+1)\rho^2} S_{rm} = \frac{4(1-R)}{3+R} S_{rm} \,.$$

The perturbation equations (2.137) and (2.138) for dust and radiation become, with $S \equiv S_{rm}$,

$$\dot{S} = -kV_{rm} \,, \tag{3.95}$$

$$k\dot{V}_{rm} + \frac{4R}{3+R}\mathcal{H}kV_{rm} = \frac{k^2}{3+R}D + \frac{k^2(1-R)}{3+R}S \,. \tag{3.96}$$

This is equivalent to the second-order equation

$$\ddot{S} + \frac{4R}{R+3}\mathcal{H}\dot{S} = -\frac{k^2}{3+R}\left[(1-R)S + D\right] \,. \tag{3.97}$$

In addition we have the second-order equation (2.119) for D. Using our expressions for w, c_s^2, Γ and $\Pi = K = 0$ we obtain

$$\ddot{D} + \frac{3 - R - 2R^2}{R+3}\mathcal{H}\dot{D} - \frac{9 + 3R + 5R^2 - R^3}{2(R+3)}\mathcal{H}^2 D + \frac{4R}{3(R+3)}k^2 D = -\frac{4R(1-R)}{3(R+3)}k^2 S \,. \tag{3.98}$$

We now want to transform this equation into a differential equation w.r.t. the variable R. For this we need $\dot{R} = -\dot{a}R^2 = \mathcal{H}R(R-1)$ and

$$\ddot{R} = \dot{\mathcal{H}}(R^2 - R) + \mathcal{H}(2R-1)\dot{R} = \frac{3}{2}\mathcal{H}^2(R-1)^2 R \,.$$

For the second equals sign we made use of $\dot{\mathcal{H}} = -(1+3w)\mathcal{H}^2/2$ (see Eq. (2.111)). A lengthy but straightforward calculation gives

$$\frac{d^2D}{dR^2} + \left[\frac{1}{2R} - \frac{1}{R+3}\right]\frac{dD}{dR} - \frac{9 + 3R + 5R^2 - R^3}{2R^2(1-R)^2(R+3)}D = \frac{-4}{3R(1-R)(R+3)}\left(\frac{k}{\mathcal{H}}\right)^2\left[\frac{1}{1-R}D + S\right]. \tag{3.99}$$

We also transform Eq. (3.97),

$$\frac{d^2S}{dR^2} + \left[\frac{3}{2R} - \frac{1}{1-R} - \frac{1}{R+3}\right]\frac{dS}{dR} = -\frac{1}{R^2(1-R)^2(R+3)}\left(\frac{k}{\mathcal{H}}\right)^2 [D + (1-R)S]\,. \tag{3.100}$$

Equation (3.99) can be simplified by writing it as a differential equation for the variable $\Delta \equiv D(1-R)R^{-3/2}$,

$$\frac{d^2\Delta}{dR^2} + \left[\frac{7}{2R} - \frac{1}{R+3} + \frac{2}{1-R}\right]\frac{d\Delta}{dR} = -\frac{4}{3R(R+3)}\left(\frac{k}{\mathcal{H}}\right)^2 \left[\frac{1}{(1-R)^2}\Delta + R^{-3/2}S\right]. \tag{3.101}$$

We want to study possible initial conditions after a generic inflationary phase and subsequent reheating (rh). We are interested in cosmological scales, $k^{-1}a_0 \sim \mathcal{O}(\text{Mpc})$. But from $\mathcal{H}_0^{-1}a_0 = H_0^{-1} \simeq 3000\,h^{-1}$ Mpc and our expressions for $\mathcal{H}$ and a one finds that $\mathcal{H}(a = 0.1)a_0 \sim \mathcal{O}$ (Mpc). For the last estimate we have used $z_{\rm eq} = a_0 \simeq 3300$ (see Appendix 1). The reheating temperature of the Universe is typically of the order $T_{\rm rh} \sim 10^{10}$ GeV so that with $T_{\rm eq} \sim 1$ eV, we obtain $a_{\rm rh} \sim 10^{-19} \ll 0.1$. Therefore, to study the initial conditions it is sufficient to consider the limit of very long wave perturbations, $k/\mathcal{H} \to 0$. In this limit we may neglect the right-hand sides of Eqs. (3.100) and (3.101) and the equations decouple completely. They can then easily be solved by quadrature leading to

$$\Delta = A_1 R^{-5/2}\left[1 - \frac{25}{9}R + \frac{5}{3}R^2 - \frac{5}{3}R^3\right] + A_2\,, \tag{3.102}$$

$$= A_1 X(R) + A_2\,, \tag{3.103}$$

$$S = B_1\left[3R^{-1/2} - 2\log\left(\frac{1+\sqrt{R}}{1-\sqrt{R}}\right)\right] + B_2\,. \tag{3.104}$$

We now transform Δ back into D and write the solutions in terms of the scale factor. If we just multiply the modes proportional to A_1 and A_2 by the factor $R^{3/2}/(1-R)$, both modes of D are singular. We would like to split the solution D into two modes $D = AU_R + BU_S$ where $U_S = R^{3/2}/(1-R)$ is decaying at late time, $R \to 0$ and $U_R = R^{3/2}/(1-R)X + bU_S$ stays regular at early times, $R \to 1$. This can be achieved by choosing $b = \frac{16}{9}$. In terms of the scale factor, using

$$a = \frac{1-R}{R} \quad \text{and} \quad R = \frac{1}{a+1}\,,$$

we get

$$D = AU_R + BU_S \quad \text{with}\,, \tag{3.105}$$

$$U_R = \frac{5}{3(1+a)} + \frac{(1+a)^2 - \frac{25}{9}(1+a) + \frac{16}{9}(1+a)^{-1/2}}{a}\,, \tag{3.106}$$

$$U_S = \frac{(1+a)^{-1/2}}{a}\,, \tag{3.107}$$

$$S = C_1\left[2\log\left(\frac{\sqrt{a+1}+1}{\sqrt{a+1}-1}\right) - 3\sqrt{a+1}\right] + C_2\,, \tag{3.108}$$

$$= C_1 V_S + C_2\,. \tag{3.109}$$

Developing U_R we obtain

$$U_R(a) \simeq \frac{10}{9}a^2\,,\ \text{if} \quad a \ll 1 \quad \text{and} \quad U_R(a) \simeq a\,,\ \text{if} \quad a \gg 1\,. \tag{3.110}$$

The singular modes behave like $U_S \simeq 1/a$ and $V_S \simeq 2\log(4/a)$ at early times, $a \ll 1$. In the most generic case right after reheating, at $a = a_{\rm rh} \ll 1$, all the modes may have comparable amplitudes so that $|2\log(a_{\rm rh}/4)C_1| \sim |C_2| \simeq |Aa_{\rm rh}^2| \sim |B/a_{\rm rh}|$. Hence $C_1 \ll C_2$ and $B \ll A$. Therefore the U_S and V_S modes cannot be relevant at late times. We neglect them in what follows, setting $B = C_1 = 0$. On super-Hubble scales we hence end up with two possible modes, namely $A \neq 0,\ C_2 = 0$ and $A = 0,\ C_2 \neq 0$. The first is called the 'adiabatic mode' and the second the 'entropy mode'. We shall also be interested in a linear combination of these modes, the so-called 'iso-curvature mode' below.

3.3.1 Adiabatic initial conditions

Let us first consider the adiabatic mode, given by the initial condition

$$D = A\left[\frac{5}{3(1+a)} + \frac{(1+a)^2 - \frac{25}{9}(1+a) + \frac{16}{9}(1+a)^{-1/2}}{a}\right], \quad S = 0 \tag{3.111}$$

on super-Hubble scales, $k/\mathcal{H} \ll 1$. Here $A = A(k)$ is a function of the wave number which determines the spectrum. On sub-Hubble scales, radiation perturbations oscillate while matter perturbations which exert no pressure do not, therefore, adiabaticity, $S = 0$ cannot be maintained. The term 'adiabatic perturbations' is however used for perturbations which have adiabatic initial conditions, i.e. which satisfy $S = 0$ at early times, when $k/\mathcal{H} \ll 0$. From the constraint Einstein equation we

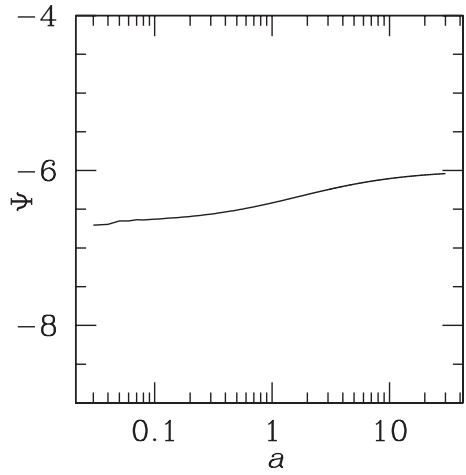


Fig. 3.2. The Bardeen potential (in units of A) for adiabatic perturbations of a mixed radiation and matter fluid on super-Hubble scales as a function of the scale factor normalized to equality, $a(t_{\rm eq}) = 1$.

have $\Psi = -(3/2)(\mathcal{H}/k)^2 D$. With (3.94) this yields

$$\Psi = -\frac{6(1+a)}{(kt_1)^2 a^2} D \,. \tag{3.112}$$

This function is nearly constant in time, see Fig. 3.2. With Eq. (3.110) we find the following asymptotic behaviour in the radiation and matter dominated eras:

$$\Psi = -6\frac{A}{(kt_1)^2} \times \begin{cases} \frac{10}{9} \,, & \text{if} \quad a \ll 1 \\ 1 \,, & \text{if} \quad a \gg 1 \,. \end{cases} \tag{3.113}$$

On super-Hubble scales we therefore have $\Psi \simeq$ constant $\equiv \Psi_0$. If the spectral index is defined as usual, $|\Psi|^2 k^3 \propto k^{n-1}$ we therefore have

$$|A|^2 \propto k^n \,, \quad \text{and} \quad |D|^2 \propto k^n \times \begin{cases} a^4 \,, & \text{if} \quad a \ll 1 \\ a^2 \,, & \text{if} \quad a \gg 1 \,. \end{cases}$$

Let us now evolve forward to the matter era, $R \ll 1$ and $a \gg 1$, but no longer require $k/\mathcal{H} \ll 1$. Neglecting the terms which are subdominant in the matter era, Eq. (3.99) reduces to

$$\frac{d^2 D}{da^2} + \frac{3}{2a}\frac{dD}{da} - \frac{3}{2a^2} D = \frac{(kt_1)^2}{9a^2}[D + S] \,. \tag{3.114}$$

We first consider modes which enter the Hubble scale only in the matter dominated era. For them $kt_1 \ll 1$ and we may always neglect the r.h.s. of Eq. (3.114). The growing mode solutions then also remain $D \simeq Aa$ and $\Psi \simeq \Psi_0$ on sub-Hubble scales.

Energy–momentum conservation for radiation (2.115) now becomes

$$D'_{gr} = -\frac{4}{3}V_r \,, \tag{3.115}$$

$$V'_r = 2\Psi_0 + \frac{1}{4}D_g^{(r)} \,, \tag{3.116}$$

where here a prime denotes a derivative w.r.t. $x \equiv kt$. Now Ψ_0 acts like a constant source term. The general solution of this system is

$$D_g^{(r)} = A\cos\left(\frac{x}{\sqrt{3}}\right) - \frac{4}{\sqrt{3}}B\sin\left(\frac{x}{\sqrt{3}}\right) + 8\Psi_0\left[\cos\left(\frac{x}{\sqrt{3}}\right) - 1\right] \,, \tag{3.117}$$

$$V_r = B\cos\left(\frac{x}{\sqrt{3}}\right) + \frac{\sqrt{3}}{4}A\sin\left(\frac{x}{\sqrt{3}}\right) + 2\sqrt{3}\Psi_0\sin\left(\frac{x}{\sqrt{3}}\right) \,. \tag{3.118}$$

In Ex. 2.3 we show that adiabaticity requires $V_r = V_m$. But in the matter dominated era $V_m \propto t \propto x$ so that

$$\lim_{x\to 0}\frac{V_r}{x} = \lim_{x\to 0}\frac{V_m}{x} = V_0 < \infty \,. \tag{3.119}$$

Therefore, we have to set $B = 0$ and $V_0 = A/4 + 2\Psi_0$. Using in addition $\Psi_0 = 3V_0$ (see (2.181)) we obtain

$$D_g^{(r)} = \frac{4}{3}\Psi_0\cos\left(\frac{x}{\sqrt{3}}\right) - 8\Psi_0 \,, \tag{3.120}$$

$$V_r = \frac{1}{\sqrt{3}}\Psi_0\sin\left(\frac{x}{\sqrt{3}}\right) \,, \tag{3.121}$$

$$D_{gm} = -\Psi_0\left(5 + \frac{1}{6}x^2\right) \,, \tag{3.122}$$

$$V_m = \frac{1}{3}\Psi_0 x \,. \tag{3.123}$$

Here, we have neglected the influence of the radiation perturbations on the matter variable and simply used the pure dust solutions (2.174) and (2.173) for D_{gm} and V_m. On super-Hubble scales, $x \ll 1$ we have

$$D_g^{(r)} \simeq -\frac{20}{3}\Psi_0 \quad \text{and} \quad V_r \simeq \frac{1}{3}x\Psi_0 \,. \tag{3.124}$$

This characterizes adiabatic initial conditions. Up to a constant, the density fluctuations oscillate like a cosine. At $x = 0$, $|D_g^{(r)}|$ has a minimum. The first maximum follows at $x = \sqrt{3}\pi$. This gives rise to the acoustic peak structure discussed in Section 2.6.

3.3.2 Iso-curvature initial conditions

Let us now turn to the 'entropy perturbations'. We first recall that the curvature perturbation in the comoving gauge, ζ, is constant for adiabatic perturbations on super-Hubble scales. We want to calculate it for the entropy perturbation mode. From Eq. (2.146) with vanishing curvature we obtain,

$$\dot{\zeta} = \frac{w}{w+1}\mathcal{H}\Gamma - \frac{2c_s^2}{3(w+1)}k^2\mathcal{H}^{-1}\Psi \ . \tag{3.125}$$

Inserting $\dot{\zeta} = \dot{R}\, d\zeta/dR = -R(1-R)\mathcal{H}\, d\zeta/dR$, $1+w = (3+R)/3$ and $c_s^2 = 4R/[3(3+R)]$, we obtain

$$\frac{d\zeta}{dR} = -\frac{4}{(R+3)^2}S + \frac{8}{3(R+3)^2(1-R)}\left(\frac{k}{\mathcal{H}}\right)^2\Psi \ , \tag{3.126}$$

$$= -\frac{4}{(R+3)^2}S - \frac{4}{(R+3)^2(1-R)}D \ , \tag{3.127}$$

where we have used the (00) Einstein equation in (2.104) for the second identity. For entropy perturbations one has $S = C =$ constant and $D = 0$ on large scales, $k \ll \mathcal{H}$, so that Eq. (3.127) can be integrated to

$$\zeta(R) = \frac{(1-R)}{(3+R)}C \ , \qquad (k/\mathcal{H}) \ll 1 \ . \tag{3.128}$$

Here we have performed the definite integral from 1 to R in order not to add a constant to the result, because such a constant simply represents an adiabatic contribution. This mode therefore satisfies $\zeta \to 0$ and $\Psi \to 0$ for $R \to 1$, i.e., in the radiation dominated era on super-Hubble scales. Equation (2.143) implies

$$D_g = D - 3(1+w)\zeta \ . \tag{3.129}$$

For the iso-curvature mode, $\zeta \to 0$ for $R \to 1$, $0 = D = RD_r + (1-R)D_m$ implies that $D_r \to 0$ and hence also $D_g^{(r)} \simeq D_r \to 0$ for $R \to 1$. Instead of the typical cosine oscillations of the adiabatic mode, we therefore obtain sine oscillations in $D_g^{(r)}$ when the scale $1/k$ enters the Hubble horizon.

As we have seen in Chapter 2, the CMB anisotropies contain a term

$$\frac{\Delta T}{T}(\mathbf{k}, t_0, \mathbf{n}) = \cdots + \frac{1}{4}D_g^{(r)}(\mathbf{k}, t_{\rm dec})\, e^{i\mathbf{k}\mathbf{n}(t_0 - t_{\rm dec})} + \cdots \ . \tag{3.130}$$

On scales where this term dominates, the peaks in $D_g^{(r)}$ translate into peaks in the angular power spectrum of CMB anisotropies.

Since $D_g^{(r)}$ oscillates like a sine for iso-curvature perturbations, we find a first peak in $D_g^{(r)} \propto \sin(c_s k t)$ at

$$x_i^{(0)} = k_i^{(0)} t_{\rm dec} = \frac{1}{c_s}\frac{\pi}{2} , \quad \lambda_i^{(0)} = \frac{\pi}{k_i^{(0)}} = 2c_s t_{\rm dec} ,$$
$$\vartheta_i^{(0)} \simeq \frac{2c_s t_{\rm dec}}{\chi\,(t_0 - t_{\rm dec})} \simeq \frac{2c_s t_{\rm dec}}{t_0} . \tag{3.131}$$

Here $\vartheta_i^{(0)}$ is the angle under which the comoving scale $\lambda_i^{(0)}$ at comoving distance $t_0 - t_{\rm dec}$ is seen, see Eqs. (1.39)–(1.49). Equation (3.131) shows clearly that $\vartheta_i^{(0)}$ strongly depends on the cosmological parameters, especially on curvature. The last $\simeq$ sign above, is only true if $K \simeq 0$.

The position of the acoustic peaks in the CMB anisotropy spectrum therefore presents an excellent means to determine the spatial curvature of the Universe. As we discussed in Chapter 2, when we expand the temperature fluctuations in terms of spherical harmonics, a fluctuation on angular scale ϑ shows up around the harmonic $\ell \sim \pi/\vartheta$. As an indication, we note that for $\Lambda = K = 0$, the harmonic of the first iso-curvature peak is

$$\ell_i^{(0)} \sim \pi/\vartheta_i^{(0)} \sim 110 .$$

In the adiabatic case the corresponding 'first peak' would actually be at $k_a^{(0)} = 0$, but we have not discussed it since it is not visible at all. Since $k = 0$ is of course a super-Hubble scale at recombination, our discussion of the peak structure does not apply at this scale. This is also nearly true for the 'first' peak of the iso-curvature mode. Furthermore, $D_g^{(r)}$ is negative for small x so that these 'first' peaks are under-densities or 'expansion peaks', and due to the gravitational attraction of the baryons (which we have neglected in this simple argument) they are less pronounced than the peaks due to over-densities, called 'compression peaks'.

These 'second' peaks are usually called the first acoustic peaks. They are the first compression peaks. We shall also adopt the convention of calling them 'first peak' for consistency with the literature. They correspond to wavelengths and angular scales

$$\left.\begin{aligned} \lambda_i^{(1)} &= \tfrac{2}{3} c_s t_{\rm dec} , \\ \vartheta_i^{(1)} &\simeq \tfrac{(2/3)c_s t_{\rm dec}}{\chi(t_0 - t_{\rm dec})} , \\ \ell_i^{(1)} &\sim 330 \end{aligned}\right\} \quad \text{(iso-curvature)} , \tag{3.132}$$

$$\left.\begin{aligned} \lambda_a^{(1)} &= c_s t_{\rm dec} , \\ \vartheta_a^{(1)} &\simeq \tfrac{c_s t_{\rm dec}}{\chi(t_0 - t_{\rm dec})} , \\ \ell_a^{(1)} &\sim 220 \end{aligned}\right\} \quad \text{(adiabatic)} . \tag{3.133}$$

Here the indicated harmonic is the one obtained in the case $\Lambda = K = 0$, for a typical baryon density inferred from nucleosynthesis.

It is interesting to note that the distance between consecutive peaks is the same for adiabatic and iso-curvature initial conditions. It is given by

$$\Delta k_i = k_i^{(1)} - k_i^{(0)} = \pi/(c_s t_{\rm dec}) = \Delta k_a \ , \quad \Delta\vartheta = \frac{c_s t_{\rm dec}}{\chi\,(t_0 - t_{\rm dec})} \ , \quad \Delta\ell \sim 200 \ . \tag{3.134}$$

Again, the numerical value indicated for $\Delta\ell$ corresponds to a universe with $\Lambda = K = 0$. The result is strongly dependent, especially on K. This is the reason why the measurement of the peak position (or better of the inter-peak distance) allows an accurate determination of curvature.

From our analysis we can draw the following important conclusions. For scales where the $D_g^{(r)}$-term dominates, the CMB anisotropies show a series of acoustic oscillations with spacing Δk. The position of the first significant peaks is at $k = k_{a/i}^{(1)}$, depending on the initial condition, however the spacing Δk is *independent* of initial conditions.

The angle $\Delta\vartheta$ onto which the scale Δk is projected in the sky is determined entirely by the matter content and the geometry of the Universe. According to our findings in Chapter 1 , ϑ will be larger if $\Omega_K < 0$ (positive curvature) and smaller if $\Omega_K > 0$ (see Fig. 1.4).

In our analysis we have neglected the presence of baryons, in order to obtain simple analytical results. Baryons have two effects: they lead to $(\rho - 3p)_{\rm rad+bar} > 0$, and therefore to an enhancement of the *compression* peaks (the first, third, etc. acoustic peak). In addition, the presence of baryons *decreases* the sound speed c_s of the baryon–photon plasma by about 10%, thereby increasing Δk and $\Delta\ell$ and decreasing $\Delta\vartheta$.

Another point which we have neglected is the fact that the Universe becomes matter dominated at $t_{\rm eq}$, only shortly before decoupling: $t_{\rm dec} \simeq 2.4 t_{\rm eq}$ for $\Omega_m \sim 0.3$. As we have seen, the gravitational potential on sub-Hubble scales is decaying in the radiation dominated era. If the radiation dominated era is not very long before decoupling, the gravitational potential is still decaying slightly and free-streaming photons fall into a deeper gravitational potential than they have to climb out of. This effect, called the 'early integrated Sachs–Wolfe effect' adds to the photon temperature fluctuations at scales which are only slightly larger than the position of the first acoustic peak for adiabatic perturbations. It therefore 'boosts' this peak and, at the same time, moves it to slightly larger scales (larger angles, lower spherical harmonics). Since $t_{\rm eq} \propto h^{-1}$, the first acoustic peak is higher if h is smaller.

A small Hubble parameter therefore *increases* the amplitude of the first acoustic peak. A similar effect is observed if a cosmological constant or negative curvature

are present, since $t_{\rm eq}$ is retarded in those cases. We shall discuss this dependence of the acoustic peak structure on cosmological parameters in detail in Chapter 6.

3.3.3 Mixed adiabatic and iso-curvature perturbation

In general, inflation (from more than one scalar field) can lead to a mixture of adiabatic and iso-curvature perturbations. At early time, $k/\mathcal{H} \ll 1$ and $R \to 1$, such a mixture is given by

$$D = AU_R \quad \text{and} \quad S = C \; . \tag{3.135}$$

Here the 'constants' A and C are random variables for each wave number $\mathbf{k}$. One usually assumes them to be Gaussian, so that all the expectation values are determined by $\langle A(\mathbf{k})A^*(\mathbf{k}')\rangle$, $\langle C(\mathbf{k})C^*(\mathbf{k}')\rangle$ and $\langle A(\mathbf{k})C^*(\mathbf{k}')\rangle$. Statistical homogeneity and isotropy requires

$$\langle A(\mathbf{k})A^*(\mathbf{k}')\rangle = \delta(\mathbf{k}-\mathbf{k}')P_a(k) \; , \tag{3.136}$$

$$\langle C(\mathbf{k})C^*(\mathbf{k}')\rangle = \delta(\mathbf{k}-\mathbf{k}')P_i(k) \; , \tag{3.137}$$

$$\langle A(\mathbf{k})C^*(\mathbf{k}')\rangle = \delta(\mathbf{k}-\mathbf{k}')P_{ai}(k) \; . \tag{3.138}$$

Clearly, $P_{ia}(k) = P_{ai}^*(k)$. Furthermore, Schwarz' inequality requires

$$\left|\langle A(\mathbf{k})C^*(\mathbf{k}')\rangle\right|^2 \leq \langle A(\mathbf{k})A^*(\mathbf{k}')\rangle\langle C(\mathbf{k})C^*(\mathbf{k}')\rangle \; . \tag{3.139}$$

Hence the Hermitean 2×2 matrix (P_{mn}) is positive semi-definite. One calls A and C completely (anti-)correlated if

$$\langle A(\mathbf{k})C^*(\mathbf{k}')\rangle = \pm\sqrt{\langle A(\mathbf{k})A^*(\mathbf{k}')\rangle\langle C(\mathbf{k})C^*(\mathbf{k}')\rangle} \; .$$

To define such generic initial conditions one has, in principle, to specify four real functions, namely $P_a(k)$, $P_i(k)$, $\mathrm{Re}(P_{ai}(k))$ and $\mathrm{Im}(P_{ai}(k))$, which satisfy the inequality (3.139). The present data are fully compatible with purely adiabatic perturbations, $C = 0$. Nevertheless, a considerable iso-curvature contribution of about 10% is still possible. (The precise percentage depends strongly on the definition of the ratio of iso-curvature to adiabatic perturbations, e.g. on the scale at which it is defined, and whether it is the ratio of the CMB anisotropies or of some other perturbation variables.) It is interesting to note that the iso-curvature contribution cannot be severely limited by CMB anisotropy data alone. The above constraint comes mainly from the dark matter spectrum, to which iso-curvature modes contribute very little on large scales. More details can be found in the literature (Trotta *et al.*, 2001, 2003; Moodley *et al.*, 2004). Another possibility for constraining the iso-curvature mode will be CMB polarization data, once they are available with sufficient accuracy (Bucher *et al.*, 2001).

In reality, the situation is even more complicated. The real Universe contains not only photons and dark matter, but also neutrinos and baryons (and maybe quintessence). It has been found recently (Bucher *et al.*, 2000) that a mixture of photons, dark matter, baryons and neutrinos allows five different modes which grow or stay constant, i.e. which are 'regular' in the sense that they do not grow very large into the past. These are the adiabatic mode and the dark matter iso-curvature mode which we have just discussed, a similar baryon iso-curvature mode (where only the baryon density is perturbed) and two neutrino modes (where only the neutrino density or velocity is perturbed). The acoustic peaks from the most generic initial conditions which allow for arbitrary correlations between the different modes are very unpredictable. For example, in a flat universe with a vanishing cosmological constant and fixed cosmic parameters we can obtain a first peak position in the range of $150 \leq \ell^{(1)} \leq 350$. However, combining CMB data and galaxy catalogues (LSS data) allows us to constrain the total contribution from all non-adiabatic modes to less than about 15%. This number will certainly still improve in the future, when accurate polarization data are available.

In the remainder of this book, we only discuss adiabatic perturbations, which are by far the most studied and which are in very good agreement with present data. However, one should keep in mind that all the results, especially those concerning the estimation of cosmological parameters, are not valid if we allow for more generic initial conditions (Bucher *et al.*, 2001; Trotta *et al.*, 2001, 2003; Moodley *et al.*, 2004).

Exercises

(The exercise marked with an asterisk is solved in Appendix A10.3.)

Ex. 3.1 Power law expansion*

Consider a FL universe filled with a (minimally coupled) scalar field with vanishing spatial curvature, $K = 0$. Show that the universe expands like a power law, $a \propto t^q$ so that $\mathcal{H} = q/t$ if and only if the scalar field potential is of the form

$$W(\varphi) = W_0 \exp\left(\alpha \frac{\varphi}{m_P}\right) . \tag{3.140}$$

Determine $\alpha(q)$ and $w(q) = P/\rho$. Determine also $p(q)$ such that $a \propto \tau^p$ and $\alpha(p)$. For which values of α do you obtain an inflationary universe?

Hint: Show that power law expansion implies $w = P/\rho =$ constant. Use this to derive that $\dot{W}/W \propto \mathcal{H}/\dot{\varphi}$ and $\ddot{\varphi} \propto \mathcal{H}\dot{\varphi}$. Replacing $\dot{W}$ and $\ddot{\varphi}$ in the equation of motion, you can now show that $a^2 W \propto \dot{\varphi}^2$. Inserting this in the Friedmann equation leads to $\mathcal{H} \propto \dot{\varphi}$ and therefore to $\dot{W}/W =$ constant.

Ex. 3.2 A canonical variable for tensor perturbations

Consider a spatially flat FL universe with pure tensor perturbations,

$$ds^2 = a^2\left[-dt^2 + (\delta_{ij} + H_{ij})\,dx^i\,dx^j\right] . \tag{3.141}$$

Consider only the gravitational part of the action,

$$S + \delta S = \frac{m_P^2}{16\pi}\int d^4x\,\sqrt{-(g+\delta g)}(R + \delta R) .$$

Show that up to second order in H_{ij} the perturbed action is given by

$$\delta S^{(2)} = \frac{1}{2}\int d^4x\left[\dot{h}_{ij}\dot{h}^{ij} - h_{ij,l}h^{ij,l} + \frac{\ddot{a}}{a}h_{ij}h^{ij}\right] . \tag{3.142}$$

Indices in h_{ij} are raised and lowered with the flat metric δ_{ij} and

$$h_{ij} = \frac{m_p}{\sqrt{8\pi}}aH_{ij} .$$

Hint: Use the variational principle of general relativity, Eq. (A2.15) and insert the first-order expression for G^{ij} given in Eq. (2.109) to obtain $\delta S^{(2)}$. This yields

$$\delta S^{(2)} = -\frac{m_P^2}{8\pi}\int d^4x\,a^2H^{ij}\left[\ddot{H}_{ij} + 2\frac{\dot{a}}{a}\dot{H}_{ij} - \Delta H^{ij}\right] .$$

After a partial integration (subtraction of a total derivative) this leads to Eq. (3.142). We can now Fourier transform h_{ij} and set $h_{ij}(\mathbf{k},t) = e_{ij}(\mathbf{k},+)h_+(\mathbf{k},t) + e_{ij}(\mathbf{k},\times)h_\times(\mathbf{k},t)$, where $e_{ij}(\mathbf{k},+)$ and $e_{ij}(\mathbf{k},\times)$ denote the two polarizations of the gravity wave which satisfy $e^i_i(\mathbf{k},\lambda) = k^i e_{ij}(\mathbf{k},\lambda) = 0$ and $e_{ij}(\mathbf{k},\lambda)e^{ij}(\mathbf{k},\lambda') = \delta_{\lambda\lambda'}$. Calculate $e_{ij}(\mathbf{k},\lambda)$ for $\mathbf{k}$ along the z direction. Show that h satisfies Eq. (3.73) for each of the two polarizations λ.

Ex. 3.3 A mixture of matter and radiation

Consider a mixture of a relativistic fluid, $P_r = (1/3)\rho_r$ and a non-relativistic fluid, with energy density ρ_m and pressure $P_m = 0$ in a Friedmann universe with negligible curvature and cosmological constant. Assume that the fluids are non-interacting. As in Eq. (3.93)

$$R = \frac{\rho_r}{\rho} = \frac{\rho_r}{\rho_r + \rho_m} . \tag{3.143}$$

(i) Show that

$$a = \frac{\rho_r}{\rho_m} = R^{-1} - 1 = \frac{1-R}{R}$$

is the scale factor normalized to $a(t_{\rm eq}) = 1$, where $t_{\rm eq}$ is defined by $\rho_r(t_{\rm eq}) = \rho_m(t_{\rm eq}) \equiv \rho_{\rm eq}$.

(ii) Show that

$$a(t) = \left(\frac{t}{t_1}\right)^2 + 2\left(\frac{t}{t_1}\right) \quad \text{where} \quad t_1 \equiv \sqrt{\frac{3}{\pi G\rho_{\rm eq}}} . \tag{3.144}$$

(iii) Also derive the following useful relations which we have used throughout this chapter:

$$\mathcal{H}^2 = \frac{4(1+a)}{t_1^2 a^2} \,,$$
$$w = \frac{R}{3} \,,$$
$$c_s^2 = \frac{4R}{3(R+3)} \,.$$

Using $\tau_0 = 2/H_0$, $z_{\rm eq} = 2300h^2$ and $z_{\rm dec} = 1090$ calculate t_0, t_1, $t_{\rm eq}$ and $t_{\rm dec}$ in our flat model. Keep the explicit dependence on the Hubble parameter, h, in the expressions.

4

CMB anisotropies

4.1 Introduction to kinetic theory

As we saw in Chapter 1, and as we know from statistical mechanics, the distribution function of photons in thermal equilibrium is given by

$$f(\omega) = \frac{1}{e^{\omega/T} - 1}\,, \tag{4.1}$$

where $\omega = a|\tilde{\mathbf{p}}|$ is the physical photon energy. The comoving photon energy and momentum are denoted by $\tilde{p}^0$ and $\tilde{\mathbf{p}}$ and we have $\tilde{p} = |\tilde{\mathbf{p}}| = \tilde{p}^0 = a^{-1}\omega$. As long as interactions are sufficiently frequent to keep photons in thermal equilibrium, this distribution is maintained. Once there are very few interactions, the distribution is affected only by redshifting photon momenta, this follows from Eq. (1.89) and was discussed in Chapter 1. As we saw there, if we define $T(a) = T_D a_D/a$ after decoupling, where $a(t_D) \equiv a_D$ is the scale factor at decoupling, the distribution retains its form even after decoupling. Of course, after decoupling $T(a)$ is no longer a temperature in the thermodynamical sense but merely a parameter of the distribution function. This point is especially interesting for neutrinos: even if they may have masses of the order of $m_\nu \sim 1\,\text{eV} \gg T_0$, their distribution is an extremely relativistic Fermi–Dirac distribution, since this is what it was at decoupling and it has only changed since by redshifting of neutrino momenta.

4.1.1 Generalities

We first present a brief introduction to relativistic kinetic theory. More details can be found in Ehlers (1971) and Stewart (1971).

In the context of general relativity on a spacetime $\mathcal{M}$, for a particle species with mass m we define the mass-shell, mass-bundle or 1-particle phase space as the part

of tangent space given by

$$P_m \equiv \{(x,\, p) \in T\mathcal{M} \mid g_{\mu\nu}(x) p^\mu p^\nu = -m^2\} \,. \tag{4.2}$$

This is a seven-dimensional subspace of the tangent space $T\mathcal{M}$. A (three-dimensional) 'fibre' of the mass-bundle at a fixed event $x \in \mathcal{M}$ is defined by

$$P_m(x) \equiv \{p \in T_x\mathcal{M} \mid g_{\mu\nu}(x) p^\mu p^\nu = -m^2\} \,. \tag{4.3}$$

Here $T_x\mathcal{M}$ is the tangent space of $\mathcal{M}$ at point $x \in \mathcal{M}$. The 1-particle distribution function is defined on P_m,

$$f : P_m \to \mathbb{R} : (x,\, p) \mapsto f(x,\, p) \,. \tag{4.4}$$

The distribution function is non-negative and represents the phase-space density of particles with respect to the invariant measure $d\mu = 2\delta(p^2 + m^2)|g|\, d^4p\, d^4x$. Here g is the determinant of the metric and $p^2 = g_{\mu\nu}\tilde{p}^\mu \tilde{p}^\nu$. The factor 2 is a convention which we adopt here for convenience. We have chosen the coordinate basis $\partial_\mu = \partial/\partial x^\mu$ in tangent space, so that $p = \tilde{p}^\mu \partial_\mu$. We integrate over p^0 to get rid of the Dirac-δ. This yields the measure $d\mu$ on phase space in terms of the phase space coordinates $(\tilde{p}^i, x^\mu)$,

$$d\mu_m = \frac{|g(x)|}{|\tilde{p}_0(x, \tilde{\mathbf{p}})|}\, d^4x\, d^3\tilde{p} = \sqrt{|g(x)|}\, d\pi_m\, d^4x \,, \qquad \text{where} \tag{4.5}$$

$$d\pi_m = \frac{\sqrt{|g(x)|}}{|\tilde{p}_0(x, \tilde{\mathbf{p}})|}\, d^3\tilde{p} \,. \tag{4.6}$$

Here $\tilde{\mathbf{p}} = (\tilde{p}^1, \tilde{p}^2, \tilde{p}^3)$ and $x = (x^0, x^1, x^2, x^3)$; $\tilde{p}_0 = g_{0\mu}\tilde{p}^\mu$ is determined as a function of $(x, \tilde{\mathbf{p}})$ via the mass-shell condition, $p^2 = -m^2$. The measure $\sqrt{|g(x)|}\, d^4x$ is the usual invariant measure on $\mathcal{M}$. Therefore densities on spacetime are obtained by integrating over the momenta with the measure $d\pi$. For example the particle flux density is given by

$$n^\mu(x) = \int_{P_m(x)} \frac{\sqrt{|g(x)|}}{|\tilde{p}_0(x, \tilde{\mathbf{p}})|} \frac{\tilde{p}^\mu}{\tilde{p}^0} f(x, \tilde{\mathbf{p}})\, d^3\tilde{p} \,. \tag{4.7}$$

More importantly, the energy–momentum tensor is given by

$$T^{\mu\nu}(x) = \int_{P_m(x)} \frac{\sqrt{|g(x)|}}{|\tilde{p}_0(x, \tilde{\mathbf{p}})|} \tilde{p}^\mu \tilde{p}^\nu f(x, \tilde{\mathbf{p}})\, d^3\tilde{p} \,. \tag{4.8}$$

If the particles are non-interacting, they move along geodesics,

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\alpha} \dot{x}^\nu \dot{x}^\alpha = 0 \,. \tag{4.9}$$

Here the dot denotes the derivative with respect to proper time s defined by the condition $g_{\mu\nu}(x)\dot{x}^\mu \dot{x}^\nu = \dot{x}^2 = -1$. In the case of massless (light-like) particles, the

arc length cannot be defined. In this case the dot can be the derivative with respect to some arbitrary affine parameter. The geodesic equation (4.9) for massless particles, $\dot{x}^2 = 0$, is invariant under affine reparametrizations, $s \to As + B$, where A and B are constants.

Equation (4.9) is obtained as the Euler–Lagrange equation of the Lagrangian

$$\mathcal{L}(x, \dot{x}) = \frac{m}{2} g_{\mu\nu}(x) \dot{x}^\mu \dot{x}^\nu \, .$$

For massive particles m denotes the mass, for massless particles it is an arbitrary non-vanishing constant normally set to one. The canonical momentum is then given by

$$\tilde{p}_\mu = \frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = m \dot{x}_\mu \quad \text{and} \quad \tilde{p}^\mu = m \dot{x}^\mu \, .$$

From the geodesic equation (4.9) we therefore have

$$m \dot{\tilde{p}}^\mu = -\Gamma^\mu_{\nu\alpha} \tilde{p}^\alpha \tilde{p}^\nu \, .$$

If there are no collisions, i.e., no interactions other than gravity, the distribution function remains constant in a 'comoving' volume element of phase space. Therefore

$$\frac{d}{ds} f = \left[\dot{x}^\mu \partial_\mu + \dot{\tilde{p}}^i \frac{\partial}{\partial \tilde{p}^i} \right] f = 0 \, , \tag{4.10}$$

$$\leftrightarrow \quad \left[\tilde{p}^\mu \partial_\mu - \Gamma^i_{\mu\nu} \tilde{p}^\mu \tilde{p}^\nu \frac{\partial}{\partial \tilde{p}^i} \right] f = 0 \, . \tag{4.11}$$

This is the Liouville equation for collisionless particles. If collisions cannot be neglected, we have to replace the right-hand side by a collision term. Since collisions involve more than one particle, in principle the collision term depends on the 2- or even 3- and 4-particle distribution functions. To continue, one then has to derive an equation of motion for the 2-particle distribution function and so forth. This leads to the well known BBGKY (Bogoliubov–Born–Green–Kirkwood–Yvon) hierarchy of equations. Often, if the particles are sufficiently diluted, the 2-particle distribution function can be approximated by the product of the 1-particle distribution functions,

$$f_2(x, y, \mathbf{p}_x, \mathbf{p}_y) \simeq f(x, \mathbf{p}_x) f(y, \mathbf{p}_y) \, . \tag{4.12}$$

This corresponds to the assumption that the particle positions in phase space are uncorrelated and is called 'molecular chaos'. In this case, the collision term becomes an integral over the momentum of the colliding particle and we obtain the Boltzmann equation,

$$\left[\tilde{p}^\mu \partial_\mu - \Gamma^i_{\mu\nu} \tilde{p}^\mu \tilde{p}^\nu \frac{\partial}{\partial \tilde{p}^i} \right] f = C[f] \, . \tag{4.13}$$

The collision integral $C[f]$ depends on the details of the interactions. We will calculate it for Thomson scattering of electrons and photons.

What we have discussed so far remains valid in the context of general relativity under some conditions on the number of collisions within a small volume which have to be satisfied in order for a coordinate-independent collision integral to exist (Ehlers, 1971).

In the kinetic approach it is often very useful to use a tetrad basis of vector fields, $e_\mu(x) = e_\mu^\nu \partial_\nu$ with $g(e_\mu, e_\nu) = g_{\alpha\beta} e_\mu^\alpha e_\nu^\beta = \eta_{\mu\nu}$. Here $\eta_{\mu\nu}$ denotes the flat Minkowski metric. With respect to such an orthonormal basis, $p = p^\mu e_\mu$ we have $|p_0| = |p^0| = \sqrt{m^2 - \mathbf{p}^2}$, where $\mathbf{p}^2 = \sum_{i=1}^3 (p^i)^2$, and $d\pi_m = d^3p/|p^0|$, as in flat Minkowski spacetime. This can also be written as

$$\eta_{\mu\nu} p^\mu p^\nu = g_{\mu\nu} \tilde{p}^\mu \tilde{p}^\nu \ .$$

4.1.2 Liouville's equation in a FL universe

We now want to discuss the Liouville equation in a FL universe. We choose the tetrad basis (orthonormal basis of four vector fields)

$$e_0 = a^{-1} \partial_t \quad \text{and} \quad e_i = a^{-1} \epsilon_i \ , \tag{4.14}$$

where (ϵ_i) is an orthonormal basis of vector fields for the metric of the 3-space of constant curvature γ_{ij}. If $K = 0$ we can choose $\epsilon_i = \partial_i$. But also if $K \neq 0$ we can always choose vector fields (ϵ_i) which form an orthonormal basis, i.e., a basis which satisfies

$$\gamma(\epsilon_i, \epsilon_j) = \delta_{ij} \ . \tag{4.15}$$

The expression for the energy–momentum tensor (with respect to the usual coordinate basis ∂_μ) in a Friedmann universe becomes

$$T^{\mu\nu}(x) = a^4 \sqrt{|\gamma(x)|} \int_{P_m(x)} \frac{1}{|\tilde{p}_0|} \tilde{p}^\mu \tilde{p}^\nu f(x, \tilde{\mathbf{p}}) \, d^3\tilde{p} \ , \tag{4.16}$$

where γ is the determinant of the three-dimensional metric (γ_{ij}) and we have used $|g| = a^8 |\gamma|$.

The Liouville equation in a Friedmann universe in terms of the coordinates $(x^\mu, \tilde{p}^i)$, is given by

$$\tilde{p}^\mu \partial_\mu f|_{\tilde{p}} - \Gamma^i_{\mu\nu} \tilde{p}^\mu \tilde{p}^\nu \frac{\partial f}{\partial \tilde{p}^i} = 0 \ . \tag{4.17}$$

Here we write $\partial_\mu f|_{\tilde{p}}$ in order to indicate that the components $\tilde{p}^i$ are fixed when the derivative w.r.t. x^μ is taken. Next we transform Eq. (4.17) into an equation for

f with respect to the new coordinates (x^μ, p^i), i.e., we consider f as a function of (x^μ, p^i). Since the FL universe is isotropic, f depends on the momentum only via[1] $p = \sqrt{\delta_{ij} p^i p^j} = \sqrt{a^2 \gamma_{ij} \tilde{p}^i \tilde{p}^j} = a\tilde{p}$. The derivative of the distribution function with respect to t or $\mathbf{x}$ depends on the momentum variable which we keep constant when performing this derivative. We denote by $\partial_\mu f|_p$ the derivative w.r.t. x^μ while keeping constant the momentum components p^i w.r.t. the orthonormal basis $\mathbf{e}_i$ and $\partial_\mu f|_{\tilde{p}}$ the derivative w.r.t. x^μ while keeping constant the momentum components $\tilde{p}^i$ w.r.t. the coordinate basis ∂_i. We then have

$$\partial_0 f|_{\tilde{p}} = \partial_0 f|_p + (\partial_0 p^j)|_{\tilde{p}} \frac{\partial f}{\partial p^j} = \partial_0 f|_p + \mathcal{H} p \frac{\partial f}{\partial p}\,, \tag{4.18}$$

$$\partial_i f|_{\tilde{p}} = \partial_i f|_p + (\partial_i p)|_{\tilde{p}} \frac{\partial f}{\partial p} = \partial_i f|_p + a^2 \frac{\tilde{p}^k \tilde{p}^j \gamma_{kj,i}}{2p} \frac{\partial f}{\partial p}\,,$$

$$\tilde{p}^i \partial_i f|_{\tilde{p}} = \tilde{p}^i \partial_i f|_p + a^2 \frac{\tilde{p}^k \tilde{p}^j \tilde{p}^i \gamma_{kj,i}}{2p} \frac{\partial f}{\partial p}\,, \tag{4.19}$$

$$\frac{\partial f}{\partial \tilde{p}^i} = \frac{\partial p}{\partial \tilde{p}^i} \frac{\partial f}{\partial p} = a^2 \frac{\gamma_{im} \tilde{p}^m}{p} \frac{\partial f}{\partial p}\,. \tag{4.20}$$

With Eq. (4.20) the terms with spatial Christoffel symbols in Eq. (4.17) become

$$\Gamma^i_{jm} \tilde{p}^m \tilde{p}^j \frac{\partial f}{\partial \tilde{p}^i} = a^2 \Gamma^i_{jm} \gamma_{ik} \frac{\tilde{p}^m \tilde{p}^j \tilde{p}^k}{p} \frac{\partial f}{\partial \tilde{p}} = a^2 \frac{1}{2} \gamma_{mj,k} \frac{\tilde{p}^m \tilde{p}^j \tilde{p}^k}{p} \frac{\partial f}{\partial \tilde{p}}\,. \tag{4.21}$$

In the last equals sign we have used the fact that $\tilde{p}^m \tilde{p}^j \tilde{p}^k$ is symmetrical in the indices m, k, j and we may therefore also symmetrize the term $\Gamma^i_{jm}\gamma_{ik} = 1/2(\gamma_{kj,m} + \gamma_{km,j} - \gamma_{jm,k})$. With the help of Eq. (4.19) the terms $-\Gamma^i_{jk} \tilde{p}^j \tilde{p}^k (\partial f/\partial \tilde{p}^i)$ and $\tilde{p}^i (\partial_i p)|_{\tilde{p}} (\partial f/\partial p)$ in Eq. (4.17) cancel and we obtain

$$\tilde{p}^\mu \partial_\mu f|_p + \mathcal{H} \tilde{p}^0 p \frac{\partial f}{\partial p} - 2\Gamma^i_{0j} \tilde{p}^0 \tilde{p}^j \frac{\partial f}{\partial \tilde{p}^i} = 0\,. \tag{4.22}$$

Inserting the Christoffel symbols of the FL universe (see Appendix A2.3) $\Gamma^i_{0j} = \Gamma^i_{j0} = \mathcal{H}\delta^i_j$, we find

$$\tilde{p}^\mu \partial_\mu f - \mathcal{H} \tilde{p}^0 p \frac{\partial f}{\partial p} = 0\,. \tag{4.23}$$

In an unperturbed FL universe we assume the distribution function to be homogeneous and isotropic, hence to depend on x^j and on p^i only via p. When using the coordinates p^i in momentum space we therefore expect f not to depend on the spatial coordinates x^i anymore. Therefore, the Liouville equation simplifies

[1] Here we use p to denote the absolute value of the physical momentum while before we used it to denote the 4-vector p. Since these are very different objects we hope that there is no danger of confusion.

further to

$$\partial_0 f - \mathcal{H} p \frac{\partial f}{\partial p} = 0 \,. \tag{4.24}$$

Or, setting $v = ap$ so that $\partial_0 f|_v = \partial_0 f|_p - \mathcal{H} p(\partial f/\partial p)$ and interpreting f as a function of (t, v), we obtain simply

$$\partial_0 f(t, v) = 0 \,. \tag{4.25}$$

The Liouville equation in a FL universe therefore just requires that the distribution function of collisionless particles changes in time only by redshifting of the physical momentum p and therefore is simply a function of the redshift corrected momentum $v = ap$. Normally we shall use the same letter f for $f(t, p)$ and $f(v)$.

4.2 The Liouville equation in a perturbed FL universe

Let us consider small (first-order) deviations from a FL universe. This does not only imply a small perturbation of the distribution function, but also its domain of definition, the mass-shell (4.2) is modified due to the perturbations of the metric. We will keep track of this by modifying the tetrad fields (e_μ).

4.2.1 Scalar perturbations

We derive the linear perturbation of Liouville's equation in the longitudinal gauge. The perturbed metric is given by

$$ds^2 = -a^2(1 + 2\Psi)\, dt^2 + a^2(1 - 2\Phi)\gamma_{ij}\, dx^i\, dx^j \,. \tag{4.26}$$

The perturbed distribution function is $f = \bar{f}(v) + F^{(S)}(x^\mu, v, \theta, \phi)$, where (θ, ϕ) define the direction of the momentum $\mathbf{p}$. Liouville's equation now becomes, to first order, in the perturbations

$$\tilde{p}^\mu \partial_\mu f - \bar{\Gamma}^i_{\mu\nu} \tilde{p}^\mu \tilde{p}^\nu \frac{\partial f}{\partial \tilde{p}^i} - \delta\Gamma^i_{\mu\nu} \tilde{p}^\mu \tilde{p}^\nu \frac{\partial \bar{f}}{\partial \tilde{p}^i} = 0 \,, \tag{4.27}$$

where the perturbations of the Christoffel symbols are given in Appendix 3, Eqs. (A3.2)–(A3.5). We have denoted background quantities by an over-bar. For simplicity, and also since this is the most relevant case, we restrict ourselves here to $K = 0$. The curved cases, $K \neq 0$ are treated in Appendix 9. However, in order to connect these results in a more straightforward manner to the $K \neq 0$ case, we do not yet make a Fourier decomposition of Φ, Ψ and $F^{(S)}$. We again use a tetrad basis which is now given by

$$e_0 = a^{-1}(1 - \Psi)\partial_t \quad \text{and} \quad e_i = a^{-1}(1 + \Phi)\partial_i \,. \tag{4.28}$$

We want to transform Eq. (4.27) to the coordinates (x^μ, p^i) with $p^\mu e_\mu = \tilde{p}^\mu \partial_\mu$. So that

$$p^0 = a(1+\Psi)\tilde{p}^0 \quad \text{and} \quad p^i = a(1-\Phi)\tilde{p}^i \; . \tag{4.29}$$

For the transformation we use the derivatives

$$\begin{aligned}
\partial_0 p^i|_{\tilde{p}} &= \left[\mathcal{H}(1-\Phi) - \dot{\Phi}\right] a\tilde{p}^i \quad \text{so that ,} \\
\partial_0 f|_{\tilde{p}} &= \partial_0 f|_p + [\mathcal{H}(1-\Phi) - \dot{\Phi}] a\tilde{p}^i \frac{\partial f}{\partial p^i} \; , \\
\partial_0 f|_{\tilde{p}} &= \partial_0 f|_p + [\mathcal{H} - \dot{\Phi}] p \frac{\partial f}{\partial p} \; , \qquad (4.30) \\
\tilde{p}^j \partial_j f|_{\tilde{p}} &= \tilde{p}^j \partial_j f|_p - \tilde{p}^j \partial_j \Phi p \frac{\partial \bar{f}}{\partial p}. \qquad (4.31)
\end{aligned}$$

As in the previous section, we indicate the momentum variable kept constant. With the help of the Liouville equation for $\bar{f}$, we then find

$$\begin{aligned}
&\tilde{p}^\mu \partial_\mu F^{(S)}\big|_p - \mathcal{H}\tilde{p}^0 p \frac{\partial F^{(S)}}{\partial p} \\
&\quad = a^{-1} v \frac{d\bar{f}}{dv} [p^i \partial_i \Phi + p^0 \dot{\Phi}] + a^{-1} \delta\Gamma^i_{\mu\nu} p^\mu p^\nu \frac{\partial \bar{f}}{\partial p^i} \; . \qquad (4.32)
\end{aligned}$$

Inserting the perturbation of the Christoffel symbols (Eqs. (A3.2)–(A3.5) of Appendix 3), the right-hand side becomes

$$a^{-1} v \frac{d\bar{f}}{dv} \left[-p^0 \dot{\Phi} + \frac{(\tilde{p}^0)^2}{\tilde{p}^2} p^k \partial_k \Psi \right] \; ,$$

where $\tilde{p}^2 = \sum_k (\tilde{p}^k)^2$ and we have used $p^i(\partial \bar{f}/\partial p^i) = v(d\bar{f}/dv)$.

We now rewrite the Liouville equation in terms of a new variable defined by $\mathcal{F} = F^{(S)} + \Phi v(d\bar{f}/dv)$. In most of the literature (Hu & Sugiayma, 1995; Hu *et al.*, 1995, 1998; Hu & White, 1997a, 1997b) the variable $F^{(S)}$ is used directly. Note, however, that $\mathcal{F}$ and $F^{(S)}$ only differ by an isotropic (direction-independent) term. Hence, once we determine the CMB anisotropies this difference will only be present in the unmeasurable monopole term. The advantage of the variable $\mathcal{F}$ will become clear later.

Setting $\tilde{p}^j = \tilde{p} n^j$ with $1 = \delta_{ij} n^i n^j$ we have to lowest order, $\tilde{p} = p/a = v/a^2$. Defining also

$$q = a^2 \tilde{p}^0 = a\omega = a\sqrt{p^2 + m^2} = \sqrt{v^2 + a^2 m^2} \; , \tag{4.33}$$

we can rewrite the Liouville equation for the function $\mathcal{F}(t, \mathbf{x}, v, \mathbf{n})$ in the form

$$q\partial_0\mathcal{F} + vn^i\partial_j\mathcal{F} = n^i\partial_i\left[q^2\Psi + v^2\Phi\right]\frac{d\bar{f}}{dv}\,. \tag{4.34}$$

Here $v^j = ap^j$ are the redshift corrected physical momentum components and $\mathcal{F}$ is understood as a function of the variables x^μ and $v^j \equiv vn^j$. Since $\mathcal{F}$ and Φ , Ψ are already perturbations, we can use the background relations between $\mathbf{p}$ and $\mathbf{v}$ as well as q.

This is the Liouville equation for collisionless (massive) particles. We shall see that the equation can be simplified in the massless case where $q = v$, which is relevant for the study of photons.

4.2.2 Vector perturbations

Next we consider vector perturbations. For simplicity, here we do not use the vector gauge, but we set $B^i = 0$ so that

$$ds^2 = a^2\left(-dt^2 + (\gamma_{ij} + 2H_{ij})\,dx^i\,dx^j\right), \quad 2H_{ij} = H_{i|j} + H_{j|i}\,. \tag{4.35}$$

We use this gauge instead of the vector gauge, because it has a simpler perturbed orthonormal basis. The vector fields

$$e_0 = a^{-1}\partial_t \quad \text{and} \quad e_i = a^{-1}\left(\delta_i^j - H_i^{\,j}\right)\partial_j\,, \tag{4.36}$$

are orthonormal. If we used a gauge with $B^i \neq 0$ (non-vanishing 'shift vector') this would lead to a mixing of time and space directions in the orthonormal basis and would complicate the calculations.

In the chosen basis the components of $p = \tilde{p}^\mu\partial_\mu = p^\mu e_\mu$ are related by

$$\begin{aligned}
\tilde{p}^0 &= a^{-1}p^0\,,\\
\tilde{p}^i &= a^{-1}p^j(\delta_j{}^i - H_j{}^i)\,,\\
p^0 &= a\tilde{p}^0\,,\\
p^i &= a\tilde{p}^j(\delta_j{}^i + H_j{}^i)\,.
\end{aligned}$$

The indices of H_{ij} are raised and lowered with the trivial metric $\delta_j{}^i$. In the gauge chosen in Eq. (4.35), the only non-vanishing perturbations of the Christoffel symbols are

$$\delta\Gamma^i_{j0} = \dot{H}^i{}_j\,, \quad \delta\Gamma^i_{jm} = H^i{}_{j|m} + H^i{}_{m|j} - H_{jm}{}^{|i}\,, \tag{4.37}$$

where in the spatially flat case, $|$ is simply the ordinary partial derivative. Again, we want to write the Liouville equation $\tilde{p}^\mu\partial_\mu f - \Gamma^i_{\alpha\beta}\tilde{p}^\alpha\tilde{p}^\beta(\partial f/\partial\tilde{p}^i)$ for f as a function of (x^μ, p^i). The difference to the scalar case comes from the different basis

and hence the difference in the relation between p^μ and $\tilde{p}^\mu$ and from the different Christoffel symbols. A short calculation gives for $f = \bar{f}(v) + F^{(V)}(t, \mathbf{x}, v, \mathbf{n})$

$$\begin{aligned}
\tilde{p}^i \partial_i f|_{\tilde{p}} &= a^{-1} p^i \left[\partial_i f|_p + \partial_i p^j|_{\tilde{p}} \frac{\partial f}{\partial p^j} \right] = a^{-1} p^i \left[\partial_i F^{(V)}|_p \; + p^k H^j_{k|i} \frac{\partial \bar{f}}{\partial p^j} \right] \\
&= a^{-1} p^i \left[\partial_i F^{(V)}|_p \; + \frac{p^j p^k}{p} H^j_{k|i} \frac{\partial \bar{f}}{\partial p} \right], \\
\Gamma^i_{jk} \tilde{p}^j \tilde{p}^k \frac{\partial f}{\partial \tilde{p}^i} &= a^{-1} \frac{p^i p^j}{p} p^k H^i_{k|j} \frac{\partial \bar{f}}{\partial p} \, .
\end{aligned}$$

Here we have used that the background contribution to f, $\bar{f}$ depends on momentum only via p so that $\partial \bar{f}/\partial p^j = (p^j/p)(\partial \bar{f}/\partial p)$. The other terms of the Liouville equation are

$$\begin{aligned}
\partial_0 f|_{\tilde{p}} &= \partial_0 f|_p + \partial_0 p^i|_{\tilde{p}} \frac{\partial f}{\partial p^i} \\
\partial_0 p^i &= \mathcal{H} a \tilde{p}^j (\delta^i_j + H_j{}^i) + a \tilde{p}^j \dot{H}^i_j = \mathcal{H} p^i + a \tilde{p}^j \dot{H}^i_j \\
\tilde{p}^0 \partial_0 f|_{\tilde{p}} &= a^{-1} \left[p^0 \partial_0 f|_p + p^0 \mathcal{H} p^i \frac{\partial f}{\partial p^i} + \dot{H}_{ij} \frac{p^i p^j}{p} \frac{\partial \bar{f}}{\partial p} \right] .
\end{aligned}$$

Furthermore,

$$2\Gamma^i_{0j} \tilde{p}^j \tilde{p}^0 \frac{\partial f}{\partial \tilde{p}^i} = 2\mathcal{H} p^0 p^i \frac{\partial f}{\partial p^i} + 2\dot{H}_{ij} \frac{p^i p^j}{p} \frac{\partial \bar{f}}{\partial p} \, .$$

Together these results yield

$$\begin{aligned}
&\tilde{p}^\mu \partial_\mu f|_{\tilde{p}} - \Gamma^i_{\mu\nu} \tilde{p}^\mu \tilde{p}^\nu \frac{\partial f}{\partial \tilde{p}^i} \\
&\quad = a^{-1} \left[p^0 \partial_0 f|_p + p^i \partial_i f|_p - \mathcal{H} p^0 p^i \frac{\partial f}{\partial p^i} - p^0 \dot{H}^i_m \frac{p^m p^i}{p} \frac{\partial f}{\partial p} \right] . \qquad (4.38)
\end{aligned}$$

Using the zeroth-order Liouville equation, and transforming to the redshift corrected momentum variable $v = ap$, all this finally leads to the following Liouville equation for $F^{(V)}(t, \mathbf{x}, v, \mathbf{n})$,

$$q \partial_0 F^{(V)} + v n^i \partial_i F^{(V)} = q v n^i n^j \dot{H}_{ij} \frac{d\bar{f}}{dv} \, . \qquad (4.39)$$

where for $B = 0$, $\sigma^{(V)}_{\ell m} = a \dot{H}^{(V)}_{\ell m}$ as defined in Chapter 2, Eq. (2.52).

Note that it was useful to choose the redshift corrected momentum v and the directions n^i as our momentum variables. Otherwise the Liouville equation would be significantly more complicated.

4.2.3 Tensor perturbations

For tensor perturbations the perturbed metric is given by

$$ds^2 = a^2\left(-dt^2 + (\gamma_{ij} + 2H_{ij})\,dx^i\,dx^j\right)\,, \quad H_i^i = H_{i|j}^j = 0\,. \tag{4.40}$$

As above we define the perturbation of the distribution function by

$$f = \bar{f}(v) + F^{(T)}(t, \mathbf{x}, v, \mathbf{n})\,.$$

The situation is exactly the same as for vector perturbations and we find the same Liouville equation,

$$q\,\partial_0 F^{(T)} + vn^i\,\partial_i F^{(T)} = q v n^i n^j \dot{H}_{ij}\frac{d\bar{f}}{dv}\,. \tag{4.41}$$

4.3 The energy–momentum tensor

From the perturbed distribution function and metric, we can determine the perturbed energy–momentum tensor. We start from the general expression

$$T_\nu^\mu(x) = \int_{P_m(x)} \frac{\sqrt{|g(x)|}}{|\tilde{p}_0(x, \tilde{\mathbf{p}})|}\tilde{p}^\mu \tilde{p}_\nu f(x, \tilde{\mathbf{p}})\,d^3\tilde{p}\,. \tag{4.42}$$

Observe that the components $\tilde{p}^\mu$ are the momentum components w.r.t. the coordinate basis ∂_μ. We now use

$$\begin{aligned}
\tilde{p}^0 &= a^{-2}(1-\Psi)q\,,\\
\tilde{p}_0 &= -(1+\Psi)q\,,\\
\tilde{p}^i &= a^{-2}vn^j(\delta_j{}^i - H_j{}^i)\,,\\
\tilde{p}_i &= vn_j(\delta^j{}_i + H^j{}_i)\,.
\end{aligned}$$

Here we consider scalar, vector and tensor perturbations together so that $H_{ij} = -\Phi\delta_{ij} + H_{ij}^{(V)} + H_{ij}^{(T)}$. In the following subsections we then isolate the contributions for the scalar, vector and tensor perturbations of the energy–momentum tensor. We note that to first order $\det g = -a^8(1 + 2\Psi - 6\Phi)$. In order to transform the integration $d^3\tilde{p}$ in Eq. (4.42) into an integration w.r.t. d^3p we use

$$\det\left(\frac{d\tilde{p}}{dp}\right) = a^{-3}\det\left(\delta_m{}^j - H_m{}^j\right) = a^{-3}[1 + 3\Phi]\,. \tag{4.43}$$

With this we find that the metric perturbations in T_0^0 and T_j^0 cancel and we obtain the following expressions for the energy–momentum tensor

$$T_0^0 = -\int_{P_m(x)} p^0 p^2 f(x, \mathbf{p})\,dp\,d\Omega_{\mathbf{n}} = \frac{-1}{a^4}\int q v^2 f\,dv\,d\Omega_{\mathbf{n}}\,, \tag{4.44}$$

$$T_0^{\ j} = -\int_{P_m(x)} n^j p^3 f(x,\mathbf{p})\, dp\, d\Omega_{\mathbf{n}} = \frac{-1}{a^4}\int n^j v^3 f\, dv\, d\Omega_{\mathbf{n}}\,, \tag{4.45}$$

$$\begin{aligned} T_j^{\ i} &= (\delta^i{}_\ell - H^i{}_\ell)(\delta_{jk} + H_{jk})\frac{1}{a^4}\int n^\ell n^k \frac{v^4}{q} f\, dv\, d\Omega_{\mathbf{n}}\,, \\ &= \bar{T}^i{}_j + \frac{1}{a^4}\int n^i n_j \frac{v^4}{q} F\, dv\, d\Omega_{\mathbf{n}} = \bar{P}\delta^i{}_j + \delta T^i{}_j\,. \end{aligned} \tag{4.46}$$

To find the last expression we use the fact that the background stress tensor is diagonal, $\int n^\ell n^k \bar{f}\, d\Omega_{\mathbf{n}} = \frac{4\pi}{3}\delta^{\ell k}\bar{f}$ and that to first order

$$(\delta^{ik} - H^{ik})(\delta_k{}^j + H_k^j) = \delta^{ij}\,,$$

since $H^{km} = H^k{}_m = H_{km}$ is symmetric. In Eq. (4.45) we have neglected the term proportional to $H^\ell{}_m$ since in the direction integral $\int n^j f$ only the perturbation of the distribution function contributes, so that this term would be second order. The surface element $d\Omega_{\mathbf{n}}$ denotes the integral over the sphere of momentum directions, $p^i = pn^i$.

Before turning to the different modes, let us split the stress tensor into a trace and a traceless part,

$$T^i{}_j = P\delta^i{}_j + \bar{P}\Pi^i{}_j \qquad \text{with} \tag{4.47}$$

$$P = \frac{1}{3}T^i{}_i = \bar{P} + \frac{1}{3a^4}\int \frac{v^4}{q} F\, dv\, d\Omega_{\mathbf{n}} \qquad \text{and} \tag{4.48}$$

$$\bar{P}\Pi^i{}_j = T^i{}_j - P\delta^i{}_j = \frac{1}{a^4}\int \frac{v^4}{q}\left(n^i n_j - \frac{1}{3}\delta^i_j\right) F\, dv\, d\Omega_{\mathbf{n}}\,. \tag{4.49}$$

Here we have used the fact that

$$\bar{P} = \frac{4\pi}{3a^4}\int \frac{v^4}{q}\bar{f}\, dv\,.$$

4.3.1 Scalar perturbations

We now use the general expressions above to determine the variables defined in Chapter 2 which specify scalar perturbations of the energy–momentum tensor. We consider a Fourier mode $F^{(S)}(t,\mathbf{k},\mathbf{n})e^{i\mathbf{k}\cdot\mathbf{x}}$. As before, we denote background quantities with an overbar. Equation (4.44) implies

$$\begin{aligned} \rho(t,\mathbf{k}) = \bar{\rho}(t) + \delta\rho^{(\mathrm{long})}(t,\mathbf{k}) &= \frac{4\pi}{a^4}\int q v^2 \bar{f}(v)\, dv \\ &+ \frac{1}{a^4}\int q v^2 F^{(S)}(t,\mathbf{k},\mathbf{n})\, dv\, d\Omega_{\mathbf{n}}\,. \end{aligned} \tag{4.50}$$

Hence

$$D_s = \frac{\delta\rho^{(\mathrm{long})}}{\bar\rho} = \frac{1}{\bar\rho a^4}\int q v^2 F^{(S)}\, dv\, d\Omega_{\mathbf{n}}\,. \tag{4.51}$$

To determine the integral of $\mathcal{F} = F + \Phi v(d\bar f/dv)$ we use

$$\frac{1}{a^4}\int q v^3 \frac{d\bar f}{dv} dv\, d\Omega_{\mathbf{n}} = -\frac{4\pi}{a^4}\int \left(3qv^2 + \frac{v^4}{q}\right)\bar f\, dv = -3(\bar\rho + \bar P)\,.$$

For the first equals sign we have integrated by parts and used

$$\frac{dq}{dv} = \frac{d}{dv}\sqrt{a^2m^2 + v^2} = \frac{v}{q}\,.$$

There is no boundary term since f decays rapidly for large momenta. With this we obtain

$$\frac{1}{\bar\rho a^4}\int q v^2 \mathcal{F}\, dv\, d\Omega_{\mathbf{n}} = D_s - 3(1+w)\Phi = D_g\,. \tag{4.52}$$

The gauge-invariant velocity perturbation is given by T_i^0 in longitudinal gauge. Hence

$$T_i^{0(S)} = \frac{1}{a^4}\int n_i v^3 F^{(S)}\, dv\, d\Omega_{\mathbf{n}} = \frac{1}{a^4}\int n_i v^3 \mathcal{F}\, dv\, d\Omega_{\mathbf{n}} = (\bar\rho + \bar P)V_i\,. \tag{4.53}$$

Taking the divergence on both sides we obtain, with $V_j \equiv -i(k_j/k)V$,

$$\begin{aligned} kV &= \frac{i}{a^4(\bar\rho + \bar P)}\int n^i k_i v^3 \mathcal{F}\, dv\, d\Omega_{\mathbf{n}}\,,\\ V &= \frac{i}{a^4(\bar\rho + \bar P)}\int \mu v^3 \mathcal{F}\, dv\, d\Omega_{\mathbf{n}}\,, \end{aligned} \tag{4.54}$$

where we have introduced the direction cosine between $\mathbf{n}$ and $\mathbf{k}$, $\mu = n^i k_i/k = n^i \hat k_i$.

In order to determine the variable Γ we first write

$$\pi_L = \frac{\delta P}{\bar P} = \frac{1}{3\bar P a^4}\int \frac{v^4}{q} F^{(S)}\, dv\, d\Omega_{\mathbf{n}}\,,$$

and therefore

$$\frac{1}{3\bar P a^4}\int \frac{v^4}{q}\mathcal{F}\, dv\, d\Omega_{\mathbf{n}} = \pi_L + \Phi\frac{4\pi}{3\bar P a^4}\int \frac{v^5}{q}\frac{d\bar f}{dv}\, dv \tag{4.55}$$

We use the background identity $\dot{\bar\rho} = -3\mathcal{H}\bar\rho(1+w) = -3\mathcal{H}\frac{(1+w)}{w}\bar P$ to replace $\bar P$ in the second term. After integration by parts we find

$$\Phi\frac{4\pi}{3\bar P a^4}\int \frac{v^5}{q}\frac{d\bar f}{dv}\, dv = \Phi\frac{\mathcal{H}(1+w)}{w\bar\rho}\frac{4\pi}{a^4}\int\left(\frac{5v^4}{q} - \frac{v^6}{q^3}\right)\bar f\, dv\,.$$

On the other hand, using $\dot{q} = \mathcal{H}m^2a^2/q = \mathcal{H}\frac{q^2-v^2}{q}$, we obtain

$$\dot{\bar{P}} = -\mathcal{H}\frac{4\pi}{3a^4}\int\left(\frac{5v^4}{q} - \frac{v^6}{q^3}\right)\bar{f}\,dv\ .$$

With $\dot{\bar{P}}/\dot{\bar{\rho}} = c_s^2$, these two equations together yield

$$\frac{1}{3\bar{P}a^4}\int\frac{v^4}{q}\mathcal{F}\,dv\,d\Omega_{\mathbf{n}} = \pi_L^{(\mathrm{long})} - 3(1+w)\frac{c_s^2}{w}\Phi\ . \tag{4.56}$$

Furthermore

$$\frac{c_s^2}{\bar{P}a^4}\int qv^2\mathcal{F}\,dv\,d\Omega_{\mathbf{n}} = \frac{c_s^2}{w}D_g = \frac{c_s^2}{w}\delta^{(\mathrm{long})} - 3(1+w)\frac{c_s^2}{w}\Phi\ .$$

Combining this with Eq. (4.56) results in

$$\frac{1}{\bar{P}a^4}\int\left(\frac{v^4}{3q} - c_s^2qv^2\right)\mathcal{F}\,dv\,d\Omega_{\mathbf{n}} = \pi_L^{(\mathrm{long})} - \frac{c_s^2}{w}\delta^{(\mathrm{long})} = \Gamma\ . \tag{4.57}$$

The scalar anisotropic stress tensor is simply given by Eq. (4.49). It is related to its potential Π via $\Pi_{ij} = \left(-k^{-2}k_ik_j + \frac{1}{3}\delta_{ij}\right)\Pi$, so that $\Pi^{ij}{}_{|ij} = \frac{2}{3}k^2\Pi$. In Eq. (4.49) this leads to

$$\Pi = \frac{3}{2a^4\bar{P}}\int\frac{v^4}{q}\left(-(\mathbf{n}\cdot\mathbf{k})^2/k^2 + \frac{1}{3}\right)\mathcal{F}\,dv\,d\Omega_{\mathbf{n}} \tag{4.58}$$

$$= \frac{3}{2a^4\bar{P}}\int\frac{v^4}{q}\left(\frac{1}{3} - \mu^2\right)\mathcal{F}\,dv\,d\Omega_{\mathbf{n}}\ . \tag{4.59}$$

4.3.2 Vector perturbations

Vector perturbations are given by divergence free vector fields. For a fixed Fourier component $\mathbf{k}$, we expand them in the basis functions $Q_j^{(V)}$ which have two independent modes. Let us choose two basis vectors $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ so that $(\mathbf{e}^{(1)},\ \mathbf{e}^{(2)},\ \hat{\mathbf{k}})$ form an orthonormal basis. The $\mathbf{k}$-Fourier mode of an arbitrary vector perturbation is then of the form $A_j = (A^{(1)}\mathbf{e}_j^{(1)} + A^{(2)}\mathbf{e}_j^{(2)})\,e^{i\mathbf{kx}}$. We can also write it in terms of the helicity basis (see Eq. (2.13))

$$\mathbf{e}^{(\pm)} = \frac{1}{\sqrt{2}}\left(\mathbf{e}^{(1)} \pm i\mathbf{e}^{(2)}\right)\ , \tag{4.60}$$

$$A_j = \left(A^{(+)}\mathbf{e}_j^{(+)} + A^{(-)}\mathbf{e}_j^{(-)}\right)e^{i\mathbf{kx}}\ ,$$

where $A^{(\pm)} = \frac{1}{\sqrt{2}}\left(A^{(1)} \mp iA^{(2)}\right)$.

We write the vector perturbations of the distribution function for a given Fourier mode $\mathbf{k}$ in this form

$$F^{(V)}(t,\mathbf{x},\mathbf{n},v)=\left[F^{(V+)}(t,\mathbf{k},\mathbf{n},v)\mathbf{e}^{(+)}\cdot\mathbf{n}+F^{(V-)}(t,\mathbf{k},\mathbf{n},v)\mathbf{e}^{(-)}\cdot\mathbf{n}\right]e^{i\mathbf{k}\cdot\mathbf{x}}\,. \tag{4.61}$$

The functions $F^{(V+)}$ and $F^{(V-)}$ no longer depend on $\mathbf{e}^{(\pm)}$. Therefore, if the process which generated the fluctuations is isotropic, the components $F^{(V\pm)}$ depend on the direction $\mathbf{n}$ only via $\mu=\hat{\mathbf{k}}\cdot\mathbf{n}$. With respect to spherical coordinates where $\mathbf{k}$ points in the z-direction, the components of $\mathbf{n}$ are $\mathbf{n}=(\sqrt{1-\mu^2}\cos\varphi,\ \sqrt{1-\mu^2}\sin\varphi,\ \mu)$. With $\mathbf{e}^{(1)}=(1,0,0)$ and $\mathbf{e}^{(2)}=(0,1,0)$ we obtain

$$\mathbf{e}^{(\pm)}\cdot\mathbf{n}=n^{\mp}=\sqrt{\frac{1-\mu^2}{2}}e^{\pm i\varphi}\,, \tag{4.62}$$

so that

$$F^{(V)}(t,\mathbf{x},\mathbf{n},v)=\sqrt{\frac{1-\mu^2}{2}}\left[F^{(V+)}(t,k,\mu,v)e^{i\varphi}+F^{(V-)}(t,k,\mu,v)e^{-i\varphi}\right]e^{i\mathbf{k}\cdot\mathbf{x}}\,. \tag{4.63}$$

We now write the vorticity vector perturbation of the energy–momentum tensor in the helicity basis, $\Omega_i(t,\mathbf{k})=\Omega^{(+)}(t,\mathbf{k})e_i^{(+)}+\Omega^{(-)}(t,\mathbf{k})e_i^{(-)}$.

$$\begin{aligned}\Omega^j(t,\mathbf{k})&=\frac{-1}{\bar\rho+\bar P}T_0^{(V)j}\,,\\ \Omega^{(\pm)}(t,\mathbf{k})&=\frac{1}{(\bar\rho+\bar P)a^4}\int\mathbf{e}^{\mp}\cdot\mathbf{n}v^3F^{(V)}(t,\mathbf{k},\mathbf{n},v)\,dv\,d\Omega_{\mathbf{n}}\\ &=\frac{\pi}{(\bar\rho+\bar P)a^4}\int v^3F^{(V\pm)}(t,k,\mu,v)(1-\mu^2)\,dv\,d\mu\,.\end{aligned} \tag{4.64}$$

In the chosen gauge, $B^i=0$, we obtain $T_0^i=(\bar\rho+\bar P)\Omega^i$, hence the first moment, $\int n^iF^{(V)}$ gives rise to the vorticity Ω^i and not to the shear $V^{(V)i}$ (for details see Section 2.2.4).

We introduce the helicity decomposition of the vector potential for anisotropic stresses,

$$\Pi_j^{(V)}=\left(\Pi^{(V+)}\mathbf{e}_j^{(+)}+\Pi^{(V-)}\mathbf{e}_j^{(-)}\right)e^{i\mathbf{kx}}\,. \tag{4.65}$$

The anisotropic stress tensor is defined by $\Pi^{(V)i}{}_j=\frac{-1}{2k}\left(\Pi^{(V)i}{}_{|j}+\Pi_j^{(V)|i}\right)$. But $\Pi^{(V)i}{}_j$ is also given by the integral of the distribution function over momentum space,

$$\Pi^{(V)i}{}_j=\frac{1}{a^4\bar P}\int\frac{v^4}{q}\left(n^in_j-\frac{1}{3}\delta^i_j\right)F^{(V)}\,dv\,d\Omega_{\mathbf{n}}\,. \tag{4.66}$$

Taking the divergence of both expressions we obtain

$$\begin{aligned}\Pi^{(V)i}{}_{j|i} &= \frac{k}{2}\left(\Pi^{(V+)}\mathbf{e}_j^{(+)} + \Pi^{(V-)}\mathbf{e}_j^{(-)}\right)e^{i\mathbf{kx}} \\ &= \frac{ik}{a^4\bar{P}}\int \frac{v^4}{q}\left(n_j\mu - \frac{1}{3}\hat{k}_j\right)F^{(V)}\,dv\,d\Omega_{\mathbf{n}}\,. \end{aligned} \tag{4.67}$$

We multiply this vector with $\mathbf{e}^{\mp}$ to isolate the modes $\Pi^{\pm}$. We also make use of the helicity decomposition of the distribution function, Eq. (4.61)

$$\begin{aligned}\Pi^{(V\pm)} &= \frac{2i}{a^4\bar{P}}\int \frac{v^4}{q} n^{\pm}\mu F^{(V)}\,dv\,d\Omega_{\mathbf{n}} \\ &= \frac{2\pi i}{a^4\bar{P}}\int \frac{v^4}{q}\mu(1-\mu^2)F^{(V\pm)}(t,k,\mu,v)\,dv\,d\mu\,. \end{aligned} \tag{4.68}$$

For the second equals sign we made use of the decomposition $\mathbf{n} = n^+\mathbf{e}^{(+)} + n^-\mathbf{e}^{(-)} + \mu\hat{\mathbf{k}} = \sqrt{1-\mu^2}e^{-i\varphi}\mathbf{e}^{(+)} + \sqrt{1-\mu^2}e^{i\varphi}\mathbf{e}^{(-)} + \mu\hat{\mathbf{k}}$ introduced above.

4.3.3 Tensor perturbations

For tensor perturbations only the anisotropic stresses survive. The ansatz for a tensor-type Fourier mode of the distribution function is

$$F^{(T)}(t,\mathbf{x},\mathbf{n},v) = \left[F^{(T\times)}(t,\mathbf{k},\mathbf{n},v)Q_{ij}^{(T\times)} + F^{(Td)}(t,\mathbf{k},\mathbf{n},v)Q_{ij}^{(Td)}\right]n^i n^j\,,$$

where

$$Q_{ij}^{(T\times)} = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{\sqrt{2}}\left[e_i^{(1)}e_j^{(2)} + e_i^{(2)}e_j^{(1)}\right] \quad \text{and} \quad Q_{ij}^{(Td)} = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{\sqrt{2}}\left[e_i^{(1)}e_j^{(1)} - e_i^{(2)}e_j^{(2)}\right]\,.$$

These tensors form a basis of the symmetric traceless tensors normal to $\mathbf{k}$. Usually, when discussing gravity waves, the second mode function is denoted $Q_{ij}^{(T+)}$. Here we use $Q_{ij}^{(Td)}$ in order not to confuse this basis with the helicity basis which we shall use later also for tensor perturbations. The superscript d indicates that this tensor is non-zero only on the diagonal with principal axes (eigenvectors) $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$, while $Q_{ij}^{(T\times)}$ is purely off-diagonal. Its principal axes are rotated by 45° with respect to $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$.

$$\Pi^{(T)i}{}_j = \frac{1}{a^4\bar{P}}\int \frac{v^4}{q}\left(n^i n_j - \frac{1}{3}\delta^i_j\right)F^{(T)}\,dv\,d\Omega_{\mathbf{n}}\,. \tag{4.69}$$

With the above decomposition of the distribution function and $\Pi_{ij}^{(T)} = \Pi^{(T\times)}(t,\mathbf{k})Q_{ij}^{(T\times)} + \Pi^{(Td)}(t,\mathbf{k})Q_{ij}^{(Td)}$ we obtain

$$\Pi^{(T\bullet)} = \frac{\pi}{2a^4\bar{P}}\int \frac{v^4}{q}F^{(T\bullet)}(t,k,v,\mu)(1-\mu^2)^2\,dv\,d\mu\ . \tag{4.70}$$

As for vector perturbations, we assume that the process generating the perturbation is isotropic, so that $F^{(T\bullet)}$ depends on the direction $\mathbf{n}$ only via $\mu = \mathbf{n}\cdot\hat{\mathbf{k}}$.

In the massless case, which is most relevant for us, the energy–momentum tensor simplifies considerably. This is the subject of the next section.

4.4 The ultra-relativistic limit, the Liouville equation for massless particles

The Liouville equation and the expression for the perturbations of the energy–momentum tensor derived in the previous section are actually more important for massive collisionless particles, e.g., massive neutrinos, than for massless particles. In the massless (or ultra-relativistic) case we have $q = v$ and the equations simplify significantly. Before discussing the different modes, let us introduce the 'longitudinal temperature fluctuation' for a thermal bath of massless particles. 'Longitudinal' indicates that we consider perturbations in the longitudinal gauge. We integrate the perturbed distribution function over energy so that only the dependence on momentum directions, $\mathbf{n}$, remains,

$$\frac{4\pi}{a^4}\int v^3 f\,dv \equiv \bar{\rho}\,(1+4\Delta(\mathbf{n}))\ . \tag{4.71}$$

We call the variable $\Delta(\mathbf{n})$ the longitudinal temperature fluctuation in direction $\mathbf{n}$. $\Delta(\mathbf{n})$ depends also on $(t,\mathbf{x})$ which we suppress here for brevity. This definition is motivated by the following consideration: for a Planck distribution of photons which has a slightly direction-dependent temperature, but is otherwise unperturbed (especially, it has a perfect blackbody spectrum, $f_B(p,T) = (\exp(p/T)+1)^{-1}$), the perturbed distribution function can be expanded to first order as

$$f(p,\mathbf{n}) = f_B(p,T(\mathbf{n})) = f_B(p,\bar{T}) - \frac{\delta T}{\bar{T}}p\partial_p f_B(p,\bar{T})\ . \tag{4.72}$$

Observe that f_B is purely a function of p/T so that $\partial_T f_B = -(p/T)\partial_p f_B$. The energy density of this photon distribution is given by

$$\begin{aligned}\rho_\gamma &= \frac{1}{a^4}\int v^3 f(v,\mathbf{n})\,dv\,d\Omega_{\mathbf{n}} = \bar{\rho}_\gamma - \frac{1}{a^4}\int \frac{\delta T}{\bar{T}}v^4\partial_v f_B(v,\bar{T})\,dv\,d\Omega_{\mathbf{n}}\\ &= \bar{\rho}_\gamma\left(1+\frac{4}{4\pi}\int\frac{\delta T}{\bar{T}}d\Omega_{\mathbf{n}}\right) = \bar{\rho}_\gamma\left(1+\frac{1}{\pi}\int\Delta(\mathbf{n})\,d\Omega_{\mathbf{n}}\right)\ .\end{aligned} \tag{4.73}$$

For the third equals sign we have performed an integration by parts to evaluate the integral over v. We shall see that the Liouville equation for photons leads to a perturbation which can be described entirely by such a direction-dependent temperature fluctuation. Of course f and also $\Delta = \delta T/\bar{T}$ also depend on position and time, arguments which we suppress here for brevity. The fact that the perturbation of the photon distribution can be described in such a simple way is not surprising. It is an expression of the 'a-chromaticy' of gravity which is a consequence of the equivalence principle: the deflection and redshift of a photon in a gravitational field are independent of its energy.

4.4.1 Scalar perturbations

For massless particles, $v = q$, the Liouville equation (4.34) reduces to

$$\partial_0 \mathcal{F} + n^i \partial_i \mathcal{F} = n^j \left[\Psi_{,j} + \Phi_{,j}\right] v \frac{d\bar{f}}{dv} \,. \tag{4.74}$$

We define

$$\mathcal{M}^{(S)}(t, \mathbf{x}, \mathbf{n}) = \frac{\pi}{a^4 \bar{\rho}} \int v^3 \mathcal{F}\, dv \,. \tag{4.75}$$

In terms of the temperature fluctuation $\Delta(\mathbf{n})$ defined in Eq. (4.71) we get

$$\mathcal{M}^{(S)}(\mathbf{n}) = \Delta^{(S)}(\mathbf{n}) - \Phi \,. \tag{4.76}$$

Up to a (irrelevant) monopole contribution, the momentum integrated distribution function $\mathcal{M}$ is nothing other than the temperature perturbation in the longitudinal gauge. It is not surprising that the monopole terms of $\mathcal{M}(\mathbf{n})$ and $\Delta(\mathbf{n})$ do not agree because they are gauge dependent. Also the dipole terms might differ since they too are gauge dependent. (In a gauge with non-vanishing shear, the dipole contributions to Δ and $\mathcal{M}$ do differ.)

Integrating the Liouville equation (4.74) over momenta and performing an integration by parts on the right-hand side, we obtain the evolution equation for $\mathcal{M}$.

For the scalar part of the distribution function we obtain

$$\partial_t \mathcal{M}^{(S)} + n^i \partial_i \mathcal{M}^{(S)} = -n^j \left[\Psi_{,j} + \Phi_{,j}\right] \,. \tag{4.77}$$

This equation can be solved formally for any given source term $\Phi + \Psi$. One easily checks that the solution with initial condition $\mathcal{M}^{(S)}(t_{\rm in}, \mathbf{x}, \mathbf{n})$ is

$$\begin{aligned} \mathcal{M}^{(S)}(t, \mathbf{x}, \mathbf{n}) = {} & \mathcal{M}^{(S)}\left(t_{\rm in}, \mathbf{x} - \mathbf{n}(t - t_{\rm in}), \mathbf{n}\right) \\ & - \int_{t_{\rm in}}^{t} dt'\, n^i \partial_i (\Psi + \Phi)(t', \mathbf{x} - \mathbf{n}(t - t')) \,. \end{aligned} \tag{4.78}$$

Using

$$\frac{d}{dt'}(\Psi+\Phi)(t',\mathbf{x}-\mathbf{n}(t-t')) = \partial_{t'}(\Psi+\Phi)(t',\mathbf{x}-\mathbf{n}(t-t')) + n^i\,\partial_i(\Psi+\Phi)(t',\mathbf{x}-\mathbf{n}(t-t'))\,,$$

we can replace the second term on the right-hand side to obtain

$$\begin{aligned}\mathcal{M}^{(S)}(t,\mathbf{x},\mathbf{n}) = {}& \mathcal{M}^{(S)}\,(t_{\rm in},\mathbf{x}-\mathbf{n}(t-t_{\rm in}),\mathbf{n}) \\ & + (\Psi+\Phi)(t_{\rm in},\mathbf{x}-\mathbf{n}(t-t_{\rm in})) \\ & + \int dt'\,\partial_{t'}(\Psi+\Phi)(t',\mathbf{x}-\mathbf{n}(t-t')) + \text{monopole}\,. \end{aligned} \tag{4.79}$$

By 'monopole' we denote an uninteresting $\mathbf{n}$-independent contribution which does not affect the CMB anisotropy spectrum. The Bardeen potentials Ψ and Φ, however, are given via Einstein's equation in terms of the perturbations of the energy–momentum tensor which contain contributions from the photons which are in turn the momentum integrals of $\mathcal{M}$ given below. Therefore, even though it might look like it, this is not really a solution of the Liouville equation. The term on the right-hand side also depends on $\mathcal{M}^{(S)}$.

Let us compare Eq. (4.79) with the result from the integration of light-like geodesics after decoupling in Eqs. (2.229) and (2.231). Here we have solved the Liouville equation which also does not take into account the scattering of photons and is therefore equivalent to our approach in Chapter 2. They both correspond to the 'sudden decoupling' approximation, where we assume that photons behave like a perfect fluid before decoupling and are entirely free after decoupling. This is a relatively good approximation for all scales which are much larger than the duration of the process of recombination which we shall estimate in the next section. The comparison with Eqs. (2.229) and (2.231) yields

$$\mathcal{M}^{(S)}\,(t_{\rm dec},\mathbf{x}-\mathbf{n}(t-t_{\rm dec}),\mathbf{n}) = \left(\frac{1}{4}D_g + \mathbf{n}\cdot\mathbf{V}^{(b)}\right)(t_{\rm dec},\mathbf{x}-\mathbf{n}(t-t_{\rm dec}))\,, \tag{4.80}$$

and

$$\mathcal{M}^{(S)}(t,\mathbf{x},\mathbf{n}) \equiv \frac{\delta T}{T}(t,\mathbf{x},\mathbf{n})\,. \tag{4.81}$$

In other words, the temperature fluctuation defined via the energy shift of photons moving along geodesics corresponds to $\mathcal{M}^{(S)}$ while the temperature fluctuation defined via the energy density fluctuation in longitudinal gauge corresponds to $\Delta^{(S)} = \mathcal{M}^{(S)} + \Phi$. In addition to the energy shift, the latter includes a contribution from the perturbation of the volume element, $\sqrt{|\det(g_{ij})|}\,d^3x = a^3(1-3\Phi)\,d^3x$. The distinction is not very important since it is a monopole which does not show

up in the angular power spectrum. However, the corresponding evolution equations are of course different. The variable $\Delta^{(S)}$ is used, for example, in Hu & Sugiayma (1995), Hu *et al.* (1995, 1998) and Seljak & Zaldarriaga (1996), while the variable $\mathcal{M}^{(S)}$ is used, for example, in Durrer & Straumann (1988), Durrer (1990, 1994), Durrer *et al.* (2002), Doran (2005) and Bashinsky (2006).

The initial condition in the sudden decoupling approximation is a distribution function which contains only a monopole and a dipole. Higher multipoles do not build up in a perfect fluid. In the next section we shall take into account the process of decoupling by studying the Boltzmann equation.

The scalar perturbations of the energy–momentum tensor of the radiation fluid for a given Fourier mode $\mathbf{k}$ can be found by integrating the r.h.s. of Eqs. (4.52), (4.54), (4.57) and (4.59) over energy,

$$D_g = 2\int_{-1}^{1} \mathcal{M}^{(S)}(\mu)\, d\mu\,, \tag{4.82}$$

$$V = \frac{3i}{2}\int_{-1}^{1} \mu\mathcal{M}^{(S)}(\mu)\, d\mu\,, \tag{4.83}$$

$$\Gamma = 0\,, \tag{4.84}$$

$$\Pi = 3\int_{-1}^{1} \left(1-3\mu^2\right)\mathcal{M}^{(S)}(\mu)\, d\mu\,. \tag{4.85}$$

We have assumed that $\mathcal{M}^{(S)}(\mathbf{n})$ depends on the direction $\mathbf{n}$ only via $\mu = \hat{\mathbf{k}}\cdot\mathbf{n}$ and have performed the integration over φ which simply gives a factor 2π. For isotropic perturbations there is no other vector which could single out a direction and therefore this assumption reflects statistical isotropy.

The exact equality $w = c_s^2 = \frac{1}{3}$ does not allow for any entropy perturbation in a pure radiation fluid.

4.4.2 Vector perturbations

Vector perturbations of the distribution function are not gauge dependent. We have directly $\mathcal{M}^{(V)} \equiv \Delta^{(V)}$. The Liouville equation for vector perturbations of the radiation fluid is obtained by integrating Eq. (4.39) over energies,

$$\mathcal{M}^{(V)}(\mathbf{n}) = \frac{\pi}{a^4\bar{\rho}}\int v^3 F^{(V)}(\mathbf{n}, v)\, dv\,,$$

$$\mathcal{M}^{(V\pm)}(\mu) = \frac{\pi}{a^4\bar{\rho}}\int v^3 F^{(V\pm)}(\mu, v)\, dv\,, \tag{4.86}$$

$$\partial_t\mathcal{M}^{(V)} + n^i\partial_i\mathcal{M}^{(V)} = -n^i n^j \dot{H}_{ij}^{(V)}\,. \tag{4.87}$$

The formal solution to this equation is

$$\mathcal{M}^{(V)}(t,\mathbf{x},\mathbf{n}) = \mathcal{M}^{(V)}(t_{\rm in},\mathbf{x}-\mathbf{n}(t-t_{\rm in}),\mathbf{n}) - \int_{t_{\rm in}}^{t} dt'\, n^i n^j a(t')^{-1}\sigma_{ij}^{(V\pm)}(t',\mathbf{x}-\mathbf{n}(t-t'))\,. \tag{4.88}$$

After Fourier transforming $\mathcal{M}^{(V)}$ and $\sigma_{ij}^{(V)}$ we can expand them in the helicity basis,

$$\dot{H}_j^{(V)} = \dot{H}^{(V+)}e_j^{(+)} + \dot{H}^{(V-)}e_j^{(-)} = \sigma^{(V+)}e_j^{(+)} + \sigma^{(V-)}e_j^{(-)},$$

so that

$$\begin{aligned}\dot{H}_{ij}^{(V)} &\equiv \frac{-1}{2k}\left(\dot{H}_{i|j}^{(V)} + \dot{H}_{j|i}^{(V)}\right)\\ &= \frac{-i}{2}\left(\dot{H}^{(V+)}\left[e_i^{(+)}\hat{k}_j + e_j^{(+)}\hat{k}_i\right] + \dot{H}^{(V-)}\left[e_i^{(-)}\hat{k}_j + e_j^{(-)}\hat{k}_i\right]\right)\\ n^i n^j \dot{H}_{ij}^{(V)} &= \frac{-i}{\sqrt{2}}\mu\sqrt{1-\mu^2}\left(\dot{H}^{(V+)}e^{i\varphi} + \dot{H}^{(V-)}e^{-i\varphi}\right)\,.\end{aligned} \tag{4.89}$$

In the last equality we have introduced the representation of $\mathbf{n}$ in the helicity basis, Eq. (4.62). The φ dependence on the left- and right-hand sides of Eq. (4.87) shows that $\mathcal{M}^{(V\pm)}$ couples only to $\sigma^{(V\pm)}$ and both helicity components satisfy the equation

$$\partial_t \mathcal{M}^{(V\pm)} + ik\mu\mathcal{M}^{(V\pm)} = -i\mu\sigma^{(V\pm)}\,. \tag{4.90}$$

From Eqs. (4.64) and (4.68) we obtain the vector perturbations of the energy–momentum tensor in terms of $\mathcal{M}^{(V)}$,

$$\Omega^{(\pm)} = \frac{3}{4}\int_{-1}^{1}(1-\mu^2)\mathcal{M}^{(V\pm)}(\mu)\,d\mu\,, \tag{4.91}$$

$$\Pi^{(V\pm)} = 6i\int_{-1}^{1}(1-\mu^2)\mu\mathcal{M}^{(V\pm)}(\mu)\,d\mu\,. \tag{4.92}$$

4.4.3 Tensor perturbations

For tensor fluctuations, the perturbed Liouville equation becomes

$$\partial_t \mathcal{M}^{(T)} + n^i\partial_i\mathcal{M}^{(T)} = -n^i n^j \dot{H}_{ij}^{(T)}\,. \tag{4.93}$$

For a given source term $H_{ij}^{(T)}$ this is solved by

$$\mathcal{M}^{(T)}(t,\mathbf{x},\mathbf{n}) = \mathcal{M}^{(T)}(t_{\rm in},\mathbf{x}-\mathbf{n}(t-t_{\rm in}),\mathbf{n}) - \int_{t_{\rm in}}^{t} dt'\, n^i n^j \dot{H}_{ij}^{(T)}(t',\mathbf{x}-\mathbf{n}(t-t'))\,. \tag{4.94}$$

We decompose also $H^{(T)}$ and $\mathcal{M}^{(T)}$ in the basis $Q_{ij}^{(T\bullet)}$ defined in Eq. (2.12),

$$H_{ij}^{(T)} = H^{(T\times)} Q_{ij}^{(T\times)} + H^{(Td)} Q_{ij}^{(Td)} , \tag{4.95}$$

$$\mathcal{M}^{(T\bullet)}(\mu) = \frac{\pi}{a^4\bar{\rho}} \int dv\, v^3 F^{(T\bullet)}(\mu, v) , \tag{4.96}$$

$$\partial_t \mathcal{M}^{(T\bullet)} + ik\mu \mathcal{M}^{(T\bullet)} = -\dot{H}^{(T\bullet)} . \tag{4.97}$$

The tensor anisotropic stresses are

$$\Pi^{(T\bullet)} = \frac{3}{2} \int_{-1}^{1} (1-\mu^2)^2 \mathcal{M}^{(T\bullet)}(\mathbf{n}) d\Omega . \tag{4.98}$$

4.4.4 The Liouville equation in terms of the Weyl tensor

We know that the motion of photons in a gravitational field is conformally invariant. Therefore, the evolution of the photon distribution, once the redshift is taken out by using the conformally invariant momentum variable v, should depend only on the Weyl tensor. To find the Liouville equation in terms of the Weyl tensor, we first consider only scalar and tensor perturbations and assume that the vector perturbations vanish. Adding together the scalar and tensor parts of the Liouville equation $\mathcal{M} = \mathcal{M}^{(S)} + \mathcal{M}^{(T)}$ the Liouville equation becomes

$$(\partial_t + n^j \partial_j)\mathcal{M} = n^i \partial_i(\Psi + \Phi) - n^j n^i \dot{H}_{ij}^{(T)} \equiv S_G . \tag{4.99}$$

Now we apply the Laplacian on both sides. Using the expressions (A3.21) and (A3.57)–(A3.59) for the Weyl tensor, one finds that this corresponds to

$$(\partial_t + n^i \partial_i)\Delta\mathcal{M} = -3n^i\, \partial^j E_{ij} - n^k n^j \epsilon_k{}^{i\ell}\, \partial_\ell B_{ij} \equiv \Delta S_G , \tag{4.100}$$

where $\epsilon_{ki\ell}$ is the totally anti-symmetric tensor in three dimensions, and E_{ij} and B_{ij} are the electric and magnetic parts of the Weyl tensor. We shall sometimes use S_G to denote the gravitational source term on the right-hand side of the Liouville equation. As before, this equation is easily solved for a given source term, S_G. In terms of our variable $\mathcal{M}$ the Liouville equation is easily written in terms of the Weyl tensor, while this is not possible with the variable Δ. This variable $\mathcal{M}$ manifests the conformal invariance of photon propagation. It remains zero if the Weyl curvature vanishes and therefore photons are not deflected.

If we want to include also vector perturbations a subtlety occurs. With the help of (A3.43)–(A3.45) one finds

$$-3n^i \partial^j E_{ij}^{(V)} - n^k n^j \epsilon_k{}^{i\ell} \partial_\ell B_{ij}^{(V)} = \frac{3}{4} n^j \Delta\dot{\sigma}_j + \frac{1}{4} n^i n^j \Delta\sigma_{i|j} , \tag{4.101}$$

which does not correspond to the r.h.s. of Eq. (4.87). However, if we transform $\mathcal{M}^{(V)}$ by the addition of a simple dipole term which does not show up in the CMB multipoles for $\ell \geq 2$ to

$$\mathcal{M}^{(V2)} \equiv \mathcal{M}^{(V)} + \frac{3}{4} n^j \dot{H}_j \,, \tag{4.102}$$

one finds easily that

$$(\partial_t + n^i \partial_i) \Delta\mathcal{M}^{(V2)} = -3 n^i \partial^j E^{(V)}_{ij} - n^k n^j \epsilon_k{}^{i\ell} \partial_\ell B^{(V)}_{ij} \,. \tag{4.103}$$

Hence with this redefinition, the variable

$$\mathcal{M} \equiv \mathcal{M}^{(S)} + \mathcal{M}^{(V2)} + \mathcal{M}^{(T)} \,,$$

satisfied the Liouville equation

$$(\partial_t + n^i \partial_i) \Delta\mathcal{M} = -3 n^i \, \partial^j E_{ij} - n^k n^j \epsilon_k{}^{i\ell} \, \partial_\ell B_{ij} \,. \tag{4.104}$$

It may be interesting to note that in a generic gauge this variable can be written as

$$\Delta\mathcal{M} = \Delta M + \Delta\mathcal{R} + \frac{3}{2} n^i \, \partial^j \sigma_{ij} \,. \tag{4.105}$$

The first term, M is the simply the momentum integration of the perturbation of the distribution function, F, while the second term is the perturbation of the spatial curvature given in Eq. (2.48). Only scalar perturbations contribute to it. The last term is related to the shear and both, scalar and vector perturbations contribute to it. Note that for scalar perturbation in longitudinal gauge the shear term vanishes while the vector part of the shear is gauge invariant. By construction, this variable is perfectly gauge invariant.

The right-hand side of Eq. (4.104) is written entirely in terms of tensor fields with vanishing background contribution. It would be interesting to attempt the same for the left-hand side, the variable $\Delta\mathcal{M}$.

4.4.5 The Liouville equation in Fourier space

A Fourier mode of $\mathcal{M}(t, \mathbf{x}, \mathbf{n})$ is given by

$$\mathcal{M}(t, \mathbf{k}, \mathbf{n}) \equiv \int d^3x \, e^{-i\mathbf{k}\cdot\mathbf{x}} \mathcal{M}(t, \mathbf{x}, \mathbf{n}) \,, \text{ and its inverse is}$$

$$\mathcal{M}(t, \mathbf{x}, \mathbf{n}) = \frac{1}{(2\pi)^3} \int d^3k \, e^{i\mathbf{k}\cdot\mathbf{x}} \mathcal{M}(t, \mathbf{k}, \mathbf{n}) \,.$$

We have seen that the Liouville equation for a Fourier mode is given by

$$(\partial_t + ik\mu)\mathcal{M}(t, \mathbf{k}, \mathbf{n}) = S_G(t, \mathbf{k}, \mu) \,, \tag{4.106}$$

where $\mu = \hat{\mathbf{k}} \cdot \mathbf{n}$ is the cosine between the unit vectors $\hat{\mathbf{k}} = \mathbf{k}/k$ and $\mathbf{n}$. The general solution to this equation for a given source term S_G can be written as

$$\mathcal{M}(t, \mathbf{k}, \mathbf{n}) = e^{-ik\mu(t-t_{\rm in})}\mathcal{M}(t_{\rm in}, \mathbf{k}, \mathbf{n}) + \int_{t_{\rm in}}^{t} dt' e^{-ik\mu(t-t')} S_G(t', \mathbf{k}, \mathbf{n}) \,. \quad (4.107)$$

The function S_G can be decomposed into scalar, vector and tensor perturbations.

As already mentioned, the source term usually depends on $\mathcal{M}$ via Einstein's equations and Eq. (4.107) is not really a solution but simply corresponds to rewriting Eq. (4.106) as an integral equation. But as we shall see, this has serious advantages especially for numerical computations.

From Eq. (4.107) using the decomposition (see Appendix A4.3)

$$e^{i\mathbf{k}\cdot\mathbf{n}(t-t_{\rm in})} = \sum_{\ell=0}^{\infty} (2\ell+1) i^{\ell} j_{\ell}(k(t-t_{\rm in})) P_{\ell}(\mu) \,,$$

one finds the CMB power spectrum, exactly as in Chapter 2, Eqs. (2.249)–(2.253) and (2.259). Before we do this, we go on to study Thomson scattering which is the relevant scattering process just before recombination. We will then derive the power spectrum taking into account this scattering process.

4.5 The Boltzmann equation

At very early times, long before recombination, scattering of photons with free electrons is very frequent. During recombination, however, the number density of free electrons, i.e., of electrons not bound to an atom, drops drastically and soon the mean free path of photons is much larger than the Hubble scale so that, effectively, photons do not scatter any more. In the previous treatment we assumed this process of decoupling to be instantaneous; now we want to reconsider it in more detail.

The only scattering process which is relevant briefly before decoupling, i.e., at temperatures of a few electron volts and less, is elastic Thomson scattering, where the photon energy is conserved and only its direction is modified. The Thomson scattering rate is

$$\Gamma_T = \sigma_T n_e \,,$$

where $\sigma_T = 6.6524 \times 10^{-25}\,\mathrm{cm}^2$ is the Thomson scattering cross section and n_e is the number density of free electrons.

Before decoupling, in a matter dominated universe, we find

$$\begin{aligned} \Gamma_T &\simeq 7 \times 10^{-30}\,\mathrm{cm}^{-1}\Omega_b h^2 (1+z)^3 \quad \text{while} \\ H &\simeq 10^{-28}\,\mathrm{cm}^{-1} h (1+z)^{3/2} \\ \Gamma_T/H &\simeq 0.07\Omega_b h (z+1)^{3/2} \,. \end{aligned}$$

Hence before recombination, which corresponds to redshifts $z > 1100$, say, Thomson scattering is much faster than expansion. During recombination, the free electron density drops and eventually the Thomson scattering rate drops below the expansion rate. To take scattering into account we add a so-called 'collision integral' to the right-hand side of the Liouville equation, which leads us to the Boltzmann equation. To learn more about the Boltzmann equation and the approximations going into it see, e.g., Lifshitz & Pitajewski (1983). The collision integral $C[f]$ takes into account that the 1-particle distribution function can change due to collisions which scatter a particle into, f_+, or out of, f_-, a volume element $d^3x\, d^3p$ in phase space,

$$\left[\tilde{p}^\mu \partial_\mu - \Gamma^i_{\alpha\beta}\tilde{p}^\alpha\tilde{p}^\beta \frac{\partial}{\partial \tilde{p}^i}\right] f = C[f] = \frac{df_+}{dt} - \frac{df_-}{dt} \,. \tag{4.108}$$

Here f_+ and f_- denote the distribution of photons scattered into and out of the beam of photons at position $\mathbf{x}$ at time t with momentum $\mathbf{p}$ respectively.

In the baryon rest frame, which we denote by a prime, the photons scattered into the beam in direction $\mathbf{n}$ per unit of time are

$$\frac{df'_+}{dt'}(\mathbf{n}) = \frac{\sigma_T n_e}{4\pi}\int f'(p', \mathbf{n}')\omega(\mathbf{n}, \mathbf{n}')\, d\Omega' \,,$$

where $\omega(\mathbf{n}, \mathbf{n}')$ denotes the normalized angular dependence of Thomson scattering after averaging over photon polarizations (Jackson, 1975)

$$\begin{aligned}\frac{d\sigma}{d\Omega} &= \frac{\sigma_T}{4\pi}\omega(\mathbf{n}, \mathbf{n}') = \frac{3\sigma_T}{16\pi}\left[1 + (\mathbf{n}\cdot\mathbf{n}')^2\right] \,, \\ &= \frac{\sigma_T}{4\pi}\left[1 + \frac{3}{4}n_{ij}n'_{ij}\right] \quad \text{with} \quad n_{ij} = n_i n_j - \frac{1}{3}\delta_{ij} \,.\end{aligned} \tag{4.109}$$

Here we have averaged over incoming polarizations and summed over final polarizations of the photons, see Jackson (1975). In this chapter we neglect the polarization dependence of Thomson scattering which we discuss fully in Chapter 5.

In the baryon rest frame which moves with 4-velocity u^μ, the photon energy is

$$p' = -\tilde{p}_\mu u^\mu = p(1 - n_i v^i) \,.$$

Since Thomson scattering is energy independent, we may integrate f_+ over photon energies $p' = v'/a$ to obtain again an equation for $\mathcal{M}$. With $v'^3\, dv' = (1 - 4n_i v^i)v^3\, dv$ and Eq. (4.71), we obtain

$$\frac{4\pi}{a^4}\int v'^3 \frac{df'_+}{dt'}\, dv' = \bar{\rho}_r \sigma_T n_e \left[1 - 4n_i v^i + \frac{1}{\pi}\int \Delta(\mathbf{n}')\omega(\mathbf{n}, \mathbf{n}')\, d\Omega'\right] \,. \tag{4.110}$$

Here $\bar{\rho}_r$ is the background radiation density. The term $4\mathbf{n}\cdot\mathbf{v}$ is a Doppler term from the velocity of the electrons with respect to the longitudinal rest frame. The factor of 4 comes from the fact that we have to transform $p'^3\,dp' = p^3(1+4\mathbf{n}\cdot\mathbf{v})\,dp$ from the electron rest frame into the 'laboratory' frame.

The distribution of photons scattered out of the beam per unit time is simply the scattering rate multiplied by the distribution function,

$$\frac{df_-}{dt'} = \sigma_T n_e f'(p', \mathbf{n})\,.$$

Also integrating this term over photon energies we obtain the collision term which enters the energy integrated Boltzmann equation for $\mathcal{M}$ in the baryon rest frame,

$$\begin{aligned} C'[\mathcal{M}] &= \frac{\pi}{a^4\bar{\rho}_r}\int v^3\,dv\left(\frac{df_+}{dt'} - \frac{df_-}{dt'}\right) \\ &= \sigma_T n_e\left[\frac{1}{4}\delta_r^{(\mathrm{long})} - \Delta(\mathbf{n}) - n_i V^{(b)i} + \frac{3n_{ij}}{16\pi}\int \Delta(\mathbf{n}')n'_{ij}\,d\Omega'\right]. \end{aligned} \quad (4.111)$$

Here $\delta_r^{(\mathrm{long})}$ is the density perturbation in longitudinal gauge. (We work, as usual, in longitudinal gauge.) In order to replace Δ in the collision term with $\mathcal{M}$ we use the relation, Eq. (4.76) and $\delta_r^{(\mathrm{long})} = D_g^{(r)} + 4\Phi$ together with the fact that Δ and $\mathcal{M}$ differ only by a monopole term with does not contribute to the angular integral in Eq. (4.111). We introduce also

$$M_{ij} = \frac{3}{8\pi}\int n_{ij}\mathcal{M}(\mathbf{n})\,d\Omega$$

and observe that to lowest order $C = (dt'/dt)C' = aC'$. With all this the Boltzmann equation becomes

$$\left(\partial_t + n^i\partial_i\right)\mathcal{M} = S_G(\mathbf{n}) + a\sigma_T n_e\left[\frac{1}{4}D_g^{(r)} - \mathcal{M} - n^i V_i^{(b)} + \frac{1}{2}n_{ij}M^{ij}\right], \quad (4.112)$$

where S_G is the gravitational term defined in Eq. (4.100). Note that the perturbation of the electron density, $n_e = \bar{n}_e + \delta n_e$ does not contribute to first order, since the isotropic background photon distribution $\bar{f}$ annihilates the collision term.

For the Fourier transform of $\mathcal{M}$ we obtain the equation

$$\begin{aligned} &(\partial_t + ik\mu + a\sigma_T n_e)\,\mathcal{M}(\mathbf{k},\mathbf{n}) = S_G(\mathbf{k},\mathbf{n}) \\ &\quad + a\sigma_T n_e\left[\frac{1}{4}D_g^{(r)}(\mathbf{k}) - n^i V_i^{(b)}(\mathbf{k}) + \frac{1}{2}n_{ij}M^{ij}(\mathbf{k},\mathbf{n})\right]. \end{aligned} \quad (4.113)$$

This can be converted to the integral equation

$$\mathcal{M}(t,\mathbf{k},\mathbf{n}) = e^{-ik\mu(t-t_{\rm in})-\kappa(t_{\rm in},t)}\mathcal{M}(t_{\rm in},\mathbf{k},\mathbf{n}) + \int_{t_{\rm in}}^{t} dt'\, e^{ik\mu(t'-t)-\kappa(t',t)}\left[S_G(\mathbf{k},\mathbf{n}) + \dot\kappa\left(\frac{1}{4}D_g^{(r)}(\mathbf{k}) - n^i V_i^{(b)}(\mathbf{k}) + \frac{1}{2}n_{ij}M^{ij}(\mathbf{k},\mathbf{n})\right)\right] . \tag{4.114}$$

Here $\kappa(t_1,t_2) = \int_{t_1}^{t_2} a\sigma_T n_e\, dt$ is the optical depth and $\dot\kappa(t_1,t_2) = \partial_{t_2}\kappa(t_1,t_2) = a\sigma_T n_e(t_2)$ is independent of the initial value t_1.

We now decompose Eq. (4.114) into its scalar, vector and tensor contributions.

4.5.1 Scalar perturbation

We first consider scalar perturbations. Since the direction dependence enters the evolution equation only via the cosine $\mu = \hat{\mathbf{k}}\cdot\mathbf{n}$, we assume consistently that this is the only direction dependence of the Fourier transform $\mathcal{M}^{(S)}(t,\mathbf{k},\mathbf{n})$, so that $\mathcal{M}(t,\mathbf{k},\mathbf{n}) = \mathcal{M}(t,\mathbf{k},\mu)$. It therefore makes sense to expand $\mathcal{M}$ in Legendre polynomials,

$$\mathcal{M}^{(S)}(t,\mathbf{k},\mu) = \sum (2\ell+1)(-i)^\ell \mathcal{M}_\ell^{(S)}(t,\mathbf{k})P_\ell(\mu) . \tag{4.115}$$

Using the orthogonality and normalization of Legendre polynomials, see Appendix A4.1, we obtain the expansion coefficients,

$$\mathcal{M}_\ell^{(S)}(t,\mathbf{k}) = \frac{i^\ell}{2}\int_{-1}^{1} d\mu\, \mathcal{M}^{(S)}(t,\mathbf{k},\mu)P_\ell(\mu) . \tag{4.116}$$

Statistical homogeneity and isotropy imply that the coefficients $\mathcal{M}_\ell$ for different values of ℓ and $\mathbf{k}$ are uncorrelated,

$$\left\langle \mathcal{M}_\ell^{(S)}(t,\mathbf{k})\mathcal{M}_{\ell'}^{(S)*}(t,\mathbf{k}')\right\rangle = M_\ell^{(S)}(t,k)(2\pi)^3\,\delta^3(\mathbf{k}-\mathbf{k}')\,\delta_{\ell\ell'} . \tag{4.117}$$

We want to relate the spectrum $M_\ell^{(S)}(t,k)$ to the scalar CMB power spectrum $C_\ell^{(S)}$. We use the definition given in Eq. (2.241),

$$\left\langle \frac{\Delta T}{T}(t_0,\mathbf{x}_0,\mathbf{n})\frac{\Delta T}{T}(t_0,\mathbf{x}_0,\mathbf{n}')\right\rangle^{(S)} = \frac{1}{4\pi}\sum_\ell (2\ell+1)C_\ell^{(S)}P_\ell(\mathbf{n}\cdot\mathbf{n}')$$
$$= \frac{1}{(2\pi)^6}\int d^3k\, d^3k' \sum_{\ell_1\ell_2}(2\ell_1+1)(2\ell_2+1)(-i)^{\ell_1-\ell_2}e^{i\mathbf{x}_0\cdot(\mathbf{k}-\mathbf{k}')} \times \left\langle \mathcal{M}_{\ell_1}^{(S)}(t_0,\mathbf{k})\mathcal{M}_{\ell_2}^{(S)*}(t_0,\mathbf{k}')\right\rangle P_{\ell_1}(\mu)P_{\ell_2}(\mu') ,$$

where $\mu = \hat{\mathbf{k}} \cdot \mathbf{n}$ and $\mu' = \hat{\mathbf{k}}' \cdot \mathbf{n}'$. With Eq. (4.117) we obtain

$$\begin{aligned}
&\frac{1}{4\pi} \sum_{\ell} (2\ell+1) C_\ell^{(S)} P_\ell(\mathbf{n}\cdot\mathbf{n}') \\
&\quad = \frac{1}{(2\pi)^3} \sum_{\ell_1} \int d^3k \, M_{\ell_1}^{(S)}(t_0,k)(2\ell_1+1)^2 P_{\ell_1}(\mu) P_{\ell_1}(\mu') \\
&\quad = \frac{2}{\pi} \sum_{\ell_1} \int d^3k \, M_{\ell_1}^{(S)}(t_0,k) \sum_{m_1 m_2} Y_{\ell_1 m_1}(\mathbf{n}) Y^*_{\ell_1 m_1}(\hat{\mathbf{k}}) Y^*_{\ell_1 m_2}(\mathbf{n}') Y_{\ell_1 m_2}(\hat{\mathbf{k}}) \\
&\quad = \frac{2}{\pi} \sum_{\ell_1 m_1} \int dk \, k^2 M_{\ell_1}^{(S)}(t_0,k) Y_{\ell_1 m_1}(\mathbf{n}) Y^*_{\ell_1 m_1}(\mathbf{n}') \\
&\quad = \frac{1}{2\pi^2} \sum_{\ell_1} \left(\int dk \, k^2 M_{\ell_1}^{(S)}(t_0,k) \right) (2\ell_1+1) P_\ell(\mathbf{n}\cdot\mathbf{n}') \, .
\end{aligned}$$

In several steps in this derivation we have applied the addition theorem of spherical harmonics derived in Appendix A4.2.3. Comparing the first and the last expressions in the series of equalities above, we infer

$$C_\ell^{(S)} = \frac{2}{\pi} \int dk \, k^2 M_\ell^{(S)}(k) \, . \tag{4.118}$$

To calculate the CMB power spectrum, we therefore have to determine the random variables $\mathcal{M}_\ell$. We now derive a hierarchical set of equations for them, the so-called Boltzmann hierarchy.

With Eqs. (4.82)–(4.85), Eq. (4.116) and the explicit expressions of the Legendre polynomials for $\ell \leq 2$ given in Appendix A4.1, one finds the relations of the scalar perturbations of the photon energy–momentum tensor to the expansion coefficients $\mathcal{M}_\ell(t, \mathbf{k})$, $\ell \leq 2$,

$$D_g^{(r)} = 4\mathcal{M}_0^{(S)} \, , \tag{4.119}$$

$$V_r^{(S)} = 3\mathcal{M}_1^{(S)} \, , \tag{4.120}$$

$$\Pi_r^{(S)} = 12\mathcal{M}_2^{(S)} . \tag{4.121}$$

Inserting Eq. (4.115) in the definition of M_{ij} and choosing the coordinate system such that $\mathbf{k}$ points in the z direction one can easily compute the integrals $M_{33} = -\mathcal{M}_2^{(S)}$ and $M_{11} = M_{22} = \mathcal{M}_2^{(S)}/2$ and all off diagonal contributions vanish. With $n_1^2 + n_2^2 = 1 - \mu^2$ this yields

$$\frac{1}{2} n_{ij} M^{ij} = -\frac{1}{2} \mathcal{M}_2^{(S)} P_2(\mu) \, .$$

Also using the fact that for scalar perturbations $\mathbf{V} = i\hat{\mathbf{k}}V$ we obtain the scalar Boltzmann equation

$$(\partial_t + ik\mu)\,\mathcal{M}^{(S)}(\mathbf{k},\mathbf{n}) = ik\mu(\Phi+\Psi) + \dot{\kappa}\left[\frac{1}{4}D_g^{(r)}(\mathbf{k}) - \mathcal{M}^{(S)} - i\mu V^{(b)}(\mathbf{k}) - \frac{1}{2}\mathcal{M}_2(\mathbf{k})P_2(\mu)\right]. \tag{4.122}$$

With the recurrence relation (see Appendix A4.1)

$$\mu P_\ell(\mu) = \frac{\ell+1}{2\ell+1}P_{\ell+1}(\mu) + \frac{\ell}{2\ell+1}P_{\ell-1}(\mu)\,,$$

we can convert Eq. (4.122) into the following hierarchy of equations

$$\dot{\mathcal{M}}_\ell^{(S)} + k\frac{\ell+1}{2\ell+1}\mathcal{M}_{\ell+1}^{(S)} - k\frac{\ell}{2\ell+1}\mathcal{M}_{\ell-1}^{(S)} + \dot{\kappa}\mathcal{M}_\ell^{(S)} = \delta_{\ell0}\dot{\kappa}\mathcal{M}_0^{(S)} + \frac{1}{3}\delta_{\ell1}\left[-k(\Phi+\Psi) + \dot{\kappa}V^{(b)}\right] + \dot{\kappa}\frac{1}{10}\delta_{\ell2}\mathcal{M}_2^{(S)}. \tag{4.123}$$

Here the source terms on the right-hand side contribute only for $\ell = 0$, 1 and $\ell = 2$ respectively. In Eq. (4.123) each variable $\mathcal{M}_\ell^{(S)}$ couples to its neighbours, $\mathcal{M}_{\ell-1}^{(S)}$ and $\mathcal{M}_{\ell+1}^{(S)}$ via the left-hand side. From Eq. (4.122) it is clear, that the left-hand side actually just describes the free streaming of photons after decoupling.

If we want to determine the CMB power spectrum via the Boltzmann hierarchy, Eq. (4.123), in order to calculate, e.g., C_{1000} we have to know all the other $\mathcal{M}_\ell^{(S)}$s which may influence $\mathcal{M}_{1000}^{(S)}$ via free streaming during a Hubble time, which is certainly more than 1000. Furthermore, at the beginning, when coupling is still relatively tight, we may simply take into account $\mathcal{M}_0^{(S)}$ and $\mathcal{M}_1^{(S)}$ given by the perfect fluid initial conditions and set all the other $\mathcal{M}_\ell^{(S)}$s to zero. They then gradually build up mainly due to free streaming. But using the Boltzmann hierarchy (4.123), we cannot calculate $\mathcal{M}_{1000}^{(S)}$ with any accuracy if we have not determined all the $\mathcal{M}_\ell^{(S)}$s with $\ell < 1000$ with the same (or rather better) accuracy.

On the other hand, if we knew the source term, the right-hand side of Eq. (4.123), we could simply write down the solution, Eq. (4.114). As the source term only depends on the first two moments of the hierarchy, it can usually be obtained with a precision of about 0.1% (see Seljak & Zaldarriaga, 1996) by solving the hierarchy only up to $\ell \simeq 10$. Inserting the corresponding moments into Eq. (4.114) one finds

$$\mathcal{M}^{(S)}(t_0,\mathbf{k},\mu) = e^{-ik\mu(t_0-t_{\rm in})-\kappa(t_{\rm in},t_0)}\mathcal{M}^{(S)}(t_{\rm in},\mathbf{k},\mu) + \int_{t_{\rm in}}^{t_0} dt\, e^{ik\mu(t-t_0)-\kappa(t,t_0)} \times \left[ik\mu(\Phi+\Psi)(\mathbf{k}) + \dot{\kappa}\left(\frac{1}{4}D_g^{(r)}(\mathbf{k}) - i\mu V^{(b)}(\mathbf{k}) - \frac{1}{2}P_2(\mu)\mathcal{M}_2^{(S)}(\mathbf{k},t)\right)\right].$$

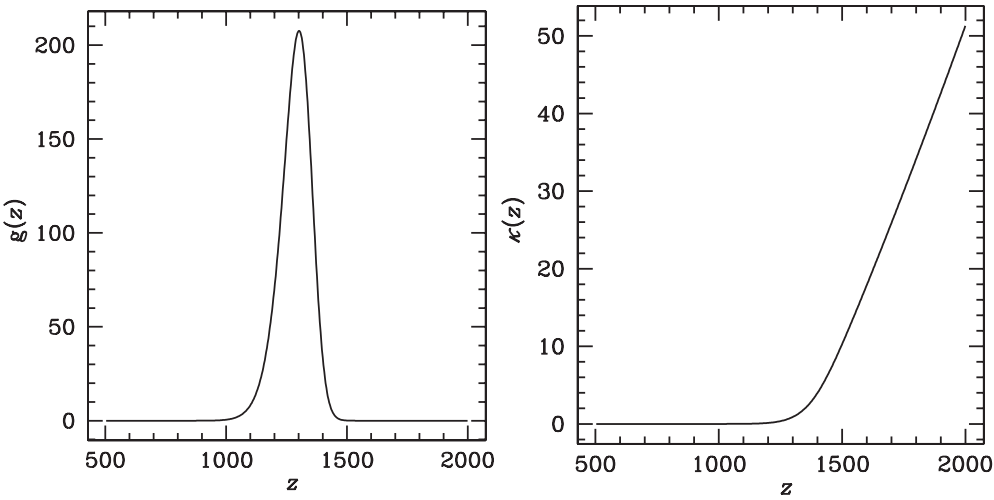


Fig. 4.1. The visibility function g (left) is plotted in units of H_0 as a function of redshift. For comparison we show also $\kappa(z)$ (right).

If the only μ-dependent term was the exponential, we could use its representation in terms of spherical Bessel functions, using Eq. (A4.101), to isolate $\mathcal{M}_\ell^{(S)}$. With this in mind, we use

$$e^{ik\mu(t-t_0)}\mu f(t) = -ik^{-1}\partial_t\left(e^{ik\mu(t-t_0)}\right) f(t)\,,$$

to get rid of all the μ-dependence in the term in square brackets above. Furthermore, we move the derivative $\partial_{t'}$ onto the function f via integration by parts. We want to choose the initial time $t_{\rm in}$ long before decoupling and t_0 denotes today. Therefore, $\kappa(t_{\rm in}, t_0)$ is huge and we can completely neglect the term from the initial condition. Since early times do not contribute, we can formally start the integral at $t_{\rm in} = 0$. We can also neglect the boundary terms since the terms from the upper boundary $t' = t_0$ contribute only to the uninteresting monopole and dipole terms.

Let us introduce the visibility function g, defined by

$$g(t) \equiv a\sigma_T n_e e^{-\kappa(t,t_0)} \equiv \dot{\kappa} e^{-\kappa(t)}\,. \tag{4.124}$$

This function is very small at early times, when the optical depth, κ is very large. During decoupling, κ becomes smaller but also the pre-factor, $a\sigma_T n_e = \dot{\kappa}$ then becomes small. Therefore, g is strongly peaked during decoupling and small both before and after, see Fig. 4.1. With the above mentioned integration by parts we then find

$$\mathcal{M}^{(S)}(t_0, \mathbf{k}, \mu) = \int_0^{t_0} dt\, e^{ik\mu(t-t_0)} S^{(S)}(t, \mathbf{k})\,, \tag{4.125}$$

with

$$S^{(S)} = -e^{-\kappa}(\dot{\Phi} + \dot{\Psi}) + g\left(+\Phi + \Psi + k^{-1}\dot{V}^{(b)} + \frac{1}{4}D_g^{(r)} + \frac{1}{4}\mathcal{M}_2\right) + k^{-1}\dot{g}V^{(b)} - \frac{3}{4k^2}\frac{d^2}{dt^2}\left(g\mathcal{M}_2^{(S)}\right) . \quad (4.126)$$

Rewriting the exponential in terms of spherical Bessel functions, we now obtain simply

$$\mathcal{M}_\ell^{(S)}(t_0, \mathbf{k}) = \int_0^{t_0} dt\; j_\ell(k(t_0 - t))S^{(S)}(t, \mathbf{k}) . \quad (4.127)$$

Together with Eq. (4.118) this yields the scalar contributions to the CMB power spectrum, once the scalar source term is given. In this chapter we still neglect the effect of polarization. As we shall see in the next chapter, including it simply leads to a slight modification of the source term $S^{(S)}$. Apart from the gravitational contribution $\dot{\Phi} + \dot{\Psi}$ which gives rise to the integrated Sachs–Wolfe effect, all the terms are multiplied with the visibility function g or its derivatives, which are strongly peaked around the decoupling era, see Fig. 4.1. In the limit when we neglect the angular dependence of Thomson scattering (the terms containing $\mathcal{M}_2$) and approximate g by a delta-function at decoupling, we recover the tight coupling approximation discussed in the previous section and in Chapter 2.

For a numerical calculation of the CMB anisotropy power spectrum, this method has become the method of choice: first, the source term is calculated via the Boltzmann hierarchy truncated at about $\ell = 10$. Then, the C_ℓs are computed via the line-of-sight integral (4.127) followed by integration over k, Eq. (4.118). Free streaming is now taken care of by the spherical Bessel functions which can be computed just once and then be stored. This is especially useful if one wants to compute many models as in the context of parameter estimation, see Chapter 6. Another advantage is that the source term varies much more slowly than the Bessel functions both in k and in time and it can therefore be sampled relatively sparsely and still lead to good accuracy. Also, not all the $\mathcal{M}_\ell$s have to be computed. It is usually sufficient to calculate every tenth ℓ and to interpolate smoothly between them. All these numerical advantages have been used in the publicly available codes CMBfast (Seljak & Zaldarriaga, 1996), CAMB (Lewis *et al.*, 2000) and CMBeasy (Doran, 2005). CMBfast is 'the original' from which the others are drawn. CAMB is presently the best maintained, updated of these codes and CMBeasy is the most user friendly.

The time integral of the source term in Eq. (4.127) will smear out and damp fluctuations with wavelengths smaller than the width of the visibility function g. This phenomenon, called 'Silk damping' will be discussed in more detail in Section 4.6.

But first we want to derive the Boltzmann hierarchy and its solution via line-of-sight integration also for vector and tensor perturbations.

4.5.2 Vector perturbations

For vector perturbations, the Boltzmann equation (4.113) becomes

$$\dot{\mathcal{M}}^{(V)} + ik\mu\mathcal{M}^{(V)} = -n^i n^j a^{-1}\sigma_{ij}^{(V)} + a\sigma_T n_e\left[n^i\Omega_i^{(b)} - \mathcal{M}^{(V)} + \frac{1}{2}n^{ij}M_{ij}\right]. \tag{4.128}$$

As before we decompose $\mathbf{n}$ into

$$\begin{aligned}\mathbf{n} &= \mu\hat{\mathbf{k}} + \sqrt{1-\mu^2}\,(\cos\varphi\mathbf{e}_1 + \sin\varphi\mathbf{e}_2) = \mu\hat{\mathbf{k}} + n^{(+)}\mathbf{e}^{(+)} + n^{(-)}\mathbf{e}^{(-)} \\ &= \mu\hat{\mathbf{k}} + \sqrt{\frac{1-\mu^2}{2}}\left(e^{-i\varphi}\mathbf{e}^{(+)} + e^{i\varphi}\mathbf{e}^{(-)}\right) .\end{aligned} \tag{4.129}$$

We use the splitting of $\mathcal{M}^{(V)}$ and $-n^i n^j\sigma_{ij}^{(V)}$ into helicity modes as in Eqs. (4.86) and (4.89)

$$\mathcal{M}^{(V)} = \sqrt{\frac{1-\mu^2}{2}}\left[\exp(i\varphi)\mathcal{M}^{(V+)} + \exp(-i\varphi)\mathcal{M}^{(V-)}\right], \tag{4.130}$$

$$\mathcal{M}^{(V+)} = \sum_\ell (-i)^\ell(2\ell+1)\mathcal{M}_\ell^{(V+)}P_\ell(\mu), \tag{4.131}$$

$$\mathcal{M}^{(V-)} = \sum_\ell (-i)^\ell(2\ell+1)\mathcal{M}_\ell^{(V-)}P_\ell(\mu)\,, \tag{4.132}$$

$$a^{-1}n^i n^j\sigma_{ij}^{(V)} = \frac{-i}{\sqrt{2}}\mu\sqrt{1-\mu^2}\left(\sigma^{(V+)}e^{i\varphi} + \sigma^{(V-)}e^{-i\varphi}\right) . \tag{4.133}$$

For the expansion in Legendre polynomials we have used the fact that the coefficients $\mathcal{M}^{(V\pm)}$ depend on $\mathbf{n}$ only via μ.

As in the case of scalar perturbations, statistical homogeneity and isotropy require that the random variables $\mathcal{M}_\ell^{(V\pm)}(\mathbf{k})$ are uncorrelated for different values $\mathbf{k}$ and ℓ. Furthermore, we want to consider parity invariant perturbations, hence also $\mathcal{M}_\ell^{(V+)}(\mathbf{k})$ and $\mathcal{M}_\ell^{(V-)}(\mathbf{k})$ are uncorrelated and they have the same spectrum,

$$\begin{aligned}\left\langle\mathcal{M}_\ell^{(V+)}(\mathbf{k})\mathcal{M}_{\ell'}^{(V+)*}(\mathbf{k}')\right\rangle &= \left\langle\mathcal{M}_\ell^{(V-)}(\mathbf{k})\mathcal{M}_{\ell'}^{(V-)*}(\mathbf{k}')\right\rangle \\ &= (2\pi)^3\,\delta^3(\mathbf{k}-\mathbf{k}')\delta_{\ell\ell'}M_\ell^{(V)}(k)\,.\end{aligned} \tag{4.134}$$

To relate this spectrum to the vector C_ℓs we use, as for scalar perturbations,

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(t_0,\mathbf{n})\frac{\Delta T}{T}(t_0,\mathbf{n}')\right\rangle^{(V)} \\
&\quad = \frac{1}{4\pi}\sum_\ell (2\ell+1)C_\ell^{(V)}P_\ell(\mathbf{n}\cdot\mathbf{n}') \\
&\quad = \frac{1}{(2\pi)^6}\int d^3k\, d^3k'\, \langle \mathcal{M}^{(V)}(\mathbf{k},\mathbf{n})\mathcal{M}^{(V)*}(\mathbf{k}',\mathbf{n}')\rangle e^{i\mathbf{x}_0(\mathbf{k}-\mathbf{k}')} \\
&\quad = \sum_\ell \frac{(2\ell+1)^2}{(2\pi)^3}\int d^3k\, P_\ell(\mu)P_\ell(\mu')\sqrt{(1-\mu^2)(1-\mu'^2)}\cos(\varphi-\varphi')M_\ell^{(V)}(k)\,,
\end{aligned} \tag{4.135}$$

where $\mu = \hat{\mathbf{k}}\cdot\mathbf{n}$ and $\mu' = \hat{\mathbf{k}}\cdot\mathbf{n}'$ and φ and φ' are the angles defined in the decomposition on $\mathbf{n}$ and $\mathbf{n}'$ respectively according to Eq. (4.129). The first equals sign is just the definition of the $C_\ell^{(V)}$s and after the second equals sign we have inserted the Fourier representation of $\Delta T/T$. Using the decomposition (4.129) one finds

$$\mathbf{n}\cdot\mathbf{n}' = \mu\mu' + \sqrt{(1-\mu^2)(1-\mu'^2)}\cos(\varphi-\varphi')\,. \tag{4.136}$$

We therefore have $\sqrt{(1-\mu^2)(1-\mu'^2)}\cos(\varphi-\varphi') = \mathbf{n}\cdot\mathbf{n}' - \mu\mu'$. The term $\mathbf{n}\cdot\mathbf{n}'$ is independent of $\mathbf{k}$ and can be taken out of the integral. The terms containing additional factors μ and μ' respectively can be absorbed with the help of the recurrence relations of Legendre polynomials and again using the addition theorem of spherical harmonics. We finally arrive at

$$C_\ell^{(V)} = \frac{2\ell(\ell+1)}{\pi(2\ell+1)^2}\int dk\, k^2\left(M_{\ell+1}^{(V)} + M_{\ell-1}^{(V)}\right)\,. \tag{4.137}$$

The details of the derivation are developed in Ex. 4.3.

A short calculation shows that the vector perturbations of the energy–momentum tensor are given in terms of the expansion coefficients $\mathcal{M}_\ell^{(V+)}$ and $\mathcal{M}_\ell^{(V-)}$ by

$$\Omega^\pm = \mathcal{M}_0^{(V\pm)} + \mathcal{M}_2^{(V\pm)}\,, \tag{4.138}$$

$$\Pi^{(V\pm)} = \frac{24}{5}\left[\mathcal{M}_1^{(V\pm)} + \mathcal{M}_3^{(V\pm)}\right]\,. \tag{4.139}$$

To write the Boltzmann equation with the help of moments of $\mathcal{M}^{(V\pm)}$ we still need $n^{ij}M_{ij}$. A short calculation shows that only $M_{13} = M_{31}$ and $M_{23} = M_{32}$ do not vanish. Using the basic properties of the Legendre polynomials (see Appendix A4.1)

we obtain ($\mathbf{k} = k\mathbf{e}_3$)

$$M_{\pm 3} \equiv M_{13} \mp i M_{23} = \frac{3}{8}\int_{-1}^{1} d\mu\, \mu(1-\mu^2)\mathcal{M}^{(V\pm)} = -\frac{3i}{10}\left(\mathcal{M}_1^{(V\pm)} + \mathcal{M}_3^{(V\pm)}\right) .$$

With the definitions (4.130)–(4.133) and (4.138) the Boltzmann equation can then be written as

$$\dot{\mathcal{M}}^{(V\pm)} + ik\mu\mathcal{M}^{(V\pm)} + \dot{\kappa}\mathcal{M}^{(V\pm)} = i\mu\sigma^{(V\pm)} + \dot{\kappa}\left[\Omega^{(\pm)} - i\mu\frac{3}{10}\left(\mathcal{M}_1^{(V\pm)} + \mathcal{M}_3^{(V\pm)}\right)\right], \tag{4.140}$$

where κ denotes the optical depth $\kappa(t) = \sigma_T \int_t^{t_0} a n_e\, dt'$. As in the case of scalar perturbations this yields a Boltzmann hierarchy equation for the $\mathcal{M}_\ell^{(V\epsilon)}$s,

$$\dot{\mathcal{M}}_\ell^{(V\pm)} + \frac{k}{2\ell+1}\left[(\ell+1)\mathcal{M}_{\ell+1}^{(V\pm)} - \ell\mathcal{M}_{\ell-1}^{(V\pm)}\right] = -\dot{\kappa}\mathcal{M}_\ell^{(V\pm)} + \delta_{\ell 0}\dot{\kappa}\Omega^{(\pm)} + \delta_{\ell 1}\left[\frac{-1}{3}\sigma^{(V\pm)} + \dot{\kappa}\frac{1}{10}\left(\mathcal{M}_1^{(V\pm)} + \mathcal{M}_3^{(V\pm)}\right)\right]. \tag{4.141}$$

Also for vector perturbations, the most rapid way of solving the equations numerically is to solve the above hierarchy only for the lowest few multipoles in order to determine the source term, the right-hand side of Eq. (4.141). For a given source term Eq. (4.140) is then easily solved by line-of-sight integration

$$\mathcal{M}^{(V\pm)}(t_0, \mathbf{k}, \mu) = \int_0^{t_0} dt\, e^{ik\mu(t-t_0)-\kappa}\Big[- i\mu\sigma^{(V\pm)} + \dot{\kappa}\Big(\Omega^{(\pm)} - i\mu\frac{3}{10}\Big(\mathcal{M}_1^{(V\pm)} + \mathcal{M}_3^{(V\pm)}\Big)\Big)\Big]. \tag{4.142}$$

Absorbing the factors μ into time derivatives as in the scalar case, we find

$$\mathcal{M}^{(V\pm)}(t_0, \mathbf{k}, \mu) = \int_0^{t_0} dt\, e^{ik\mu(t-t_0)}\Big[+ k^{-1}e^{-\kappa}\dot{\sigma}^{(V\pm)} + g\Big(\Omega^{(\pm)} - k^{-1}\sigma^{(V\pm)} + \frac{3}{10k}\Big(\dot{\mathcal{M}}_1^{(V\pm)} + \dot{\mathcal{M}}_3^{(V\pm)}\Big)\Big) + \dot{g}\frac{3}{10k}\Big(\mathcal{M}_1^{(V\pm)} + \mathcal{M}_3^{(V\pm)}\Big)\Big]. \tag{4.143}$$

We have again used the visibility function g defined in Eq. (4.124). The expansion of the exponential in terms of spherical Bessel functions and Legendre polynomials

reproduces the $\mathcal{M}_\ell^{(V\pm)}$s,

$$\mathcal{M}_\ell^{(V\pm)}(t_0,\mathbf{k}) = \int_0^{t_0} dt\; j_\ell(k\mu(t_0-t))\Big[+ k^{-1}e^{-\kappa}\dot{\sigma}^{(V\pm)} + g\Big(\Omega^{(\pm)} - k^{-1}\sigma^{(V\pm)} + \frac{3}{10k}\Big(\dot{\mathcal{M}}_1^{(V\pm)} + \dot{\mathcal{M}}_3^{(V\pm)}\Big)\Big) + \dot{g}\,\frac{3}{10k}\Big(\mathcal{M}_1^{(V\pm)} + \mathcal{M}_3^{(V\pm)}\Big)\Big]\,. \tag{4.144}$$

4.5.3 Tensor perturbations

The Boltzmann equation (4.113) for tensor perturbations finally, has the form

$$\dot{\mathcal{M}}^{(T)} + ik\mu\mathcal{M}^{(T)} = -n^i n^j \dot{H}_{ij}^{(T)} + a\sigma_T n_e\left[\frac{1}{2}n^{ij}M_{ij}^{(T)} - \mathcal{M}^{(T)}\right]\,. \tag{4.145}$$

Since $H_{ij}^{(T)}$ is entirely orthogonal to $\mathbf{k}$, it is of the form $H_{ij}^{(T)} = \hat{H}_{ab}^{(T)} e_i^a e_j^b$, where $\mathbf{e}^1$ and $\mathbf{e}^2$ denote the two polarization directions normal to $\mathbf{k}$. Using the decomposition Eq. (4.129) for $\mathbf{n}$, the gravitational source term in Eq. (4.145) is seen to be the time derivative of

$$\begin{aligned} n^i n^j H_{ij}^{(T)} &= (1-\mu^2)\left[\widehat{H}_{11}^{(T)}\cos^2\varphi + \widehat{H}_{22}^{(T)}\sin^2\varphi + 2\widehat{H}_{12}^{(T)}\cos\varphi\sin\varphi\right] \\ &= (1-\mu^2)\left[H_d\cos(2\varphi) + H_\times\sin(2\varphi)\right]\,. \end{aligned} \tag{4.146}$$

For the last equals sign we have used that $\hat{H}_{22}^{(T)} = -\widehat{H}_{11}^{(T)}$ and we have introduced the usual notation for the two polarizations of a gravity wave propagating in direction $\hat{\mathbf{k}}$, $H_d \equiv \widehat{H}_{11}^{(T)}$ and $\hat{H}_{12}^{(T)} \equiv H_\times$.

This motivates our ansatz for the tensor perturbations of the temperature anisotropy,

$$\begin{aligned} \mathcal{M}^{(T)}(\mathbf{k},\mathbf{n}) = (1-\mu^2)\big[&\mathcal{M}^{(Td)}(\mathbf{k},\mu)\cos(2\varphi) \\ &+ \mathcal{M}^{(T\times)}(\mathbf{k},\mu)\sin(2\varphi)\big]\,. \end{aligned} \tag{4.147}$$

The coefficients $\mathcal{M}^{(T\bullet)}$ only depend on μ and can thus be expanded in Legendre polynomials,

$$\mathcal{M}^{(T\bullet)} = \sum_\ell (2\ell+1)(-i)^\ell \mathcal{M}_\ell^{(T\bullet)}(\mathbf{k})P_\ell(\mu)\,. \tag{4.148}$$

Here $\bullet$ denotes either d or $\times$. Statistical homogeneity and isotropy again imply that expansion coefficients with different values of $\mathbf{k}$ or for different ℓs are uncorrelated. Furthermore, also requiring invariance under parity, implies that the two

polarizations are uncorrelated and have the same spectra,

$$\left\langle \mathcal{M}_\ell^{(T\,\bullet)}(\mathbf{k})\mathcal{M}_{\ell'}^{(T\,\bullet)\,*}(\mathbf{k}')\right\rangle = (2\pi)^3\delta^3(\mathbf{k}-\mathbf{k}')\delta_{\ell\ell'}M_\ell^{(T)}(k)\,. \tag{4.149}$$

The relation of this tensor spectrum to the C_ℓs is obtained with the same reasoning as for the scalar and vector modes:

$$\begin{aligned}\frac{1}{4\pi}\sum_\ell (2\ell+1)C_\ell^{(T)}P_\ell(\mathbf{n}\cdot\mathbf{n}') &= \frac{1}{(2\pi)^6}\int d^3k d^3k' \langle \mathcal{M}^{(T)}(\mathbf{k},\mathbf{n})\mathcal{M}^{(T)*}(\mathbf{k}',\mathbf{n}')\rangle e^{-\mathbf{x}(\mathbf{k}-\mathbf{k}')} \\ &= \frac{1}{(2\pi)^3}\sum_\ell (2\ell+1)^2\int d^3k M_\ell^{(T)}(k)P_\ell(\mu)P_\ell(\mu') \\ &\quad\times(1-\mu^2)(1-\mu'^2)[\cos(2\varphi)\cos(2\varphi')+\sin(2\varphi)\sin(2\varphi')]\,, \end{aligned}\tag{4.150}$$

where $\mu = \hat{\mathbf{k}}\cdot\mathbf{n}$ and $\mu' = \hat{\mathbf{k}}\cdot\mathbf{n}'$. The angles φ and φ' are those appearing in the decomposition of $\mathbf{n}$ and $\mathbf{n}'$ according to Eq. (4.129).

Again using Eq. (4.136) and $\cos(2\varphi)\cos(2\varphi')+\sin(2\varphi)\sin(2\varphi') = \cos(2(\varphi-\varphi'))$, the φ-dependence can be written as a function of $\mathbf{n}\cdot\mathbf{n}'$ and μ and μ'. The recurrence relations for the Legendre polynomials can then be applied to absorb the factors μ and μ' and with the addition theorem of spherical harmonics, we arrive after a somewhat lengthy calculation at

$$C_\ell^{(T)} = \frac{2}{\pi}\frac{(\ell+2)!}{(\ell-2)!}\int dk\,k^2\frac{\Sigma^{(T)}(k)}{(2\ell+1)^2}\,, \qquad \text{with} \tag{4.151}$$

$$\Sigma^{(T)}(k) = \frac{M_{\ell-2}^{(T)}}{(2\ell-1)^2} + \frac{4(2\ell+1)^2M_\ell^{(T)}}{[(2\ell-1)(2\ell+3)]^2} + \frac{M_{\ell+2}^{(T)}}{(2\ell+3)^2}\,. \tag{4.152}$$

The details of this result are developed in Ex. 4.4.

We also express the tensor anisotropic stress in terms of the expansion coefficients $\mathcal{M}^{(T\bullet)}$, we use that it is transverse to $\mathbf{k}$.

$$\Pi_r^{(T\bullet)} = \frac{3}{2}\int_{-1}^{1} d\mu\,(1-\mu^2)^2\mathcal{M}^{(T\bullet)} = \frac{24}{35}\mathcal{M}_4^{(T\bullet)} + \frac{16}{7}\mathcal{M}_2^{(T\bullet)} + \frac{8}{5}\mathcal{M}_0^{(T\bullet)}\,. \tag{4.153}$$

With the ansatz (4.147) we find that for tensor perturbations

$$\begin{aligned} n^{ij}M_{ij} = (1-\mu^2)\Bigg\{&\cos(2\varphi)\left[\frac{3}{35}\mathcal{M}_4^{(Td)} + \frac{2}{7}\mathcal{M}_2^{(Td)} + \frac{1}{5}\mathcal{M}_0^{(Td)}\right] \\ &+ \sin(2\varphi)\left[\frac{3}{35}\mathcal{M}_4^{(Tx)} + \frac{2}{7}\mathcal{M}_2^{(TX)} + \frac{1}{5}\mathcal{M}_0^{(TX)}\right]\Bigg\}\,. \end{aligned}\tag{4.154}$$

Inserting this in the Boltzmann equation we obtain

$$\dot{\mathcal{M}}^{(T\,\bullet)} + ik\mu\mathcal{M}^{(T\,\bullet)} = -\dot{H}_\bullet + \dot{\kappa}\left[\frac{3}{70}\mathcal{M}_4^{(T\bullet)} + \frac{1}{7}\mathcal{M}_2^{(T\bullet)} + \frac{1}{10}\mathcal{M}_0^{(T\bullet)} - \mathcal{M}^{(T\bullet)}\right], \tag{4.155}$$

with the line-of-sight 'solution'

$$\mathcal{M}^{(T\,\bullet)}(t_0, \mathbf{k}, \mu) = \int_0^{t_0} dt\, e^{ik(t-t_0)-\kappa}\Bigg[-\dot{H}_\bullet + \dot{\kappa}\left(\frac{3}{70}\mathcal{M}_4^{(T\bullet)} + \frac{1}{7}\mathcal{M}_2^{(T\bullet)} + \frac{1}{10}\mathcal{M}_0^{(T\bullet)}\right)\Bigg]. \tag{4.156}$$

Of course for this to solve the equation, the first moments, $\mathcal{M}_0^{(T\bullet)}$ to $\mathcal{M}_4^{(T\bullet)}$ which also determine $\dot{H}_\bullet$ have to be calculated via the Boltzmann hierarchy which in this case is

$$\dot{\mathcal{M}}_\ell^{(T\bullet)} + \frac{k}{2\ell+1}\left[(\ell+1)\mathcal{M}_{\ell+1}^{(T\bullet)} - \ell\mathcal{M}_{\ell-1}^{(T\bullet)}\right] = -\dot{\kappa}\mathcal{M}_\ell^{(T\bullet)} + \delta_{\ell 0}\left[-\dot{H}_\bullet + \dot{\kappa}\left(\frac{3}{70}\mathcal{M}_4^{(T\bullet)} + \frac{1}{7}\mathcal{M}_2^{(T\bullet)} + \frac{1}{10}\mathcal{M}_0^{(T\bullet)}\right)\right]. \tag{4.157}$$

The coefficients $\mathcal{M}_\ell^{(T\bullet)}$ are now given simply by

$$\mathcal{M}_\ell^{(T\,\bullet)}(t_0, \mathbf{k}) = \int_0^{t_0} dt\; j_\ell(k(t_0 - t))S_\ell^{(T)}, \qquad \text{with} \tag{4.158}$$

$$S_\ell^{(T)} = e^{-\kappa}\left[-\dot{H}_\bullet + \delta_{0\ell}\dot{\kappa}\left(\frac{3}{70}\mathcal{M}_4^{(T\bullet)} + \frac{1}{7}\mathcal{M}_2^{(T\bullet)} + \frac{1}{10}\mathcal{M}_0^{(T\bullet)}\right)\right]. \tag{4.159}$$

Also here, the only modification which this solution will experience once we include polarization, is a change in the source term $S^{(T)}$, to which contributions from the polarization spectrum will have to be added (see Chapter 5).

4.6 Silk damping

In this section we want to discuss, in more detail, the damping on small scales which appears when the coupling between photons and the baryon/electron gas is still present, but no longer perfect. We therefore do not want to describe the photon–baryon system as a perfect fluid, but want to take into account the force provided by the Thomson scattering of electrons and photons. This leads to an additional force in the baryon equation of motion (Eq. (2.118) for $w = c_s^2 = 0$), the photon drag force due to Thomson scattering. For simplicity, and since this is the relevant

case, we consider only scalar perturbations in this section. For them the photon drag force is given by

$$F_j = -\frac{\rho_r}{\pi}\int C[\mathcal{M}] n_j \, d\Omega_{\mathbf{n}} \,, \tag{4.160}$$

$$\hat{\mathbf{k}}\cdot\mathbf{F} = \frac{-4i\sigma_T n_e a \rho_r}{3}\left(3\mathcal{M}_1 - V^{(b)}\right) \,. \tag{4.161}$$

For the second equals sign we have used Eq. (4.111) and integrated $\mathbf{n}C[\mathcal{M}] = \mathbf{n}C'[\mathcal{M}]a$ over angles. From the expansion of $\mathcal{M}$ in Legendre polynomials we know that

$$(-i)^\ell \mathcal{M}_\ell = \frac{1}{2}\int d\mu \, P_\ell(\mu)\mathcal{M}(\mu) \,,$$

so that

$$\mathcal{M}_1 = \frac{i}{2}\int d\mu \, \mu\mathcal{M}(\mu) \quad \text{and} \quad \frac{1}{4}D_g^{(r)} = \frac{1}{2}\int d\mu \, \mathcal{M} \,.$$

Adding the drag force to the baryon equation of motion yields (in Fourier space)

$$\dot{\mathbf{V}}^{(b)} + \mathcal{H}\mathbf{V}^{(b)} = -i\mathbf{k}\Psi + \rho_b^{-1}\mathbf{F} \,. \tag{4.162}$$

To discuss damping, we are only interested in small scales $kt \gg 1$ and therefore shall neglect the expansion of the Universe in our treatment. It then makes sense to model the time dependence of our variables with an exponential, $V, \ \mathcal{M} \propto e^{-i\omega t}$. Furthermore, we consider the epoch when there are still many collisions per Hubble expansion. Denoting the collision time by $t_c = 1/\dot{\kappa} = 1/(a\sigma_T n_e)$, this means $t \gg t_c$. For simplicity, we also neglect gravitational terms and the term $n^{ij}M_{ij}$ which is due to the direction dependence of Thomson scattering and is not important as long as scattering is sufficiently abundant. With the ansatz of a harmonic time dependence with frequency ω, we then obtain from Eqs. (4.122) and (4.162) with $\hat{\mathbf{k}}\cdot\mathbf{V} = +iV^{(b)}$,

$$-it_c(\omega - k\mu)\mathcal{M} = \frac{1}{4}D_g^{(r)} - i\mu V^{(b)} - \mathcal{M} \,, \tag{4.163}$$

$$t_c\omega V^{(b)} = \frac{-4i\rho_r}{3\rho_b}\left[3\mathcal{M}_1 - V^{(b)}\right] \,. \tag{4.164}$$

Therefore, integrating the Boltzmann equation (4.163) over μ, yields

$$\mathcal{M}_1 = \frac{i\omega}{4k}D_g^{(r)} \,. \tag{4.165}$$

Inserting this in Eq. (4.164) we find, with $R \equiv 3\rho_b/4\rho_r$,

$$V^{(b)} = \frac{+3i\omega D_g^{(r)}}{4k(1 - it_c\omega R)} \,. \tag{4.166}$$

Inserting this result for $V^{(b)}$ in Eq. (4.163) we obtain

$$\mathcal{M} = \frac{1 + \frac{3\mu\omega/k}{1-it_c\omega R}}{1 - it_c(\omega - k\mu)} \frac{D_g^{(r)}}{4} \,. \tag{4.167}$$

Integrating this equation over μ yields a dispersion relation for $\omega(k)$ in the form

$$\begin{aligned} 1 &= \frac{1}{2}\int_{-1}^{1} d\mu \, \frac{1 + \frac{3\mu\omega/k}{1-it_c\omega R}}{1 - it_c(\omega - k\mu)} \\ &= \frac{3\omega}{it_c k^2 + t_c^2 k^2 \omega R} + \frac{1}{2}\left(\frac{1}{it_c k} + \frac{3\omega}{t_c^2 k^3}\frac{1 - it_c\omega}{1 - it_c\omega R}\right) \\ &\quad \times [\ln(1 + it_c(k - \omega)) - \ln(1 - it_c(\omega + k))] \,. \end{aligned} \tag{4.168}$$

This equation cannot be solved analytically. If we expand it in $t_c k$ and $t_c\omega$ we find to lowest non-vanishing order

$$\omega(k) = k\left(\frac{1}{\sqrt{3(R+1)}} - i\frac{kt_c}{6}\frac{R^2 + \frac{4}{5}(R+1)}{(R+1)^2}\right) \,. \tag{4.169}$$

The real part of $\omega(k)$ describes oscillations and $\mathrm{Re}(\omega)/k$ is the group velocity of the oscillations. The imaginary term is a damping term. Over a time interval $t_d = \frac{6}{kt_c}\frac{(R+1)^2}{R^2+\frac{4}{5}(R+1)}$, the amplitude is reduced by one e-fold. This is Silk damping, due to the imperfect coupling of electrons and photons. It vanishes in the limit $kt_c \to 0$. It is interesting to note that one has to expand Eq. (4.168) to sixth order in ωt_c and kt_c to find this relation (see Ex. 4.5). This indicates that damping is effective only for $t_c k$ relatively close to 1.

4.7 The full system of perturbation equations

We terminate this chapter by writing down the full system of perturbation equations in a 'standard' universe containing dark matter, baryons, photons, massless neutrinos and a cosmological constant. The latter only influences the background evolution and does not appear in the perturbation equations. Even though we know that neutrinos are not truly massless, since their mass scale may be as low as 0.01 eV, it is most probably irrelevant for CMB anisotropies. We thus neglect it here. In standard inflationary models only scalar and tensor modes are generated, we therefore restrict this recapitulation to them. The fluid equations for dark matter and baryons and the Einstein equations have been derived in Chapter 2. The Boltzmann equation for photons and the evolution equation for neutrinos, which is simply the Liouville equation, have been derived in this chapter.

The evolution of cold dark matter perturbations, D_c and V_c, is determined by the energy–momentum conservation equations

$$\dot{D}_c = -kV_c \,, \tag{4.170}$$

$$\dot{V}_c + \mathcal{H}V_c = k\Psi \,. \tag{4.171}$$

For the evolution of baryons, we have also to take into account the photon drag force leading to

$$\dot{D}_b = -kV_b \,, \tag{4.172}$$

$$\dot{V}_b + \mathcal{H}V_b = k\Psi + \frac{4\dot{\kappa}\rho_r}{3\rho_b}(3\mathcal{M}_1 - V_b) \,. \tag{4.173}$$

For the low multipoles, $\ell < 10$, say, we have to solve the Boltzmann hierarchies

$$\dot{\mathcal{M}}_\ell^{(S)} + k\frac{\ell+1}{2\ell+1}\mathcal{M}_{\ell+1}^{(S)} - k\frac{\ell}{2\ell+1}\mathcal{M}_{\ell-1}^{(S)} = \frac{1}{3}\delta_{\ell 1}\left[-k(\Phi+\Psi) + \dot{\kappa}V^{(b)}\right] + \dot{\kappa}\left[\frac{1}{2}\delta_{\ell 2}\mathcal{M}_2^{(S)} - \mathcal{M}_\ell^{(S)}\right], \tag{4.174}$$

for scalar perturbations, and

$$\dot{\mathcal{M}}_\ell^{(T\bullet)} + \frac{k}{2\ell+1}\left[(\ell+1)\mathcal{M}_{\ell+1}^{(T\bullet)} - \ell\mathcal{M}_{\ell-1}^{(T\bullet)}\right] = -\dot{\kappa}\mathcal{M}_\ell^{(T\bullet)} + \delta_{\ell 0}\left[-\dot{H}_\bullet + \dot{\kappa}\left(\frac{3}{35}\mathcal{M}_4^{(T\bullet)} + \frac{2}{7}\mathcal{M}_2^{(T\bullet)} + \frac{1}{5}\mathcal{M}_0^{(T\bullet)}\right)\right], \tag{4.175}$$

for tensor perturbations. The higher multipoles, ℓ can then be obtained via the integrals Eqs. (4.127) and (4.156).

Neutrino perturbations have to be treated via the collisionless Boltzmann equation. Setting the collision term to zero and denoting the neutrino perturbation of the distribution function integrated over energies by $\mathcal{N}$, we obtain, by exactly the same steps as explained in the previous sections for photons,

$$\dot{\mathcal{N}}_\ell^{(S)} + \frac{k}{2\ell+1}\left[(\ell+1)\mathcal{N}_{\ell+1}^{(S)} - \ell\mathcal{N}_{\ell-1}^{(S)}\right] = \frac{-k}{3}\delta_{\ell 1}(\Phi+\Psi) \,, \tag{4.176}$$

for scalar perturbations, and

$$\dot{\mathcal{N}}_\ell^{(T\bullet)} + \frac{k}{2\ell+1}\left[(\ell+1)\mathcal{N}_{\ell+1}^{(T\bullet)} - \ell\mathcal{N}_{\ell-1}^{(T\bullet)}\right] = -\delta_{\ell 0}\dot{H}_\bullet \,, \tag{4.177}$$

for tensor perturbations.

The scalar and tensor metric perturbations are determined by Einstein's equations,

$$-k^2\Phi = 4\pi G a^2 \rho D \,, \tag{4.178}$$

$$k^2(\Phi - \Psi) = 4\pi G a^2 \left(P_r\Pi_r^{(S)} + P_\nu\Pi_\nu^{(S)}\right) \,, \qquad \text{and} \tag{4.179}$$

$$\ddot{H}_\bullet + 2\mathcal{H}\dot{H}_\bullet + k^2 H_\bullet = 8\pi G a^2 \left(P_r\Pi_{r\bullet}^{(T)} + P_\nu\Pi_{\nu\bullet}^{(T)}\right) \,. \tag{4.180}$$

The scalar and tensor anisotropic stresses are given by

$$\Pi_r^{(S)} = 12\mathcal{M}_2^{(S)} , \tag{4.181}$$

$$\Pi_\nu^{(S)} = 12\mathcal{N}_2^{(S)} , \tag{4.182}$$

$$\Pi_{r\bullet}^{(T)} = \frac{24}{35}\mathcal{M}_4^{(T)} + \frac{16}{7}\mathcal{M}_2^{(T)} + \frac{8}{5}\mathcal{M}_0^{(T)} , \tag{4.183}$$

$$\Pi_{\nu\bullet}^{(T)} = \frac{24}{35}\mathcal{N}_4^{(T)} + \frac{16}{7}\mathcal{N}_2^{(T)} + \frac{8}{5}\mathcal{N}_0^{(T)} . \tag{4.184}$$

The total density perturbation is

$$\rho D = \rho_c D_c + \rho_b D_b + \rho_r D_r + \rho_\nu D_\nu , \tag{4.185}$$

where

$$\begin{aligned} D_r &= D_g^{(r)} + 4k^{-1}\mathcal{H}V_r + 4\Phi , \\ &= 4(\mathcal{M}_0 + 3k^{-1}\mathcal{H}\mathcal{M}_1 + \Phi) , \end{aligned} \tag{4.186}$$

$$\begin{aligned} D_\nu &= D_{g\nu} + 4k^{-1}\mathcal{H}V_\nu + 4\Phi , \\ &= 4(\mathcal{N}_0 + 3k^{-1}\mathcal{H}\mathcal{N}_1 + \Phi) . \end{aligned} \tag{4.187}$$

For a given background evolution, the above is a closed set of perturbation equations which can be solved. One obtains a good approximation by truncating the hierarchies for the photons and neutrinos at about $\ell = 10$ and determining the higher moments via the line-of-sight integrals. For photons these are given in Eqs. (4.127) and (4.156). For neutrinos one obtains the same equations just setting $\dot{\kappa} = \kappa = 0$.

The initial conditions are determined by inflation. In order not to miss any of the physical processes which can influence perturbations once they enter the Hubble horizon, we choose the initial time $t_{\rm in}$ so that $kt_{\rm in} \ll 1$ for the mode k under study. Furthermore, we want to start deep in the radiation era so that we can use the results of Section 2.4.3. Requiring that perturbations remain regular for $t \to 0$ usually restricts us to the growing mode. Let us first consider scalar perturbations. On superhorizon scales the growing mode behaves as $\mathcal{M}_0 \propto \mathcal{N}_0 \propto$ constant. $\mathcal{M}_1 \propto \mathcal{N}_1 \propto kt$ and $\Phi = \Psi =$ constant. For the non-relativistic, subdominant species one can choose (for adiabatic perturbations) $V_c = V_b = V_r = 3\mathcal{M}_1$. The initial condition for D_b and D_c is then determined by Eqs. (4.172) and (4.170) with the condition that $D \to 0$ for $t \to 0$. Adiabaticity also requires $\mathcal{N}_1 = \mathcal{M}_1$. For adiabatic scalar perturbations this leaves us with one initial condition, which is usually given as the initial power spectrum for Ψ determined during inflation,

$$k^3\langle|\Psi|^2(k)\rangle = A_S(k/H_0)^{n_S-1} . \tag{4.188}$$

Here A_S is the square of the perturbation amplitude and n_S is the scalar spectral index. It is sufficient to start integration at $z \simeq 10^7 - 10^8$.

For tensor perturbations we can simply set $H_\bullet =$ constant and $\mathcal{N}_\ell \propto j_\ell(kt)$. At early times, the collision term suppresses the build up of higher moments in the photon distribution and imposes

$$\frac{4}{5}\mathcal{M}_0^{(T)} = \frac{2}{7}\mathcal{M}_2^{(T)} + \frac{3}{35}\mathcal{M}_4^{(T)} .$$

A simple possibility is $\mathcal{M}_0^{(T)} = \mathcal{N}_0^{(T)}$, $\mathcal{M}_2^{(T)} = \frac{14}{5}\mathcal{M}_0^{(T)}$ and $\mathcal{M}_4^{(T)} = 0$.

Of course one can also suggest some other initial conditions, e.g., the neutrino iso-curvature velocity mode, where $\mathcal{N}_1$ dominates.

From purely theoretical grounds one can define an initial perturbation which just induces a $\mathcal{N}_{13}^{(S)} \neq 0$ at some early time $t_{\rm in}$, while all other perturbation variables vanish. The system of equations presented here can then be solved given this initial condition. Such a condition is purely iso-curvature, since the energy–momentum perturbations vanish initially. But via Eq. (4.176), the perturbation will be induced in the lower moments of the neutrino distribution function and finally in the neutrino energy–momentum tensor. It then leads to perturbations of the gravitational field which in turn induce perturbations in the dark matter, the baryons and photons.

However, a physical mechanism leading to this kind of initial perturbations has not been proposed so far.

Exercises

(The exercises marked with an asterisk are solved in Appendix A10.4.)

Ex. 4.1 An orthonormal basis for a curved universe

Study the first section of Appendix 9.

Determine the transformation matrix $E_k{}^j(\mathbf{x})$ in the spatial basis $dx^2 = \gamma_{ij}\, dx^i\, dx^j = dr^2 + \chi^2(r)\left(d\theta^2 + \sin^2\theta\, d\varphi\right)$, with

$$\chi(r) = \begin{cases} \sin(r) & \text{for} \quad K = 1 \\ r & \text{for} \quad K = 0 \\ \sinh(r) & \text{for} \quad K = -1 \,. \end{cases}$$

Ex. 4.2 Vector perturbations of the CMB

Consider a vector perturbation spectrum of the form

$$\langle \sigma_i(\mathbf{k}) \sigma_j^*(\mathbf{k}') \rangle = (\delta_{ij} - \hat{\mathbf{k}}_i \hat{\mathbf{k}}_j) A k^{n_V}\, \delta(\mathbf{k} - \mathbf{k}') \,. \tag{4.189}$$

Using statistical isotropy (and symmetry under parity), explain why the k-space structure of the power spectrum has to be of this form.

Using the solution Eq. (4.144) in k-space, calculate the vector-type CMB anisotropies generated from σ_j. Which value of n_V leads to a scale-invariant

spectrum? i.e., for which n_V do you obtain $\ell(\ell+1)C_\ell \simeq$ constant for sufficiently large ℓs?

Ex. 4.3 The vector C_ℓs∗

Derive Eq. (4.137) from Eqs. (4.134) and (4.135).

Indication: use

$$\sqrt{(1-\mu^2)(1-\mu'^2)}\cos(\varphi-\varphi') = \mathbf{n}\cdot\mathbf{n}' - \mu\mu' \,, \tag{4.190}$$

where $\mu = \hat{\mathbf{k}}\cdot\mathbf{n}$ and $\mu' = \hat{\mathbf{k}}\cdot\mathbf{n}'$. Replace now terms $\mu P_\ell(\mu)$ via the recursion relations in terms of $P_{\ell+1}$ and $P_{\ell-1}$. Show, using the addition theorem for spherical harmonics given in Appendix A4.2.3 that

$$\int d\Omega_{\hat{\mathbf{k}}}\, P_\ell(\mu)P_{\ell'}(\mu') = \delta_{\ell\ell'}\frac{4\pi}{2\ell+1}P_\ell(\mathbf{n}\cdot\mathbf{n}') \,, \tag{4.191}$$

and use this relation to perform angular integrations. Using the recursion relation for $(\mathbf{n}\cdot\mathbf{n}')P_\ell(\mathbf{n}\cdot\mathbf{n}')$ and finally collecting the terms which multiply $P_\ell(\mathbf{n}\cdot\mathbf{n}')$ one obtains Eq. (4.135).

Ex. 4.4 The tensor C_ℓs

From Eqs. (4.147) and (4.150) derive Eq. (4.151).

Indication: follow exactly the same lines as for Ex. 4.3. Only this time the recursion formula has to be applied twice to reduce terms $\mu^2 P_\ell(\mu)$ and, at the end $(\mathbf{n}\cdot\mathbf{n}')^2 P_\ell(\mathbf{n}\cdot\mathbf{n}')$. The calculation is lengthy but straight forward.

Ex. 4.5 Silk damping

Derive the dispersion relation (4.169) from the integral (4.168).

Hint: use an algebraic program, like Maple or Mathematica to expand (4.168) up to sixth order in $t_c k$ and $t_c\omega$.

5

CMB polarization and the total angular momentum approach

The Thomson scattering cross section depends on the polarization of the outgoing photon. If its polarization vector lies in the scattering plane, the cross section is proportional to $\cos^2\beta$, where β denotes the scattering angle. If, however, the outgoing photon is polarized normal to the scattering plane, no such reduction by a factor $\cos^2\beta$ occurs (see Jackson (1975), Section 14.7). If photons come in isotropically from all directions, this does not lead to any net polarization of the outgoing radiation. However, if, for a fixed outgoing direction, the intensity of incoming photons from one direction is different from the intensity of photons coming in at a right angle with respect to the first direction and with respect to the direction of the outgoing photon (see Fig. 5.1), this anisotropy leads to some polarization of the outgoing photon beam. As it is clear from the figure, it is the quadrupole anisotropy in the reference frame of the scattering electron which is responsible for polarization.

In this chapter we discuss the induced polarization in detail. We derive the equations which govern the generation and propagation of polarization and we discuss their implications. This can be done by different methods, most of which are either rather involved or incomplete. Here we employ the so-called 'total angular momentum method' which has been developed in Hu & White (1997b) and Hu *et al.* (1998), based on previous work mainly by Seljak (1996b), Kamionkowski *et al.* (1997) and Zaldarriaga & Seljak (1997). Even though the derivation of the results is somewhat involved, it is straight forward, in the sense that there are no 'unexpected turns' in it. Nevertheless, readers who do not want to dwell in lengthy derivations may just read the first section and then go directly to the results which are given in the form of integral solutions at the end of the chapter. Computationally, this is the most difficult chapter of this book.

For our derivations we use spherical harmonics and spin weighted spherical harmonics. Also the basics of representation theory of the rotation group will be needed. All the notions on these topics which are employed here are presented in

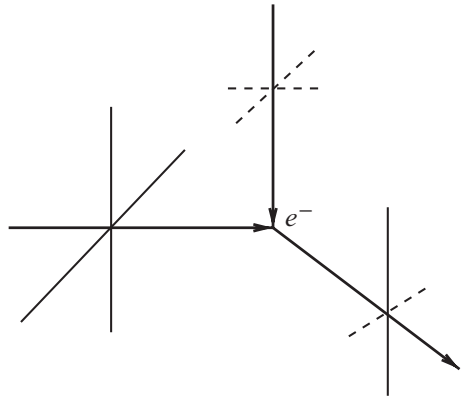


Fig. 5.1. More incoming photons from the left than from the top (indicated in the figure with longer polarization directions), lead to a net polarization of the outgoing photon beam, since the photons coming in from the left are preferentially polarized vertically, while the photons coming in from the top are preferentially polarized horizontally. In this way, an unpolarized photon distribution which exhibits a quadrupole anisotropy generates polarization on the surface of last scattering.

Appendix 4, especially A4.2. Some detailed derivations are also deferred to that appendix or to the exercises.

5.1 Polarization dependent Thomson scattering

5.1.1 The Stokes parameters

We consider an electromagnetic wave propagating in direction $\mathbf{n}$. We define the polarization directions $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\epsilon}^{(2)}$ such that $\left(\boldsymbol{\epsilon}^{(1)}, \boldsymbol{\epsilon}^{(2)}, \mathbf{n}\right)$ form a right-handed orthonormal system. The electric field of the wave is of the form $\mathbf{E} = E_1\boldsymbol{\epsilon}^{(1)} + E_2\boldsymbol{\epsilon}^{(2)}$. (The polarizations $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\epsilon}^{(2)}$ are not to be confused with $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ which were introduced in Chapter 4 to form an orthonormal system with the wave vector $\mathbf{k}$.) The polarization tensor of an electromagnetic wave is defined as

$$P_{ij} = \widetilde{\mathcal{P}}_{ab}\boldsymbol{\epsilon}_i^{(a)}\boldsymbol{\epsilon}_j^{(b)}\,, \quad \text{with} \quad \widetilde{\mathcal{P}}_{ab} = E_a^* E_b\,. \tag{5.1}$$

$\widetilde{\mathcal{P}}_{ab}$ is a hermitian 2×2 matrix and can therefore be written as

$$\begin{aligned}\widetilde{\mathcal{P}}_{ab} &= \frac{1}{2}\left[I\sigma_{ab}^{(0)} + U\sigma_{ab}^{(1)} + V\sigma_{ab}^{(2)} + Q\sigma_{ab}^{(3)}\right] \\ &= \frac{1}{2}I\sigma_{ab}^{(0)} + \mathcal{P}_{ab}\,,\end{aligned} \tag{5.2}$$

where $\sigma^{(\alpha)}$ denote the Pauli matrices and the four real functions of the photon direction $\mathbf{n}$, I, U, V and Q are the Stokes parameters.

$$\sigma^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^{(1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$\sigma^{(2)} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^{(3)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{5.3}$$

In terms of the electric field, the Stokes parameters are

$$I = |E_1|^2 + |E_2|^2, \quad Q = |E_1|^2 - |E_2|^2,$$
$$U = (E_1^* E_2 + E_2^* E_1) = 2\mathrm{Re}(E_1^* E_2), \quad V = 2\mathrm{Im}(E_1^* E_2). \tag{5.4}$$

I is simply the intensity of the electromagnetic wave. Q represents the amount of linear polarization in directions $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\epsilon}^{(2)}$, i.e., Q is the difference between the intensity of radiation polarized along $\boldsymbol{\epsilon}^{(1)}$ minus the intensity polarized in direction $\boldsymbol{\epsilon}^{(2)}$. The parameters Q and U describe the symmetric traceless part of the polarization tensor while V multiplies the anti-symmetric Pauli matrix $\sigma^{(2)}$. This part describes a phase difference between E_1 and E_2 which results in circular polarization. This is best seen by expressing $\mathcal{P}_{ab}$ in terms of the helicity basis $\boldsymbol{\epsilon}^{(\pm)} = (1/\sqrt{2})\left(\boldsymbol{\epsilon}^{(1)} \pm i\boldsymbol{\epsilon}^{(2)}\right)$, where one finds that V is the difference between the left- and right-handed circular polarized intensities (see e.g. Jackson, 1975). As we shall see below, Thomson scattering does not introduce circular polarization. We therefore expect the V-Stokes parameter of the CMB radiation to vanish. We neglect it in the following. If $V = 0$, we have $\mathcal{P}_{ab} = \mathcal{P}_{ab}^* = \mathcal{P}_{ba}$. Hence $\mathcal{P}_{ab}$ is a real symmetric matrix.

We often also use the quantities

$$P \equiv \mathcal{P}_{++} = 2\mathcal{P}^{ab}\boldsymbol{\epsilon}_a^{(+)}\boldsymbol{\epsilon}_b^{(+)} = Q + iU, \quad \text{and} \tag{5.5}$$
$$\bar{P} \equiv \mathcal{P}_{--} = 2\mathcal{P}^{ab}\bar{\boldsymbol{\epsilon}}_a^{(+)}\bar{\boldsymbol{\epsilon}}_b^{(+)} = 2\mathcal{P}^{ab}\boldsymbol{\epsilon}_a^{(-)}\boldsymbol{\epsilon}_b^{(-)} = Q - iU. \tag{5.6}$$

Up to a factor of 2, these are the components of the polarization tensor expressed in the helicity basis. One easily verifies that the off-diagonal terms vanish since they are proportional to V, $\mathcal{P}_{+-} = \mathcal{P}_{-+} \propto V = 0$.

The intensity is proportional to the energy density of the CMB, $\rho = \frac{1}{8\pi} I$ and therefore to our perturbation variable $\mathcal{M} = \delta T/T = \frac{1}{4}\delta\rho/\rho = \frac{1}{4}\delta I/I$. Correspondingly we define the dimensionless Stokes parameters

$$\mathcal{Q} \equiv \frac{Q}{4I} \quad \text{and} \quad \mathcal{U} \equiv \frac{U}{4I}. \tag{5.7}$$

Rotating the basis $\left(\boldsymbol{\epsilon}^{(1)}, \boldsymbol{\epsilon}^{(2)}\right)$ by an angle ψ around the direction $\mathbf{n}$ we obtain $\boldsymbol{\epsilon}^{(1)'} = \cos\psi\boldsymbol{\epsilon}^{(1)} + \sin\psi\boldsymbol{\epsilon}^{(2)}$ and $\boldsymbol{\epsilon}^{(2)'} = \cos\psi\boldsymbol{\epsilon}^{(2)} - \sin\psi\boldsymbol{\epsilon}^{(1)}$ so that the coefficients with

respect to the rotated basis are $E_1' = E_1 \cos\psi - E_2 \sin\psi$ and $E_2' = E_2 \cos\psi + E_1 \sin\psi$. For the Stokes parameters this implies

$$I' = I\,, \qquad V' = V \quad \text{and}$$
$$Q' = Q\cos 2\psi - U \sin 2\psi\,, \qquad U' = U\cos 2\psi + Q\sin 2\psi\,, \tag{5.8}$$

or more simply

$$Q' \pm iU' = e^{\pm 2i\psi}(Q \pm iU)\,. \tag{5.9}$$

Hence $Q \pm iU$ transform like spin-2 variables with a magnetic quantum number ± 2 under rotations around the $\mathbf{n}$ axis. They depend not only on the direction $\mathbf{n}$, but also on the orientation of the polarization basis $\left(\boldsymbol{\epsilon}^{(1)}, \boldsymbol{\epsilon}^{(2)}\right)$. For example, when rotating the polarization basis by $\pi/4$ we turn U into Q and Q into $-U$. Hence U measures the linear polarization in the basis $\left(\boldsymbol{\epsilon}^{(1)'}, \boldsymbol{\epsilon}^{(2)'}\right)$ which is rotated by $-\pi/4$ from the original basis.

The eigenvalues of $2\mathcal{P} = \begin{pmatrix} Q & U \\ U & -Q \end{pmatrix}$ are $\lambda_{1,2} = \pm\sqrt{Q^2+U^2}$ with eigenvectors

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = A \begin{pmatrix} Q + \sqrt{Q^2+U^2} \\ U \end{pmatrix}\,, \quad \text{and} \quad \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = A \begin{pmatrix} Q - \sqrt{Q^2+U^2} \\ U \end{pmatrix}\,.$$

Here $A \neq 0$ is an arbitrary constant. The first eigenvector encloses the angle ϕ_1 with the $\boldsymbol{\epsilon}^{(1)}$-axis which is given by

$$\tan(2\phi_1) = \frac{2\sin\phi_1\cos\phi_1}{\cos^2\phi_1 - \sin^2\phi_1} = \frac{2x_1y_1}{x_1^2 - y_1^2} = \frac{U}{Q}\,.$$

The same equation is fulfilled for $\phi_2 = \phi_1 + \pi$. A polarizer oriented in the directions $\phi_{1,2}$ will detect a maximal signal, while when oriented at 90° to this polarization direction the signal is minimal.

It is not very convenient to work with these basis dependent amplitudes. First of all, the results will depend on the arbitrary choice of $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\epsilon}^{(2)}$. Therefore, we shall not work directly with the Stokes parameters Q and U. But we make use of the spin weighted spherical harmonic functions ${}_sY_{\ell m}(\mathbf{n})$ which are defined for each integer s with $|s| \leq \ell$ and have the property that they transform under rotations about $\mathbf{n}$ by an angle ψ like ${}_sY_{\ell m}(\mathbf{n}) \to e^{is\psi}\,{}_sY_{\ell m}(\mathbf{n})$. The spin weighted spherical harmonics are the components of a symmetric rank $|s|$ tensor field defined on the tangent space of the sphere in the canonical basis $(\mathbf{e}_\vartheta \equiv \partial_\vartheta,\ \mathbf{e}_\varphi \equiv (1/\sin\vartheta)\partial_\varphi)$. Note that $(\mathbf{e}_\vartheta,\ \mathbf{e}_\varphi)$ are not well defined at the north and south poles. Setting

$$\mathbf{e}^{\pm} = \frac{1}{\sqrt{2}}\left(\mathbf{e}_\vartheta \pm i\mathbf{e}_\varphi\right)\,,$$

${}_sY_{\ell m}(\mathbf{n})$ transforms like the $+\cdots+$ component of a rank s tensor, if $s>0$ and like the $-\cdots-$ component of a rank $|s|$ tensor, if $s<0$. With respect to the helicity basis $\mathbf{e}^{(\pm)}$, the dimensionless parameters $\mathcal{Q}\pm i\mathcal{U}$ can be expanded as

$$(\mathcal{Q}\pm i\mathcal{U})(\mathbf{n}) = \sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m}^{(\pm 2)}\, {}_{\pm 2}Y_{\ell m}(\mathbf{n})\,, \tag{5.10}$$

$$= \sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell} (e_{\ell m}\pm i b_{\ell m})\, {}_{\pm 2}Y_{\ell m}(\mathbf{n})\,. \tag{5.11}$$

Hence

$$e_{\ell m} = \frac{1}{2}\left(a_{\ell m}^{(2)} + a_{\ell m}^{(-2)}\right)\,, \quad b_{\ell m} = \frac{-i}{2}\left(a_{\ell m}^{(2)} - a_{\ell m}^{(-2)}\right)\,. \tag{5.12}$$

Under a 'parity' transformation, $\mathbf{n}\to-\mathbf{n}$ the basis vectors $\mathbf{e}^{(\pm)}$ transform as $\mathbf{e}^{(\pm)}\to\mathbf{e}^{(\mp)}$. Hence the coefficient $a_{\ell m}^{(2)}$ turns into $a_{\ell m}^{(-2)}$ and $a_{\ell m}^{(-2)}\to a_{\ell m}^{(2)}$ so that $e_{\ell m}$ remains invariant while $b_{\ell m}$ changes sign.

The spin weighted spherical harmonics are defined in Appendix A4.2.4 where useful properties are derived. The sum over ℓ starts only at $\ell=2$. As is clear from their definition, the spin weighted spherical harmonics ${}_sY_{\ell m}$ vanish for $|s|>\ell$. Here $a_{\ell m}^{(\pm 2)}$ is a decomposition into positive and negative helicity, while $(e_{\ell m}, b_{\ell m})$ is a decomposition into the $\mathcal{Q}$ and $\mathcal{U}$ Stokes parameter with respect to the canonical basis on the sphere which requires the choice of a z axis.

In Appendix A4.2.4 we also define the spin raising and lowering operators $\eth$ and $\eth^*$, similar to the quantum mechanical angular momentum operators L_+ and L_- which raise and lower the magnetic quantum number m. The operators $\eth^{(*)}$ have the properties $\eth\, {}_sY_{\ell m}\propto {}_{s+1}Y_{\ell m}$ and $\eth^*\, {}_sY_{\ell m}\propto {}_{s-1}Y_{\ell m}$. Actually one obtains (see Appendix A4.2.4)

$$\eth^2\,({}_{-2}Y_{\ell m}) = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,Y_{\ell m}\,, \tag{5.13}$$

$$(\eth^*)^2\,({}_2Y_{\ell m}) = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,Y_{\ell m}\,. \tag{5.14}$$

Applying this to $\mathcal{Q}\pm i\mathcal{U}$ we find

$$(\eth^*)^2(\mathcal{Q}+i\mathcal{U})(\mathbf{n}) = \sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m}^{(2)}\sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,Y_{\ell m}(\mathbf{n})\,, \tag{5.15}$$

$$\eth^2(\mathcal{Q}-i\mathcal{U})(\mathbf{n}) = \sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m}^{(-2)}\sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,Y_{\ell m}(\mathbf{n})\,. \tag{5.16}$$

With this, we can define the scalar quantities

$$\mathcal{E}(\mathbf{n}) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} e_{\ell m} Y_{\ell m}(\mathbf{n}) \,, \tag{5.17}$$

$$\mathcal{B}(\mathbf{n}) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} b_{\ell m} Y_{\ell m}(\mathbf{n}) \,. \tag{5.18}$$

Like temperature fluctuations, $\mathcal{E}$ and $\mathcal{B}$ are invariant under rotation. Since the sign of $b_{\ell m}$ changes under parity, $\mathcal{B}$ has negative parity while $\mathcal{E}$ and $\mathcal{M}$ have positive parity. At the end of Section 5.3 we shall show that $\mathcal{E}$ measures gradient contributions while $\mathcal{B}$ measures curl contributions to the electric field considered as a function on the sphere. The electric field is transverse and hence tangential to the sphere of photon directions.

5.1.2 The scattering matrix and collision term

We now consider incoming radiation from direction $\mathbf{n}'$ which is then scattered into direction $\mathbf{n}$ with scattering angle β, $\mathbf{n} \cdot \mathbf{n}' = \cos\beta$. The cross section for scattering off a non-relativistic electron depends on the polarization of the photon. For photons polarized in the scattering plane it is suppressed by a factor $\cos^2\beta$, while it is unsuppressed for photons polarized normal to the scattering plane. The scattering field generated per unit of time in a plasma with electron density n_e is proportional to $\sqrt{n_e \sigma_T}\,\mathbf{E}$, where σ_T is the scattering cross section. In the rest frame of the electron we thus have (Jackson, 1975),

$$E_{\parallel}^{(c)} = \frac{\sqrt{n_e} e^2}{m_e} \cos\beta E_{\parallel} = \sqrt{\frac{3}{8\pi} n_e \sigma_T} \cos\beta E_{\parallel} \,, \tag{5.19}$$

$$E_{\perp}^{(c)} = \frac{\sqrt{n_e} e^2}{m_e} E_{\perp} = \sqrt{\frac{3}{8\pi} n_e \sigma_T} E_{\perp} \,. \tag{5.20}$$

We now choose the polarization basis such that $\boldsymbol{\epsilon}^{(1)}(\mathbf{n})$ lies in the scattering plane and $\boldsymbol{\epsilon}^{(2)}(\mathbf{n})$ is normal to it. Using $I = |E_{\parallel}|^2 + |E_{\perp}|^2$, $Q = |E_{\parallel}|^2 - |E_{\perp}^2|$ and $U = 2E_{\parallel}E_{\perp}^*$ we obtain

$$\mathcal{M}^{(c)} = \frac{3}{16\pi} n_e \sigma_T \left[(1 + \cos^2\beta)\mathcal{M} - \sin^2\beta \mathcal{Q} \right] \,, \tag{5.21}$$

$$\mathcal{Q}^{(c)} = \frac{3}{16\pi} n_e \sigma_T \left[(1 + \cos^2\beta)\mathcal{Q} - \sin^2\beta \mathcal{M} \right] \,, \tag{5.22}$$

$$\mathcal{U}^{(c)} = \frac{3}{8\pi} n_e \sigma_T \cos\beta \mathcal{U} \,. \tag{5.23}$$

Defining the vector

$$\mathcal{V} = \begin{pmatrix} \mathcal{M} \\ \mathcal{Q} + i\mathcal{U} \\ \mathcal{Q} - i\mathcal{U} \end{pmatrix} , \tag{5.24}$$

we can write the scattered amplitudes in terms of a scattering matrix, $\mathcal{V}^{(c)} = (n_e \sigma_T / 4\pi) S \mathcal{V}$ with

$$S = \frac{3}{4} \begin{pmatrix} \cos^2\beta + 1 & -\frac{1}{2}\sin^2\beta & -\frac{1}{2}\sin^2\beta \\ -\sin^2\beta & \frac{1}{2}(\cos\beta + 1)^2 & \frac{1}{2}(\cos\beta - 1)^2 \\ -\sin^2\beta & \frac{1}{2}(\cos\beta - 1)^2 & \frac{1}{2}(\cos\beta + 1)^2 \end{pmatrix} . \tag{5.25}$$

This is the scattering matrix expressed in the polarization basis $\left(\boldsymbol{\epsilon}^{(1)}(\mathbf{n}),\ \boldsymbol{\epsilon}^{(2)}(\mathbf{n})\right)$ which is chosen such that $\boldsymbol{\epsilon}^{(1)}(\mathbf{n})$ lies in the scattering plane. In the expansion (5.10), we express $\mathcal{Q} \pm i\mathcal{U}$ in the basis $(\mathbf{e}_\vartheta, \mathbf{e}_\varphi)$. To obtain the scattering matrix with respect to this basis, we first rotate $\mathcal{Q} \pm i\mathcal{U}$ by an angle γ' around $\mathbf{n}'$ to turn the basis $(\mathbf{e}_\vartheta(\mathbf{n}'), \mathbf{e}_\varphi(\mathbf{n}'))$ into $\left(\boldsymbol{\epsilon}^{(1)}(\mathbf{n}'),\ \boldsymbol{\epsilon}^{(2)}(\mathbf{n}')\right)$, only then can we apply the scattering matrix S on $\mathcal{V}$. Finally, we rotate the polarizations $\left(\boldsymbol{\epsilon}^{(1)}(\mathbf{n}),\ \boldsymbol{\epsilon}^{(2)}(\mathbf{n})\right)$ back into $(\mathbf{e}_\vartheta(\mathbf{n}), \mathbf{e}_\varphi(\mathbf{n}))$ by the rotation with angle $-\gamma$ around $\mathbf{n}$.

The rotation with angle γ' around direction $\mathbf{n}'$ multiplies $\mathcal{Q}(\mathbf{n}') \pm i\mathcal{U}(\mathbf{n}')$ by a factor $\exp(\pm 2i\gamma')$ and the rotation around $\mathbf{n}$ with angle $-\gamma$ multiplies $\mathcal{Q}^{(c)}(\mathbf{n}) \pm i\mathcal{U}^{(c)}(\mathbf{n})$ by $\exp(\mp 2i\gamma)$. The intensity perturbation is invariant under rotations. The scattering matrix which multiplies $\mathcal{V}$ with Stokes parameters oriented in the fixed polarization basis $(\mathbf{e}_\vartheta,\ \mathbf{e}_\varphi)$, is therefore simply $R(-\gamma) S R(\gamma')$ where we define the 3×3 matrix $R(\alpha) = \mathrm{diag}\left(1, e^{2i\alpha}, e^{-2i\alpha}\right)$.

Using the expressions for ${}_{\pm s}Y_{\ell m}(\vartheta, \varphi)$, $\ell \leq 2$ given in Appendix A4.2.4, straightforward comparison gives

$$\begin{aligned} &R(-\gamma) S R(\gamma') \\ &\quad = \frac{1}{2}\sqrt{\frac{4\pi}{5}} \begin{pmatrix} Y_{20}(\beta,\gamma') + 2\sqrt{5}Y_{00}(\beta,\gamma') & -\sqrt{\frac{3}{2}}Y_{2-2}(\beta,\gamma') & -\sqrt{6}Y_{22}(\beta,\gamma') \\ -\sqrt{6}\,{}_2Y_{20}(\beta,\gamma')e^{-2i\gamma} & 3\,{}_2Y_{2-2}(\beta,\gamma')e^{-2i\gamma} & 3\,{}_2Y_{22}(\beta,\gamma')e^{-2i\gamma} \\ -\sqrt{\frac{3}{2}}\,{}_{-2}Y_{20}(\beta,\gamma')e^{2i\gamma} & 3\,{}_{-2}Y_{2-2}(\beta,\gamma')e^{2i\gamma} & 3\,{}_{-2}Y_{22}(\beta,\gamma')e^{2i\gamma} \end{pmatrix} . \end{aligned} \tag{5.26}$$

Using the addition theorem for spin weighted spherical harmonics,

$${}_sY_{2s'}(\beta,\gamma')e^{-si\gamma} = \sqrt{\frac{4\pi}{5}} \sum_m {}_{-s'}Y^*_{2m}(\mathbf{n}')\,{}_sY_{2m}(\mathbf{n}) ,$$

we can write the matrix $R(-\gamma) S R(\gamma') = \frac{4\pi}{10} P(\mathbf{n}, \mathbf{n}') + \mathrm{diag}(1, 0, 0)$, where the matrix $P(\mathbf{n}, \mathbf{n}')$ is given by $({}_0Y_{\ell m} \equiv Y_{\ell m})$

$$P(\mathbf{n}, \mathbf{n}') = \sum_{m=-2}^{2} P_m(\mathbf{n}, \mathbf{n}') ,$$

where

$$P_m(\mathbf{n},\mathbf{n}') = \begin{pmatrix} Y_{2m}(\mathbf{n})Y^*_{2m}(\mathbf{n}') & -\sqrt{\frac{3}{2}}Y_{2m}(\mathbf{n})\,{}_2Y^*_{2m}(\mathbf{n}') & -\sqrt{\frac{3}{2}}Y_{2m}(\mathbf{n})\,{}_{-2}Y^*_{2m}(\mathbf{n}') \\ -\sqrt{6}\,{}_2Y_{2m}(\mathbf{n})Y^*_{2m}(\mathbf{n}') & 3\,{}_2Y_{2m}(\mathbf{n})\,{}_2Y^*_{2m}(\mathbf{n}') & 3\,{}_2Y_{2m}(\mathbf{n})\,{}_{-2}Y^*_{2m}(\mathbf{n}') \\ -\sqrt{6}\,{}_{-2}Y_{2m}(\mathbf{n})Y^*_{2m}(\mathbf{n}') & 3\,{}_{-2}Y_{2m}(\mathbf{n})\,{}_2Y^*_{2m}(\mathbf{n}') & 3\,{}_{-2}Y_{2m}(\mathbf{n})\,{}_{-2}Y^*_{2m}(\mathbf{n}') \end{pmatrix}. \quad (5.27)$$

The three component collision term for $\mathcal{V}$ in the electron rest frame is now obtained by integrating over the incoming photon directions and subtracting the photons scattered out of the beam, as in Eq. (4.111),

$$C[\mathcal{V}]_{\rm rest} = an_e\sigma_T\left[\frac{1}{10}\int \Omega_{\mathbf{n}'}\sum_{m=-2}^{2} P_m(\mathbf{n},\mathbf{n}')\mathcal{V}(\mathbf{n}') - \mathcal{V}(\mathbf{n}) + \frac{1}{4\pi}\int \Omega_{\mathbf{n}'}\mathcal{M}(\mathbf{n}')\begin{pmatrix}1\\0\\0\end{pmatrix}\right]. \quad (5.28)$$

The Y_{00} term in Eq. (5.26) results in the second integral in Eq. (5.28) which provokes isotropization in the electron rest frame. The other terms of $R(-\gamma)SR(\gamma')$ lead to $\sum_{m=-2}^{2} P_m(\mathbf{n},\mathbf{n}')$.

As we shall see, the contribution to the scattering term coming from the spin weighted spherical harmonics with $|m| = 0,\ 1$ and 2 correspond to the contributions for scalar, vector and tensor perturbations respectively. To transform the scattering term from the electron (or baryon) rest frame to our coordinate frame, we simply add the Doppler term $\mathbf{n}\cdot\mathbf{V}^{(b)}$ as in Eq. (4.112). Also as there, we obtain an additional factor a since we calculate the scattering per conformal time interval. The collision term per unit of conformal time in the coordinate frame then becomes

$$C[\mathcal{V}] = an_e\sigma_T\left[\frac{1}{10}\int \Omega_{\mathbf{n}'}\sum_{m=-2}^{2} P_m(\mathbf{n},\mathbf{n}')\mathcal{V}(\mathbf{n}') - \mathcal{V}(\mathbf{n}) + \left[\frac{1}{4\pi}\int \Omega_{\mathbf{n}'}\mathcal{M}(\mathbf{n}') + \mathbf{n}\cdot\mathbf{V}^{(b)}\right]\begin{pmatrix}1\\0\\0\end{pmatrix}\right]. \quad (5.29)$$

5.2 Total angular momentum decomposition

In the previous section we calculated the scattering term of the vector $\mathcal{V}$ at some fixed position $\mathbf{x}$ as a function of the photon direction $\mathbf{n}$. Now we also want to consider the $\mathbf{x}$ dependence.

In Chapter 4, we Fourier transformed the $\mathbf{x}$ dependence of the temperature fluctuation $\mathcal{M}$, and then decomposed $\mathcal{M}(t,\mathbf{k},\mathbf{n})$ into its scalar, vector and tensor contributions. We found that $\mathcal{M}^{(S)}$ depends on $\mathbf{n}$ only via $\mu = \hat{\mathbf{k}}\cdot\mathbf{n}$, while $\mathcal{M}^{(V)}$ and

$\mathcal{M}^{(T)}$ are of the form

$$\mathcal{M}^{(V)} = \sqrt{1-\mu^2}\frac{1}{2}\left[\exp(i\phi)\mathcal{M}_+^{(V)}(\mu) + \exp(-i\phi)\mathcal{M}_-^{(V)}(\mu)\right], \tag{5.30}$$

$$\mathcal{M}^{(T)} = (1-\mu^2)\frac{1}{2}\left[\exp(i2\phi)\mathcal{M}_+^{(T)}(\mu) + \exp(-2i\phi)\mathcal{M}_-^{(T)}(\mu)\right]. \tag{5.31}$$

Here ϕ is the angle with respect to some fixed (but arbitrary) direction in the plane normal to $\mathbf{k}$. We then expanded the functions $\mathcal{M}_\pm^{(V,T)}$ in Legendre polynomials. But, according to Appendix A4.2,

$$Y_{\ell,\pm1}(\mathbf{n}) \propto \pm e^{\pm i\phi}\sqrt{1-\mu^2}P'_\ell(\mu),$$
$$Y_{\ell,\pm2}(\mathbf{n}) \propto \pm e^{\pm 2i\phi}(1-\mu^2)P''_\ell(\mu).$$

For a fixed wave vector $\mathbf{k}$, we can therefore expand the $\mathbf{n}$ dependence of the vector contribution to $\mathcal{M}$ in terms of spherical harmonics of order $|m| = 1$ and the tensor contributions in terms of spherical harmonics of order $|m| = 2$. These are the spherical harmonics of the photon direction $\mathbf{n}$ in the coordinate system with $\mathbf{k} \parallel \mathbf{e}_z$.

For a fixed Fourier mode $\mathbf{k}$ we now introduce the basis functions

$${}_sG_{\ell m}(\mathbf{x},\mathbf{n}) = (-i)^\ell\sqrt{\frac{4\pi}{2\ell+1}}e^{i\mathbf{k}\cdot\mathbf{x}}\,{}_sY_{\ell m}(\mathbf{n}), \tag{5.32}$$

where the spin weighted spherical harmonics are evaluated in a coordinate system with $\mathbf{k} \parallel \mathbf{e}_z$. According to our findings in Chapter 4, the temperature fluctuation can now be expanded as

$$\mathcal{M}(t,\mathbf{x},\mathbf{n}) = \int\frac{d^3k}{(2\pi)^3}\sum_{\ell=0}^{\infty}\sum_{m=-2}^{2}\mathcal{M}_\ell^{(m)}(t,\mathbf{k})\,{}_0G_{\ell m}(\mathbf{x},\mathbf{n}). \tag{5.33}$$

As we have seen, the $m = 0$ term represents scalar fluctuations while the $|m| = 1$ terms are of vector-type and the $|m| = 2$ terms are tensor fluctuations. The coefficients $\mathcal{M}_\ell^{(\pm2)}$ are easily related to the expansion coefficients $\mathcal{M}_\ell^{(T\pm)}$ defined in Eq. (4.147), and $\mathcal{M}_\ell^{(\pm1)}$ are related to $\mathcal{M}_\ell^{(V\pm)}$ given in Eqs. (4.131) and (4.132) (see Ex. 5.1).

Next, we use that the polarization can be expanded in terms of spin weighted spherical harmonics ${}_{\pm2}Y_{\ell m}$ (see Eqs. (5.10) and (5.11)):

$$\mathcal{Q} \pm i\mathcal{U} = \int\frac{d^3k}{(2\pi)^3}\sum_{\ell=2}^{\infty}\sum_{m=-2}^{2}{}_{\pm2}\mathcal{A}_\ell^{(m)}(t,\mathbf{k})\,{}_{\pm2}G_{\ell m}(\mathbf{x},\mathbf{n}), \tag{5.34}$$

$$= \int\frac{d^3k}{(2\pi)^3}\sum_{\ell=2}^{\infty}\sum_{m=-2}^{2}\left(\mathcal{E}_\ell^{(m)}(t,\mathbf{k}) \pm i\mathcal{B}_\ell^{(m)}(t,\mathbf{k})\right){}_{\pm2}G_{\ell m}(\mathbf{x},\mathbf{n}). \tag{5.35}$$

Here, as in Eqs. (5.10) and (5.11), the coefficients $\mathcal{A}$ are related to $\mathcal{E}$ and $\mathcal{B}$ by

$${}_{\pm2}\mathcal{A}_\ell^{(m)}(t,\mathbf{k}) = \mathcal{E}_\ell^{(m)}(t,\mathbf{k}) \pm i\mathcal{B}_\ell^{(m)}(t,\mathbf{k}).$$

As in the case of temperature fluctuations, $m = 0$ are scalar perturbations, while $|m| = 1$ and $|m| = 2$ are vector and tensor perturbations respectively. The above $\mathcal{Q}$ and $\mathcal{U}$ polarization are defined with respect to some fixed coordinate system in real space, while the Fourier coefficients $\mathcal{E}_\ell$ and $\mathcal{B}_\ell$ correspond to the $\mathcal{Q}$ and $\mathcal{U}$ polarization with respect to the coordinate system where $\mathbf{k}$ points in the z-direction. Therefore, the inverse Fourier transform of $\mathcal{E}$ and $\mathcal{B}$ respectively, will in general not simply give $\mathcal{Q}$ and $\mathcal{U}$ respectively with respect to any fixed real space coordinate system. The basis functions ${}_sG_{\ell m}$ have three different types of indices. Let us briefly recapitulate their meaning. As we have seen, m determines the tensor character of the perturbations. The index ℓ labels the expansion in an orthonormal set of functions of $\mu = \hat{\mathbf{k}} \cdot \mathbf{n} = \cos\vartheta$. Under rotations around the photon direction $\mathbf{n}$ temperature fluctuations are tensorial quantities of rank $s = 0$ which gives them the index 0, while the polarization variables, $\mathcal{Q} \pm i\mathcal{U}$ are tensorial quantities of rank $|s| = 2$ with helicity ± 2.

It is important to note that when expanding in ${}_sG_{\ell m}$ we express the spherical harmonics $Y_{\ell m}$ with respect to a coordinate system which depends on $\mathbf{k}$.

We now consider the situation where the observer is placed at $\mathbf{x} = 0$ and the incoming photon is at a distance r from her, so that the photon position is $\mathbf{x} = -r\mathbf{n}$, where $\mathbf{n}$, as above, denotes the direction of propagation of the photon. This situation will be relevant for the line-of-sight integration which we shall use to solve the Boltzmann equation. We want to expand our basis functions ${}_sG_{\ell m}(-r\mathbf{n}, \mathbf{n})$ for fixed $\mathbf{k}$ in their total angular momentum components. The functions ${}_sG_{\ell m}$ have 'spin' ℓ but the 'orbital' angular momentum of the exponential is a sum,

$$e^{i\mathbf{k}\cdot\mathbf{x}} = e^{-ikr\mu} = \sum_{L=0}^{\infty} \sqrt{4\pi(2L+1)}(-i)^{-L} j_L(kr) Y_{L0}(\mathbf{n}) \,,$$

where we have used Eq. (A4.101) and $P_\ell(\mu) = \sqrt{4\pi/(2\ell+1)}Y_{\ell 0}(\mathbf{n})$. Hence

$${}_sG_{\ell m}(-r\mathbf{n}, \mathbf{n}) = 4\pi \sum_{L=0}^{\infty} \sqrt{\frac{2L+1}{2\ell+1}} i^{-L-\ell} j_L(kr) Y_{L0}(\mathbf{n})\, {}_sY_{\ell m}(\mathbf{n}) \,. \tag{5.36}$$

The spin weighted spherical harmonics are related to the matrix elements of the representations of the rotation group by (see Appendix A4.2)

$${}_sY_{\ell m}(\theta, \phi) = \sqrt{\frac{2\ell+1}{4\pi}} D^{(\ell)}_{-sm}(\phi, \theta, 0) \,. \tag{5.37}$$

Here $\mathbf{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ and $(\phi, \theta, 0)$ denote the Euler angles of the rotation which first rotates around the y axis with angle θ and then around the z axis with angle ϕ. This is a rotation which turns the z axis into $\mathbf{n}$. We also want to use the relation of Y_{L0} to $D^{(L)}_{00}$. But $D^{(L)}_{SM} D^{(\ell)}_{sm}$ is the matrix element $(S, M; s, m)$ of the

representation $D^{(L)} \otimes D^{(\ell)}$ in the basis $Y_{LM} \otimes Y_{\ell m}$. With the help of the Clebsch–Gordan series (see Appendix A4.2) this representation can be decomposed as a sum of irreducible representations,

$$D^{(L)} \otimes D^{(\ell)} = \sum_{j=|L-\ell|}^{L+\ell} D^{(j)} \,.$$

The basis $(Y_{LM} \otimes Y_{\ell m})_{M=-L,m=-\ell}^{M=L,m=\ell}$ is transformed into the basis $((Y_{jr})_{r=-j}^{j})_{j=|L-\ell|}^{j=L+\ell}$ with the Clebsch–Gordan coefficients $\langle L, \ell; M, m | j, r\rangle$. Using the fact that Clebsch–Gordan coefficients are non-vanishing only if $r = M + m$, we can write the matrix elements

$$D^{(L)}_{00} D^{(\ell)}_{-sm} = \sum_{j=|L-\ell|}^{j=L+\ell} \langle L, \ell; 0, m | j, m\rangle \langle L, \ell; 0, -s | j, -s\rangle D^{(j)}_{-sm} \,. \tag{5.38}$$

(For more details see Appendix A4.2.) Using Eq. (5.37) this yields

$$\begin{aligned} &4\pi \sqrt{\frac{2L+1}{2\ell+1}} Y_{L0}(\mathbf{n}) {}_sY_{\ell m}(\mathbf{n}) \\ &\quad = (2L+1) \sum_{j=|L-\ell|}^{j=L+\ell} \langle L, \ell; 0, m | j, m\rangle \langle L, \ell; 0, -s | j, -s\rangle \sqrt{\frac{4\pi}{(2j+1)}} \, {}_sY_{jm}(\mathbf{n}) \,. \end{aligned} \tag{5.39}$$

When inserting this in the sum, Eq. (5.36) we can exchange the sums over L and j. Extending the sum over j from zero to infinity, we have to sum for each given j over all Ls for which this j contributes in Eq. (5.39). These are simply the values $|j - \ell| \le L \le j + \ell$. Defining the functions ${}_sf_j^{(\ell m)}$ which represent the sums over L by

$${}_sf_j^{(\ell m)}(x) \equiv \sum_{L=|j-\ell|}^{j+\ell} (-i)^{L+\ell-j} \frac{2L+1}{2j+1} \langle L, \ell; 0, m | j, m\rangle \langle L, \ell; 0, -s | j, -s\rangle \, j_L(x) \,, \tag{5.40}$$

we can then write the sum (5.36) as

$${}_sG_{\ell m}(-r\mathbf{n}, \mathbf{n}) = \sum_{j=0}^{\infty} (-i)^j \sqrt{4\pi(2j+1)} \, {}_sf_j^{(\ell m)}(kr) \, {}_sY_{jm}(\mathbf{n}) \,. \tag{5.41}$$

We are only really interested in the cases $s = 0$ and $s = \pm 2$. For these we define

$$\alpha_j^{(\ell m)} \equiv {}_0f_j^{(\ell m)} \,, \tag{5.42}$$

$$\epsilon_j^{(\ell m)} \pm i\beta_j^{(\ell m)} \equiv {}_{\pm 2}f_j^{(\ell m)} \,. \tag{5.43}$$

We repeat Eq. (5.41) for the relevant cases $s = 0$ and $|s| = 2$:

$$ {}_0G_{\ell m}(-r\mathbf{n}, \mathbf{n}) = \sum_{j=0}^{\infty} \sqrt{4\pi(2j+1)}\,(-i)^j \alpha_j^{(\ell m)}(kr) Y_{jm}(\mathbf{n})\,, \tag{5.44} $$

$$ {}_{\pm 2}G_{\ell m}(-r\mathbf{n}, \mathbf{n}) = \sum_{j=0}^{\infty} \sqrt{4\pi(2j+1)}\,(-i)^j \left(\epsilon_j^{(\ell m)}(kr) \pm i\beta_j^{(\ell m)}(kr)\right) {}_{\pm 2}Y_{jm}(\mathbf{n})\,. \tag{5.45} $$

This is the total angular momentum expansion of ${}_sG_{\ell m}(-r\mathbf{n}, \mathbf{n})$. We want to use it to find the integral solution of the Boltzmann equation. For this we shall need the functions $\alpha_j^{(\ell m)}, \epsilon_j^{(\ell m)}$ and $\beta_j^{(\ell m)}$ only for ℓ and $|m| \leq 2$, since the 'source terms' of the Boltzmann equation, which are the collision terms collected in $C[\mathcal{V}]$ in Eq. (5.29) and the gravitational contributions which we have determined in Chapter 4 all have $\ell \leq 2$ and $|m| \leq 2$.

Using the Clebsch–Gordan coefficients given in Appendix A4.2 and the recurrence relations of spherical Bessel functions presented in Appendix A4.3 one obtains

$$ \alpha_\ell^{(00)}(x) = j_\ell(x)\,, \tag{5.46} $$

$$ \alpha_\ell^{(10)}(x) = j_\ell'(x)\,, \quad \alpha_\ell^{(1\,\pm 1)}(x) = \sqrt{\frac{\ell(\ell+1)}{2}}\,\frac{j_\ell(x)}{x}\,, \tag{5.47} $$

$$ \alpha_\ell^{(20)}(x) = \frac{1}{2}[3j_\ell''(x) + j_\ell(x)], \quad \alpha_\ell^{(2\,\pm 1)}(x) = \sqrt{\frac{3\ell(\ell+1)}{2}}\left(\frac{j_\ell(x)}{x}\right)', \tag{5.48} $$

$$ \alpha_\ell^{(2\,\pm 2)}(x) = \sqrt{\frac{3(\ell+2)!}{8(\ell-2)!}}\,\frac{j_\ell(x)}{x^2}\,, \tag{5.49} $$

$$ \epsilon_\ell^{(20)}(x) = \sqrt{\frac{3(\ell+2)!}{8(\ell-2)!}}\,\frac{j_\ell(x)}{x^2} \equiv \alpha_\ell^{(2\,\pm 2)}(x)\,, \tag{5.50} $$

$$ \epsilon_\ell^{(2\,\pm 1)}(x) = \frac{1}{2}\sqrt{(\ell-1)(\ell+2)}\left[\frac{j_\ell(x)}{x^2} + \frac{j_\ell'(x)}{x}\right], \tag{5.51} $$

$$ \epsilon_\ell^{(2\,\pm 2)}(x) = \frac{1}{4}\left[-j_\ell(x) + j_\ell''(x) + 2\frac{j_\ell(x)}{x^2} + 4\frac{j_\ell'(x)}{x}\right], \tag{5.52} $$

$$ \beta_\ell^{(20)}(x) = 0\,, \tag{5.53} $$

$$ \beta_\ell^{(2\,\pm 1)}(x) = \pm\frac{1}{2}\sqrt{(\ell-1)(\ell+2)}\,\frac{j_\ell(x)}{x}\,, \tag{5.54} $$

$$ \beta_\ell^{(2\,\pm 2)}(x) = \pm\frac{1}{2}\left[j_\ell'(x) + 2\frac{j_\ell(x)}{x}\right]. \tag{5.55} $$

In order to (hopefully) avoid confusion we have used the letter ℓ here as the total angular momentum, since j is the name of the spherical Bessel functions.

The functions $\alpha_j^{(\ell m)}$, $\epsilon_j^{(\ell m)}$ and $\beta_j^{(\ell m)}$ will be investigated in more detail when we discuss the integral solution of the Boltzmann equation. They peak around $x \simeq \ell$, like spherical Bessel functions, and then oscillate and decay like $1/x$ or faster.

From the definition of ${}_sY_{\ell m}$ it follows that under the parity operation, $\mathbf{n} \to -\mathbf{n}$, $\mathbf{e}_\vartheta(\mathbf{n}) \to \mathbf{e}_\vartheta(-\mathbf{n}) = \mathbf{e}_\vartheta(\mathbf{n})$, $\mathbf{e}_\varphi(\mathbf{n}) \to \mathbf{e}_\varphi(-\mathbf{n}) = -\mathbf{e}_\varphi(\mathbf{n})$ one finds ${}_sY_{\ell m}(-\mathbf{n}) = (-1)^\ell {}_{-s}Y_{\ell m}(\mathbf{n})$. The first factor simply reflects the behaviour of $Y_{\ell m}$ under parity, while the transformation $s \to -s$ comes from the fact that $\mathbf{e}_\varphi$ changes sign under parity, while $\mathbf{e}_\vartheta$ does not. This together with the parity of the spherical Bessel functions, $j_\ell(-x) = (-1)^\ell j_\ell(x)$ explains that $\alpha_j^{(\ell\ -m)}(x) = \alpha_j^{(\ell m)}(x)$ and $\epsilon_j^{(\ell\ -m)}(x) = \epsilon_j^{(\ell m)}(x)$ while $\beta_j^{(\ell\ -m)}(x) = -\beta_j^{(\ell m)}(x)$. Furthermore, since $\mathcal{E}_\ell^{(m)}$ couples to the sum ${}_sY_{\ell m} + {}_{-s}Y_{\ell m}$ it has parity $(-1)^\ell$, while $\mathcal{B}_\ell^{(m)}$ which couples to the difference ${}_sY_{\ell m} - {}_{-s}Y_{\ell m}$ has parity $(-1)^{\ell+1}$. With $Y_{\ell m}$, the $\mathcal{M}_\ell^{(m)}$ have parity $(-1)^\ell$.

Recalling the definition (5.11), we observe that in the coordinate system where $\mathbf{k}$ points in the z direction, the $\mathcal{E}_\ell^{(m)}(\mathbf{k})$-terms correspond to pure Q and the $\mathcal{B}_\ell^{(m)}(\mathbf{k})$-terms correspond to pure U polarization.

5.3 The spectra

To find the power spectra in terms of the random variables $\mathcal{M}_\ell^{(m)}$, $\mathcal{E}_\ell^{(m)}$ and $\mathcal{B}_\ell^{(m)}$ in Fourier space, we use the definition of the temperature perturbation spectrum given in Chapter 2,

$$C_\ell^{(\mathcal{M})} = \langle |a_{\ell m}|^2 \rangle \,, \qquad \text{where} \tag{5.56}$$

$$\mathcal{M}(\mathbf{x}, \mathbf{n}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}(x) Y_{\ell m}(\mathbf{n}) \,. \tag{5.57}$$

From this and the addition theorem of spherical harmonics,

$$Y_{\ell 0}(\cos\vartheta = \mathbf{n}\cdot\mathbf{n}') = \sqrt{\frac{4\pi}{2\ell+1}} \sum_m Y_{\ell 0}(\mathbf{n}) Y^*_{\ell m}(\mathbf{n}') \,, \tag{5.58}$$

we have derived the expression for the correlation function,

$$\langle \mathcal{M}(t, \mathbf{x}, \mathbf{n}) \mathcal{M}(t, \mathbf{x}, \mathbf{n}') \rangle = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell+1) P_\ell(\mathbf{n}\cdot\mathbf{n}') C_\ell^{(\mathcal{M})} \,, \tag{5.59}$$

where we have used $P_\ell(\mathbf{n}\cdot\mathbf{n}') = \sqrt{4\pi/(2\ell+1)} Y_{\ell 0}(\cos\vartheta = \mathbf{n}\cdot\mathbf{n}')$. In the same

way we now define the rotationally invariant spectra

$$C_\ell^{(\mathcal{E})} = \langle |e_{\ell m}|^2 \rangle \,, \tag{5.60}$$

$$C_\ell^{(\mathcal{B})} = \langle |b_{\ell m}|^2 \rangle \,, \tag{5.61}$$

$$C_\ell^{(\mathcal{ME})} = \langle a^*_{\ell m} e_{\ell m} \rangle \,, \tag{5.62}$$

with the expansion coefficients $e_{\ell m}$ and $b_{\ell m}$ defined in Eq. (5.12). The coefficients $b_{\ell m}$ have parity $(-1)^{\ell+1}$ while $a_{\ell m}$ and $e_{\ell m}$ have parity $(-1)^\ell$. We shall always assume that the random process which generates the initial fluctuations is invariant under parity, so that expectation values with negative parity such as $C_\ell^{(\mathcal{MB})}$ and $C_\ell^{(\mathcal{EB})}$ vanish. But, in principle, this has to be tested experimentally. It is possible that parity violating processes, such as weak interactions lead to effects in the CMB, see Caprini *et al.* (2004).

We now want to relate the spectra to the **k**-space expressions for the variables $\mathcal{M}$, $\mathcal{E}$ and $\mathcal{B}$. To do this we have to be careful about our use of spherical harmonics. In Eq. (5.57) we employ them with respect to some arbitrary but fixed z direction, let us call it **e**, while in the Fourier decomposition, Eq. (5.33), the spherical harmonics are to be taken in the coordinate system where $\hat{\mathbf{k}}$ denotes the z direction. To make this dependence clear, in this section we indicate the spherical harmonics with respect to a given z axis, **e** by $Y_{\ell m}(\mathbf{n};\mathbf{e})$. To relate $Y_{\ell m}(\mathbf{n};\hat{\mathbf{k}})$ to $Y_{\ell m}(\mathbf{n};\mathbf{e})$ we use the fact that a basis with $\hat{\mathbf{k}}$ in the z direction can be obtained from a basis with **e** in the z direction by first rotating with the angle $-\phi_{\mathbf{k}}$ around the z axis, **e** and then with $-\theta_k$ around the y axis. Here, $(\theta_{\mathbf{k}}, \phi_{\mathbf{k}})$ are the polar angles of **k** in the coordinate system with **e** in the z direction. We therefore rotate the basis with the rotation given by the Euler angles $(0, -\theta_{\mathbf{k}}, -\phi_{\mathbf{k}})$. This is the inverse of the rotation with Euler angles $(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0)$. Since the representation matrices are unitary,

$$D^{(\ell)}_{mm'}(0, -\theta_{\mathbf{k}}, -\phi_{\mathbf{k}}) = \bar{D}^{(\ell)}_{m'm}(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) \,.$$

Furthermore using the fact that the basis vectors $Y_{\ell m}$ transform with the transpose of the matrix with which the coefficients of vectors transform, we obtain (see also Appendix A4.2.3, Eqs. (A4.37) and (A4.41)):

$$\begin{aligned} Y_{\ell m}(\mathbf{n};\mathbf{k}) &= \sum_{m'} Y_{\ell m'}(\mathbf{n};\mathbf{e}) \bar{D}^{(\ell)}_{mm'}(\theta_{\mathbf{k}}, \phi_{\mathbf{k}}, 0) \\ &= \sqrt{\frac{4\pi}{2\ell+1}} \sum_{m'} Y_{\ell m'}(\mathbf{n};\mathbf{e}) \, {}_m\bar{Y}_{\ell m'}(\mathbf{k};\mathbf{e}) \,. \end{aligned} \tag{5.63}$$

Note how the magnetic quantum number m in the **k**-basis transfers to the spin weight in the **e**-basis. Equation (5.63) is a generalization of the addition theorem of spherical harmonics (see also Appendix A4.2).

Inserting this in the Fourier decomposition, Eq. (5.33) we can isolate the coefficient $a_{\ell m}$ as the term proportional to $Y_{\ell m}(\mathbf{n};\mathbf{e})$. We use the orthogonality of spherical harmonics which implies

$$a_{\ell m}(\mathbf{x}) = \int d\Omega_{\mathbf{n}} Y^*_{\ell m}(\mathbf{n};\mathbf{e})\mathcal{M}(\mathbf{x},\mathbf{n}) \,.$$

Inserting $\mathcal{M}(\mathbf{x},\mathbf{n})$ from Eq. (5.33) and making use of the identity, Eq. (5.63) we obtain finally

$$a_{\ell m}(\mathbf{x}) = \sqrt{\frac{4\pi}{2\ell+1}} \sum_{s=-2}^{2} \int \frac{d^3k}{(2\pi)^3} \mathcal{M}^{(s)}_\ell(\mathbf{k})\, {}_sY^*_{\ell m}(\mathbf{k};\mathbf{e}) e^{-i\mathbf{x}\cdot\mathbf{k}} \,. \tag{5.64}$$

Because of statistical homogeneity, coefficients $\mathcal{M}^{(s)}_\ell(\mathbf{k})$ with different values of $\mathbf{k}$ are uncorrelated. We introduce the power spectrum of $\mathcal{M}^{(s)}_\ell(\mathbf{k})$ which is of the form

$$\left\langle \mathcal{M}^{(s)}_\ell(\mathbf{k}) \mathcal{M}^{(s)*}_\ell(\mathbf{k}') \right\rangle \equiv (2\pi)^3 \delta^3(\mathbf{k}-\mathbf{k}') M^{(s)}_\ell(k) \,. \tag{5.65}$$

Because of statistical isotropy, $M^{(s)}_\ell(k)$ is a function of the modulus $k = |\mathbf{k}|$ only, and the $\mathcal{M}^{(s)}_\ell(\mathbf{k})$s with different ℓ or s are uncorrelated. With this and Eq. (5.64) integration over angles leads to

$$(2\ell+1)^2 C_\ell = (2\ell+1)^2 \langle |a_{\ell m}|^2 \rangle = \frac{2}{\pi} \sum_{s=-2}^{2} \int dk\, k^2 M^{(s)}_\ell(k) \,. \tag{5.66}$$

We now address the polarization spectra. Here, the situation is somewhat more complicated, since apart from the dependence of ${}_sY_{\ell m}(\mathbf{n};\mathbf{k})$ on the chosen z axis, the spin weighted spherical harmonics also depend on the polarization basis normal to $\mathbf{n}$. In addition to the rotation from the $\mathbf{k}$-basis into the $\mathbf{e}$-basis outlined above, we would also have to fix the polarization basis. To avoid this complication we again use the spin raising and lowering operator $\eth$ and its hermitian conjugate $\eth^*$. With Eqs. (5.13) and (5.14) we define the following scalar quantities which are closely related to $\mathcal{E}$ and $\mathcal{B}$,

$$\begin{aligned} \tilde{\mathcal{E}}(\mathbf{x},\mathbf{n}) &= \frac{1}{2}\left[(\eth^*)^2(\mathcal{Q}+i\mathcal{U})(\mathbf{x},\mathbf{n}) + \eth^2(\mathcal{Q}-i\mathcal{U})(\mathbf{x},\mathbf{n})\right] \\ &= \sum_{\ell=2}^{\infty} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \sum_{m=-\ell}^{\ell} e_{\ell m}(\mathbf{x}) Y_{\ell m}(\mathbf{n};\mathbf{e}) \,, \end{aligned} \tag{5.67}$$

and

$$\begin{aligned} \tilde{\mathcal{B}}(\mathbf{x},\mathbf{n}) &= \frac{-i}{2}\left[(\eth^*)^2(\mathcal{Q}+i\mathcal{U})(\mathbf{x},\mathbf{n}) - \eth^2(\mathcal{Q}-i\mathcal{U})(\mathbf{x},\mathbf{n})\right] \\ &= \sum_{\ell=2}^{\infty} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \sum_{m=-\ell}^{\ell} b_{\ell m}(\mathbf{x}) Y_{\ell m}(\mathbf{n};\mathbf{e}) \,. \end{aligned} \tag{5.68}$$

Here we have inserted the original expansion of $\mathcal{Q} \pm i\mathcal{U}$ given in Eq. (5.11).

To relate $\tilde{\mathcal{E}}(\mathbf{x},\mathbf{n})$ and $\tilde{\mathcal{B}}(\mathbf{x},\mathbf{n})$ to their Fourier transforms, which can be obtained from Eq. (5.35), we first rotate ${}_{\pm 2}G_{\ell m}(\mathbf{x},\mathbf{n})$ into the **e**-basis, using

$$ {}_sY_{\ell m}(\mathbf{n};\mathbf{k}) = \sqrt{\frac{4\pi}{2\ell+1}} \sum_{m'} {}_sY_{\ell m'}(\mathbf{n};\mathbf{e})\, {}_{-m}Y^*_{\ell m'}(\mathbf{k};\mathbf{e}) \,, \tag{5.69} $$

which is derived exactly like for $s = 0$. But since we do not take notice of the orientation of the polarization basis, the latter is still oriented in a **k** dependent manner and ${}_sY_{\ell m'}(\mathbf{n};\mathbf{e})$ still depends on **k** over the orientation of the polarization basis. We now act with the operator $\eth^2$ for $s = -2$ and $(\eth^*)^2$ for $s = 2$ on ${}_sY_{\ell m'}(\mathbf{n};\mathbf{e})$ to obtain $Y_{\ell m'}(\mathbf{n};\mathbf{e})$. Using Eq. (5.63) we then find

$$ \tilde{\mathcal{E}}(\mathbf{x},\mathbf{n}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \sum_{\ell=2}^{\infty} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \sum_{m=-2}^{2} \mathcal{E}^{(m)}_\ell(\mathbf{k}) Y_{\ell m}(\mathbf{n};\mathbf{k}) \,, \tag{5.70} $$

$$ \tilde{\mathcal{B}}(\mathbf{x},\mathbf{n}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \sum_{\ell=2}^{\infty} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \sum_{m=-2}^{2} \mathcal{B}^{(m)}_\ell(\mathbf{k}) Y_{\ell m}(\mathbf{n};\mathbf{k}) \,. \tag{5.71} $$

This is exactly the same result as when acting directly with $\eth$ and $\eth^*$ on ${}_sY_{\ell m'}(\mathbf{n};\mathbf{k})$ which is not entirely obvious since in general the operators $\eth$ and $\eth^*$ are basis dependent. However, as we have seen the relations Eqs. (5.13) and (5.14) are valid in every basis. Since both sides of these equations have spin-0, they are independent of the polarization basis.

Now that we have expressed polarization in terms of ordinary spherical harmonics, we can proceed as for the temperature anisotropies. We rotate $Y_{\ell m}(\mathbf{n};\mathbf{k})$ into spin weighted harmonics ${}_mY_{\ell m'}(\mathbf{n};\mathbf{e})$, and obtain,

$$ \begin{aligned} (2\ell+1)^2 C^{(\mathcal{E})}_\ell &\equiv (2\ell+1)^2 \langle |e_{\ell m}(\mathbf{x})| \rangle \\ &= \frac{2}{\pi} \sum_{s=-2}^{2} \int dk\, k^2 E^{(s)}_\ell(k) \,, \end{aligned} \tag{5.72} $$

$$ \begin{aligned} (2\ell+1)^2 C^{(\mathcal{B})}_\ell &\equiv (2\ell+1)^2 \langle |b_{\ell m}(\mathbf{x})| \rangle \\ &= \frac{2}{\pi} \sum_{s=-2}^{2} \int dk\, k^2 B^{(s)}_\ell(k) \,, \end{aligned} \tag{5.73} $$

and

$$ \begin{aligned} (2\ell+1)^2 C^{(\mathcal{ME})}_\ell &\equiv (2\ell+1)^2 \langle a^*_{\ell m} e_{\ell m}(\mathbf{x}) \rangle \\ &= \frac{2}{\pi} \sum_{s=-2}^{2} \int dk\, k^2 F^{(s)}_\ell(k) \,, \end{aligned} \tag{5.74} $$

where we have introduced the power spectra

$$\left\langle \mathcal{E}_\ell^{(s)}(\mathbf{k})\mathcal{E}_\ell^{(s)*}(\mathbf{k}')\right\rangle \equiv (2\pi)^3\delta^3(\mathbf{k}-\mathbf{k}')E_\ell^{(s)}(k)\,, \tag{5.75}$$

$$\left\langle \mathcal{B}_\ell^{(s)}(\mathbf{k})\mathcal{B}_\ell^{(s)*}(\mathbf{k}')\right\rangle \equiv (2\pi)^3\delta^3(\mathbf{k}-\mathbf{k}')B_\ell^{(s)}(k)\,, \tag{5.76}$$

$$\left\langle \mathcal{E}_\ell^{(s)}(\mathbf{k})\mathcal{M}_\ell^{(s)*}(\mathbf{k}')\right\rangle \equiv (2\pi)^3\delta^3(\mathbf{k}-\mathbf{k}')F_\ell^{(s)}(k)\,. \tag{5.77}$$

To relate these spectra to meaningful correlation functions, we correlate quantities which are scalars under rotations around $\mathbf{n}$ and $\mathbf{n}'$ respectively, hence quantities which can be expanded in ordinary, $s=0$ spherical harmonics. For this we use our quantities $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{B}}$. The same derivation which led to Eq. (5.59) now yields

$$\left\langle \tilde{\mathcal{E}}(t,\mathbf{x},\mathbf{n})\tilde{\mathcal{E}}(t,\mathbf{x},\mathbf{n}')\right\rangle = \frac{1}{4\pi}\sum_{\ell=0}^{\infty}\frac{(2\ell+2)!}{(2\ell-2)!}(2\ell+1)P_\ell(\mathbf{n}\cdot\mathbf{n}')C_\ell^{(\mathcal{E})}\,, \tag{5.78}$$

$$\left\langle \tilde{\mathcal{B}}(t,\mathbf{x},\mathbf{n})\tilde{\mathcal{B}}(t,\mathbf{x},\mathbf{n}')\right\rangle = \frac{1}{4\pi}\sum_{\ell=0}^{\infty}\frac{(2\ell+2)!}{(2\ell-2)!}(2\ell+1)P_\ell(\mathbf{n}\cdot\mathbf{n}')C_\ell^{(\mathcal{B})}\,, \tag{5.79}$$

$$\left\langle \mathcal{M}(t,\mathbf{x},\mathbf{n})\tilde{\mathcal{E}}(t,\mathbf{x},\mathbf{n}')\right\rangle = \frac{1}{4\pi}\sum_{\ell=0}^{\infty}\sqrt{\frac{(2\ell+2)!}{(2\ell-2)!}}(2\ell+1)P_\ell(\mathbf{n}\cdot\mathbf{n}')C_\ell^{(\mathcal{ME})}\,. \tag{5.80}$$

It is easier to interpret the scalar polarization amplitudes $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{B}}$, than $\mathcal{E}$ and $\mathcal{B}$ which are lacking the factors $(2\ell+2)!/(2\ell-2)!$ in the expansion, since the former are simply related to derivatives of the polarization tensor. In Appendix A4.2.4 we show that the operators $\eth$ and $\eth^*$ are the covariant derivatives on the sphere in the direction $\mathbf{e}_\pm = (1/\sqrt{2})(\mathbf{e}_1 \mp i\mathbf{e}_2)$. Observing that $Q \pm iU$ actually correspond to the $+,+$ and $-,-$ components of the polarization tensor $\mathcal{P}_{ab}$ defined in Eq. (5.1), we find that

$$(\eth^*)^2(Q+iU) = 2\nabla_-\nabla_-\mathcal{P}_{++}\,, \quad \eth^2(Q-iU) = 2\nabla_+\nabla_+\mathcal{P}_{--}\,. \tag{5.81}$$

Here we have used $\eth = -\sqrt{2}\nabla_+$ and $\eth^* = -\sqrt{2}\nabla_-$ which is derived in Appendix A4.2.4. We also note that in two dimensions the curl of a vector, $\mathrm{rot}V \equiv \epsilon_{ij}\nabla_i V_j$ is a (pseudo-)scalar, hence the double curl of a tensor, $\mathrm{rot\,rot}T \equiv \epsilon_{lm}\epsilon_{ij}\nabla_l\nabla_i T_{jm}$ is a scalar. A short calculation (see Appendix A4.2.4) now shows that

$$\begin{aligned}\nabla_-\nabla_-\mathcal{P}_{++} + \nabla_+\nabla_+\mathcal{P}_{--} &= 2\nabla_i\nabla_j\mathcal{P}_{ij} = 2\mathrm{div\,div}\mathcal{P}\,,\\ \nabla_-\nabla_-\mathcal{P}_{++} - \nabla_+\nabla_+\mathcal{P}_{--} &= 2\epsilon_{lm}\epsilon_{ij}\nabla_l\nabla_i\mathcal{P}_{jm} = 2\mathrm{rot\,rot}\mathcal{P}\,,\end{aligned} \tag{5.82}$$

so that

$$\tilde{\mathcal{E}} = 2\mathrm{div\,div}\mathcal{P} \quad \text{and} \quad \tilde{\mathcal{B}} = 2\mathrm{rot\,rot}\mathcal{P}\,. \tag{5.83}$$

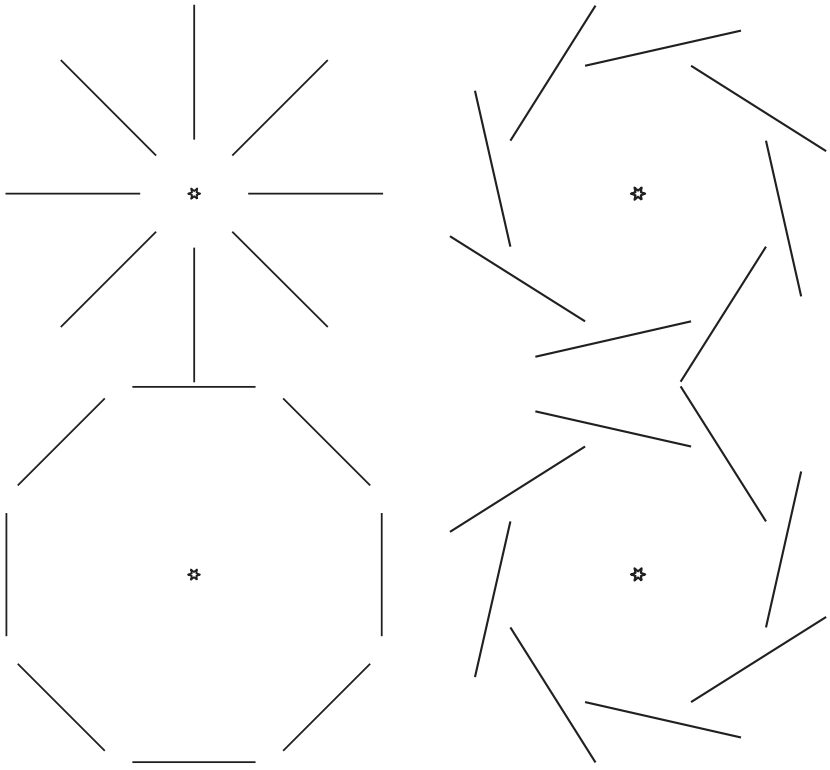


Fig. 5.2. E-polarization (left) and B-polarization (right) patterns are shown around the photon direction indicated as the centre. E-polarization can be either radial or tangential, while B-polarization is clearly of curl type.

Hence $\tilde{\mathcal{E}}$ measures 'gradient-type' polarization while $\tilde{\mathcal{B}}$ measures curl-type polarization. More precisely, if we split the electric field tangent to the sphere of directions $\mathbf{n}$ into a gradient part and a curl part, $E_i = \nabla_i f + \epsilon_{ij}\nabla_j g$, we obtain

$$\tilde{\mathcal{E}} = 2\nabla_i\nabla_j\left(\nabla_i f \nabla_j f^* - \frac{1}{2}\delta_{ij}|\nabla f|^2\right), \quad \text{and}$$
$$\tilde{\mathcal{B}} = 2\nabla_i\nabla_j\left(\nabla_i g \nabla_j g^* - \frac{1}{2}\delta_{ij}|\nabla g|^2\right). \tag{5.84}$$

Hence $\tilde{\mathcal{B}}$ (and $\mathcal{B}$) is measuring the curl component in the electric field while $\tilde{\mathcal{E}}$ (and $\mathcal{E}$) is measuring the gradient component. This is derived in Ex. 5.2. In Fig. 5.2 examples of E- and B-polarization are shown. Note that the functions $\mathcal{E}$ and $\mathcal{B}$ are not local, they do not have a direct interpretation in terms of the measurable polarization pattern $Q \pm iU$.

5.4 The small-scale limit and the physical meaning of $\mathcal{E}$ and $\mathcal{B}$

The polarization variables $\mathcal{E}$ and $\mathcal{B}$ are easier to interpret in the small-scale limit. For $\ell \gtrsim 100$ which corresponds to angles of less than about 2°, we may neglect the curvature of the sphere of directions and consider it as a plane normal to $\mathbf{e}_z$. In this approximation, the spherical harmonics can be replaced by exponentials, the eigenfunctions of the Laplacian on the plane.

$$Y_{\ell m}(\mathbf{n}) \to \frac{1}{2\pi} \exp(i\boldsymbol{\ell} \cdot \mathbf{x}) \,, \tag{5.85}$$

where $\mathbf{x}$ is a small vector in the plane normal to $\mathbf{e}_z$ and $\boldsymbol{\ell} = \ell(\cos\varphi_\ell, \sin\varphi_\ell)$ is a vector in the 'Fourier plane'. In this approximation the magnetic quantum number m is replaced by the continuous direction of the vector $\boldsymbol{\ell}$. The orthogonality relation now becomes

$$\frac{1}{(2\pi)^2} \int d^2x \, e^{i\mathbf{x}(\boldsymbol{\ell}-\boldsymbol{\ell}')} = \delta^2(\boldsymbol{\ell} - \boldsymbol{\ell}') \,.$$

The temperature anisotropy is given by

$$\frac{\Delta T}{T}(\mathbf{x}) = \mathcal{M}(\mathbf{x}) = \frac{1}{2\pi} \int d^2\boldsymbol{\ell} \, \mathcal{M}(\boldsymbol{\ell}) \, e^{i\mathbf{x}\cdot\boldsymbol{\ell}}, \tag{5.86}$$

$$\mathcal{M}(\boldsymbol{\ell}) = \frac{1}{2\pi} \int d^2\mathbf{x} \, \mathcal{M}(\mathbf{x}) e^{-i\mathbf{x}\cdot\boldsymbol{\ell}} \,. \tag{5.87}$$

The spin weighted spherical harmonics $s = 2$ become

$${}_2Y_{\ell m} = \sqrt{\frac{(\ell-2)!}{(\ell+2)!}} \, \eth^2 Y_{\ell m} \to \frac{1}{2\pi} \ell^{-2} \, \eth^2 \, e^{i\mathbf{x}\cdot\boldsymbol{\ell}}, \tag{5.88}$$

$${}_{-2}Y_{\ell m} = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \, \eth^{*2} Y_{\ell m} \to \frac{1}{2\pi} \ell^{-2} \, \eth^{*2} \, e^{i\mathbf{x}\cdot\boldsymbol{\ell}} \,. \tag{5.89}$$

Inserting this in Eq. (5.11) yields

$$(\mathcal{Q} + i\mathcal{U})(\mathbf{x}) = \frac{1}{2\pi} \int d^2\boldsymbol{\ell} \, (\mathcal{E}(\boldsymbol{\ell}) + i\mathcal{B}(\ell)) \, \frac{1}{\ell^2} \, \eth^2 \, e^{i\mathbf{x}\cdot\boldsymbol{\ell}} \,, \tag{5.90}$$

$$(\mathcal{Q} - i\mathcal{U})(\mathbf{x}) = \frac{1}{2\pi} \int d^2\boldsymbol{\ell} \, (\mathcal{E}(\boldsymbol{\ell}) - i\mathcal{B}(\ell)) \, \frac{1}{\ell^2} \, \eth^{*2} \, e^{i\mathbf{x}\cdot\boldsymbol{\ell}} \,. \tag{5.91}$$

We orient the coordinate system such that $\eth = -(\nabla_\vartheta - i\nabla_\varphi) = -(\nabla_x - i\nabla_y)$ at $\mathbf{e}_z$ and $\eth e^{i\mathbf{x}\cdot\boldsymbol{\ell}} = -i(\ell_x - i\ell_y)e^{i\mathbf{x}\cdot\boldsymbol{\ell}} = -i\ell e^{-i\varphi_\ell} e^{i\mathbf{x}\cdot\boldsymbol{\ell}}$. With this we obtain

$$\eth^2 e^{i\mathbf{x}\cdot\boldsymbol{\ell}} = -\ell^2 e^{-2i\varphi_\ell} e^{i\mathbf{x}\cdot\boldsymbol{\ell}} \,, \tag{5.92}$$

$$\eth^{*2} e^{i\mathbf{x}\cdot\boldsymbol{\ell}} = -\ell^2 e^{2i\varphi_\ell} e^{i\mathbf{x}\cdot\boldsymbol{\ell}} \,. \tag{5.93}$$

In the small-scale limit, the Stokes parameters are therefore given in terms of $\mathcal{E}(\ell)$ and $\mathcal{B}(\ell)$ by

$$\mathcal{Q}(\mathbf{x}) = \frac{-1}{2\pi}\int d^2\ell\,[\mathcal{E}(\ell)\cos(2\varphi_\ell) - \mathcal{B}(\ell)\sin(2\varphi_\ell)]\,e^{i\mathbf{x}\cdot\ell}\,, \tag{5.94}$$

$$\mathcal{U}(\mathbf{x}) = \frac{-1}{2\pi}\int d^2\ell\,[\mathcal{E}(\ell)\sin(2\varphi_\ell) + \mathcal{B}(\ell)\cos(2\varphi_\ell)]\,e^{i\mathbf{x}\cdot\ell}\,. \tag{5.95}$$

These relations were introduced by Seljak (1996b), where E- and B-polarizations have been introduced for the first time. They can be inverted to

$$\mathcal{E}(\ell) = \frac{-1}{2\pi}\int d^2x\,[\mathcal{Q}(\mathbf{x})\cos(2\varphi) + \mathcal{U}(\mathbf{x})\sin(2\varphi)]\,e^{-i\mathbf{x}\cdot\ell}\,, \tag{5.96}$$

$$\mathcal{B}(\ell) = \frac{-1}{2\pi}\int d^2x\,[\mathcal{U}(\mathbf{x})\cos(2\varphi) - \mathcal{Q}(\mathbf{x})\sin(2\varphi)]\,e^{-i\mathbf{x}\cdot\ell}\,. \tag{5.97}$$

A short calculation, see Ex. 5.3, leads to the following relation of $\mathcal{E}$, $\mathcal{B}$ and $\mathcal{Q}, \mathcal{U}$ in real space:

$$\mathcal{E}(\mathbf{x}) = \nabla^{-2}(\partial_x^2 - \partial_y^2)\mathcal{Q}(\mathbf{x}) + \nabla^{-2}2\partial_x\partial_y\mathcal{U}(\mathbf{x})\,, \tag{5.98}$$

$$\mathcal{B}(\mathbf{x}) = \nabla^{-2}(\partial_x^2 - \partial_y^2)\mathcal{U}(\mathbf{x}) - \nabla^{-2}2\partial_x\partial_y\mathcal{Q}(\mathbf{x})\,. \tag{5.99}$$

Hence $\mathcal{E}$ and $\mathcal{B}$ which are the inverse Laplacians of combinations of second derivatives of $\mathcal{Q}$ and $\mathcal{U}$ are more closely related to the latter than $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{B}}$, they have no additional factors of ℓ; but the relation is non-local. Because of the inverse Laplacians we have to know $\mathcal{U}$ and $\mathcal{Q}$ globally to determine $\mathcal{E}$ and $\mathcal{B}$, while $\tilde{\mathcal{E}} = \nabla^2\mathcal{E}$ and $\tilde{\mathcal{B}} = \nabla^2\mathcal{B}$ are locally related to $\mathcal{Q}$ and $\mathcal{U}$. Furthermore, since Q-polarization turns into U-polarization and vice versa if we rotate the coordinate system by 45°, a pure $\mathcal{E}$-polarization configuration turns into pure $\mathcal{B}$, if we turn all the polarization vectors by 45°.

If we have a pure gradient-type polarization, $E_i = \nabla_i f$, we find

$$\mathcal{E} = \nabla^{-2}\left(\nabla_i\nabla_j(\nabla_i f\nabla_j f^*)\right) - \frac{1}{2}|\nabla f|^2\,.$$

Whereas a pure curl-type polarization is related to $\mathcal{B}$ via

$$\mathcal{B} = \nabla^{-2}\left(\nabla_i\nabla_j(\nabla_i g\nabla_j g^*)\right) - \frac{1}{2}|\nabla g|^2\,,$$

if $E_j = \epsilon_{im}\nabla_m g$. Vanishing polarization corresponds to $\mathcal{B} = \mathcal{E} = 0$. Positive values of $\mathcal{E}$ around a zero indicate radial polarization patterns while negative values indicate tangential polarization. A B-polarization pattern can then be obtained by simply rotating the polarization vectors by 45°. Hence the B-polarization patterns rotate around their zeros, see Fig. 5.3.

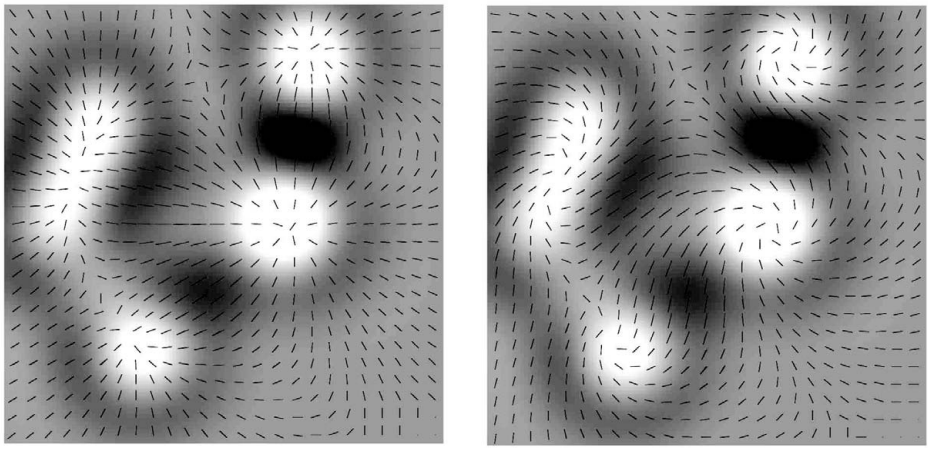


Fig. 5.3. A E-polarization pattern (left) is compared with B-polarization. The function $\tilde{\mathcal{E}}$ is indicated in grey scale, and the polarization directions are drawn. E-polarization is tangential along the dark negative regions while it is radial from the white positive regions. The B-polarization pattern is obtained by rotating the polarization directions by 45°.

We finally want to derive expressions for the correlation functions in the small-scale limit. For the temperature anisotropies we use the fact that the correlation function is simply the Fourier transform of the power spectrum. In the small-scale limit, the definition of the temperature anisotropy spectrum yields

$$\langle \mathcal{M}(\boldsymbol{\ell})\mathcal{M}^*(\boldsymbol{\ell}')\rangle = \delta(\boldsymbol{\ell}-\boldsymbol{\ell}')C_\ell^{(\mathcal{M})} \,.$$

Hence

$$\begin{aligned}\xi_{\mathcal{M}}(\mathbf{x}) \equiv \langle \mathcal{M}(\mathbf{y})\mathcal{M}(\mathbf{y}+\mathbf{x})\rangle &= \frac{1}{(2\pi)^2}\int d^2\boldsymbol{\ell}\, e^{i\boldsymbol{\ell}\mathbf{x}} C_\ell^{(\mathcal{M})} \\ &= \frac{1}{(2\pi)^2}\int d\ell\, \ell C_\ell^{(\mathcal{M})}\int_0^{2\pi} e^{\ell r\cos\phi} = \frac{1}{2\pi}\int_0^\infty \ell\, d\ell\, J_0(r\ell)C_\ell \,. \end{aligned} \quad (5.100)$$

For the integral over the angle ϕ between $\mathbf{x}$ and $\boldsymbol{\ell}$ we have set $r = |\mathbf{x}|$ and

$$\int d\phi\, e^{-ir\ell\cos\phi} = 2\pi J_0(r\ell) \,.$$

To see this we can use the formula given in Appendix A4.3,

$$e^{iy\cos\phi} = \sum_{n=-\infty}^{\infty} i^n J_n(y)\, e^{in\phi} = J_0(y) + 2\sum_{n=1}^{\infty} i^n J_n(y)\cos(n\phi) \,. \quad (5.101)$$

Integrating this expansion yields

$$\frac{1}{2\pi}\int_0^{2\pi} d\phi\, e^{iy\cos\phi}\, e^{-in\phi} = i^n J_n(y)\,. \tag{5.102}$$

Starting from the correlation function we can derive the equivalent expression for the power spectrum

$$C_\ell = 2\pi \int_0^\infty r\, dr\, J_0(r\ell)\xi(r)\,. \tag{5.103}$$

To derive the correlation functions for polarization, we introduce the variable $\mathcal{P} = \mathcal{Q} + i\mathcal{U}$ and, correspondingly $\bar{\mathcal{P}} = \mathcal{Q} - i\mathcal{U}$. According to Eqs. (5.94)–(5.97), their Fourier representations are

$$\mathcal{P} = \mathcal{Q} + i\mathcal{U} = -\int \frac{d^2\ell}{2\pi}\,[\mathcal{E}(\ell) + i\mathcal{B}(\ell)]\, e^{2i\phi}\, e^{i\boldsymbol{\ell}\cdot\mathbf{x}}\,, \tag{5.104}$$

$$\bar{\mathcal{P}} = \mathcal{Q} - i\mathcal{U} = -\int \frac{d^2\ell}{2\pi}\,[\mathcal{E}(\ell) - i\mathcal{B}(\ell)]\, e^{-2i\phi}\, e^{i\boldsymbol{\ell}\cdot\mathbf{x}}\,, \tag{5.105}$$

or, inversely

$$\mathcal{E}(\ell) + i\mathcal{B}(\ell) = -\int \frac{d^2\mathbf{x}}{2\pi}\,\mathcal{P} e^{-2i\phi}\, e^{i\boldsymbol{\ell}\cdot\mathbf{x}}\,, \tag{5.106}$$

$$\mathcal{E}(\ell) - i\mathcal{B}(\ell) = -\int \frac{d^2\mathbf{x}}{2\pi}\,\mathcal{P} e^{2i\phi}\, e^{i\boldsymbol{\ell}\cdot\mathbf{x}}\,. \tag{5.107}$$

We want to define correlation functions of $\mathcal{P}$ and $\bar{\mathcal{P}}$ in a coordinate independent way. For two given points $\mathbf{x} \neq \mathbf{x}'$, $\mathbf{r} \equiv \mathbf{x} - \mathbf{x}'$ we rotate the polarization basis by the angle ϕ_r which $\mathbf{r}$ encloses with the x axis. The new polarization basis, $\hat{\mathbf{r}}$ and the direction orthogonal to it, is uniquely defined by $\mathbf{r}$. The rotated polarization is given by

$$\mathcal{P}_r(\mathbf{x}) = e^{-2i\phi_r}\,\mathcal{P}(\mathbf{x})\,.$$

With respect to this intrinsic basis we can now define the coordinate independent correlation functions

$$\begin{aligned}\xi_+(\mathbf{r}) &= \langle \bar{\mathcal{P}}_r(\mathbf{x})\mathcal{P}_r(\mathbf{x}')\rangle = \langle \bar{\mathcal{P}}(\mathbf{x})\mathcal{P}(\mathbf{x}')\rangle \\ &= \langle \mathcal{Q}(\mathbf{x})\mathcal{Q}(\mathbf{x}')\rangle + \langle \mathcal{U}(\mathbf{x})\mathcal{U}(\mathbf{x}')\rangle\,, \end{aligned} \tag{5.108}$$

$$\begin{aligned}\xi_-(\mathbf{r}) &= \langle \mathcal{P}_r(\mathbf{x})\mathcal{P}_r(\mathbf{x}')\rangle = \langle e^{-4i\phi_r}\mathcal{P}(\mathbf{x})\mathcal{P}(\mathbf{x}')\rangle \\ &= \langle \mathcal{Q}_r(\mathbf{x})\mathcal{Q}_r(\mathbf{x}')\rangle - \langle \mathcal{U}_r(\mathbf{x})\mathcal{U}_r(\mathbf{x}')\rangle + i\left(\langle \mathcal{Q}_r(\mathbf{x})\mathcal{U}_r(\mathbf{x}')\rangle - \langle \mathcal{U}_r(\mathbf{x})\mathcal{Q}_r(\mathbf{x}')\rangle\right)\,, \end{aligned} \tag{5.109}$$

$$\begin{aligned}\xi_\times(\mathbf{r}) &= \langle \mathcal{P}_r(\mathbf{x})\mathcal{M}(\mathbf{x}')\rangle = \langle e^{-2i\phi_r}\mathcal{P}(\mathbf{x})\mathcal{M}(\mathbf{x}')\rangle \\ &= \langle \mathcal{Q}_r(\mathbf{x})\mathcal{M}(\mathbf{x}')\rangle + i\,\langle \mathcal{U}_r(\mathbf{x})\mathcal{M}(\mathbf{x}')\rangle\,. \end{aligned} \tag{5.110}$$

Under parity, $\mathbf{r} \to -\mathbf{r}$, ϕ_r and with it the imaginary part of the terms $e^{ni\phi_r}$ change sign. If we assume statistical parity invariance, they therefore have to vanish,

$$\langle \mathcal{U}_r(\mathbf{x})\mathcal{Q}_r(\mathbf{x}')\rangle \equiv \langle \mathcal{U}_r(\mathbf{x})\mathcal{M}(\mathbf{x}')\rangle \equiv 0\,.$$

This expresses the fact that $\mathcal{B}$-polarization is uncorrelated with $\mathcal{E}$-polarization and the temperature anisotropies in terms of the correlation functions.

The calculation of the correlation function ξ_+ now is exactly analogous to that for the temperature anisotropy, one just has to replace C_ℓ by $C_\ell^{(\mathcal{E})} + C_\ell^{(\mathcal{B})}$,

$$\begin{aligned}\xi_+(r) &= \langle \bar{\mathcal{P}}(\mathbf{x})\mathcal{P}(\mathbf{x}')\rangle \\ &= \frac{1}{2\pi}\int \ell\, d\ell \left[C_\ell^{(\mathcal{E})} + C_\ell^{(\mathcal{B})}\right] J_0(\ell r)\,. \end{aligned} \tag{5.111}$$

For ξ_- and $\xi_\times$, the situation is somewhat different because of the exponentials $e^{im\phi_r}$. We insert the Fourier transform of $\mathcal{P}(\mathbf{x})$ given in Eq. (5.104) in the expression for ξ_-, we find

$$\begin{aligned}\xi_-(r) &= \langle \mathcal{P}_r(\mathbf{x})\mathcal{P}_r(\mathbf{x}')\rangle \\ &= \int \frac{d^2\ell}{2\pi}\frac{d^2\ell'}{2\pi}\langle[\mathcal{E}(\ell) + i\mathcal{B}(\ell)][\mathcal{E}^*(\ell') + i\mathcal{B}^*(\ell')]\rangle\, e^{i(\ell\cdot\mathbf{x}-\ell'\cdot\mathbf{x}')}e^{i(2\phi_\ell+2\phi_{\ell'}-4\phi_r)} \\ &= \int \frac{d^2\ell}{(2\pi)^2}\left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right] e^{ir\ell\cos\phi_r}e^{4i\phi_r} \\ &= \frac{1}{2\pi}\int_0^\infty d\ell\,\ell J_4(r\ell)\left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right]\,. \end{aligned} \tag{5.112}$$

For the last equals sign we have used Eq. (5.102) and the fact that Bessel functions with an even index are even. Similarly we obtain for the cross correlation function

$$\begin{aligned}\xi_\times(r) &= \langle \mathcal{P}_r(\mathbf{x})\mathcal{M}_r(\mathbf{x}')\rangle \\ &= -\int \frac{d^2\ell}{2\pi}\frac{d^2\ell'}{2\pi}\langle[\mathcal{E}(\ell) + i\mathcal{B}(\ell)]\mathcal{M}^*(\ell')\rangle e^{i(\ell\cdot\mathbf{x}-\ell'\cdot\mathbf{x}')}e^{i(2\phi_\ell-2\phi_r)} \\ &= \frac{1}{2\pi}\int_0^\infty d\ell\,\ell J_2(r\ell)C_\ell^{(\mathcal{ME})}\,. \end{aligned} \tag{5.113}$$

As for the temperature anisotropy, the polarization power spectra and correlation functions are related via two-dimensional Fourier transforms. Taking into account the correct factors $e^{i(\phi_\ell-\phi_r)}$ coming from the definitions, Eqs. (5.109) and (5.110) and the expression (5.104), we find

$$C_\ell^{(\mathcal{E})} + C_\ell^{(\mathcal{B})} = 2\pi\int r\,dr\, J_0(\ell r)\xi_+(r)\,, \tag{5.114}$$

$$C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})} = 2\pi\int r\,dr\, J_4(\ell r)\xi_-(r)\,, \tag{5.115}$$

$$C_\ell^{(\mathcal{EM})} = 2\pi\int r\,dr\, J_2(\ell r)\xi_\times(r)\,. \tag{5.116}$$

These small-scale expressions for the temperature and polarization power spectra and for the correlation functions will be especially useful when we discuss lensing in Chapter 7.

5.5 The Boltzmann equation

In this section we write the Boltzmann equation for the mode functions $\mathcal{M}_\ell^{(m)}$, $\mathcal{E}_\ell^{(m)}$ and $\mathcal{B}_\ell^{(m)}$ introduced in Section 5.3. First we note that the usual free-streaming term is given by

$$i\mu k \equiv ik\sqrt{\frac{4\pi}{3}}Y_{10}(\mathbf{n})\,. \tag{5.117}$$

Furthermore, as we show in Appendix A4.2.4,

$$\sqrt{\frac{4\pi}{3}}Y_{10}\cdot\,{}_sY_{\ell m} = \sqrt{\frac{\left[(\ell+1)^2-m^2\right]\left[(\ell+1)^2-s^2\right]}{(\ell+1)^2(2\ell+3)(2\ell+1)}}\,{}_sY_{\ell+1\,m} - \frac{ms}{\ell(\ell+1)}\,{}_sY_{\ell m} + \sqrt{\frac{(\ell^2-m^2)(\ell^2-s^2)}{\ell^2(2\ell+1)(2\ell-1)}}\,{}_sY_{\ell-1\,m}\,. \tag{5.118}$$

This determines the free streaming of the modes ${}_0G_{\ell m}$ which hence couple to ${}_0G_{\ell+1,m}$ and ${}_0G_{\ell-1,m}$. The $\mathcal{E}_\ell^{(m)}$-mode, which is proportional to ${}_2G_{\ell m}+{}_{-2}G_{\ell m}$, couples to the $\mathcal{E}_{\ell\pm1}^{(m)}$-modes and to the $\mathcal{B}_\ell^{(m)}$-mode, which multiplies ${}_2G_{\ell m}-{}_{-2}G_{\ell m}$. Correspondingly, free streaming couples $\mathcal{B}_\ell^{(m)}$ to $\mathcal{B}_{\ell\pm1}^{(m)}$ and $\mathcal{E}_\ell^{(m)}$. Therefore, even if B-modes are not generated by Thomson scattering, as we shall see below, they are generated by free streaming from the $\mathcal{E}$-modes. Only the scalar $\mathcal{B}$- and $\mathcal{E}$-modes for which $m=0$, are not coupled. Therefore, if $\mathcal{B}_\ell^{(0)}$ vanishes initially it will remain zero. With Eq. (5.118), the left-hand side of the Boltzmann equation turns into the mode equations

$$(\partial_t+\mathbf{n}\cdot\nabla)\mathcal{M}_\ell^{(m)}\,Y_{\ell m} = \left[\partial_t+\sqrt{\frac{\left[(\ell+1)^2-m^2\right]}{(2\ell+3)(2\ell+1)}}Y_{\ell+1,m}+\sqrt{\frac{(\ell^2-m^2)}{(2\ell+1)(2\ell-1)}}Y_{\ell-1,m}\right]\mathcal{M}_\ell^{(m)}\,, \tag{5.119}$$

$$(\partial_t+\mathbf{n}\cdot\nabla)\left[\mathcal{E}_\ell^{(m)}\pm i\mathcal{B}_\ell^{(m)}\right]({}_{\pm2}Y_{\ell m}) = \left[\partial_t+\sqrt{\frac{\left[(\ell+1)^2-m^2\right]\left[(\ell+1)^2-4\right]}{(\ell+1)^2(2\ell+3)(2\ell+1)}}\left({}_{\pm2}Y_{\ell+1,m}\right)\mp\frac{2m}{\ell(\ell+1)}\left({}_{\pm2}Y_{\ell m}\right) + \sqrt{\frac{(\ell^2-m^2)(\ell^2-4)}{\ell^2(2\ell+1)(2\ell-1)}}\left({}_{\pm2}Y_{\ell-1,m}\right)\right]\left(\mathcal{E}_\ell^{(m)}\pm i\mathcal{B}_\ell^{(m)}\right)\,. \tag{5.120}$$

To obtain the scattering term we integrate $P_m(\mathbf{n},\mathbf{n}')\mathcal{V}(\mathbf{n}')$ given in Eq. (5.27) over the $\mathbf{n}'$-sphere. For this we use the mode expansion

$$\begin{aligned}\mathcal{V}(\mathbf{k},\mathbf{n}') &= \begin{pmatrix} \mathcal{M} \\ \mathcal{Q}+i\mathcal{U} \\ \mathcal{Q}-i\mathcal{U} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{\ell=0}^{\infty}\sum_{m=-2}^{2}\mathcal{M}_\ell^{(m)}(\mathbf{k})\,{}_0G_{\ell m}(\mathbf{n}') \\ \sum_{\ell=2}^{\infty}\sum_{m=-2}^{2}\left(\mathcal{E}_\ell^{(m)}(\mathbf{k})+i\mathcal{B}_\ell^{(m)}(\mathbf{k})\right){}_2G_{\ell m}(\mathbf{x},\mathbf{n}') \\ \sum_{\ell=2}^{\infty}\sum_{m=-2}^{2}\left(\mathcal{E}_\ell^{(m)}(\mathbf{k})-i\mathcal{B}_\ell^{(m)}(\mathbf{k})\right){}_{-2}G_{\ell m}(\mathbf{x},\mathbf{n}') \end{pmatrix}. \end{aligned} \tag{5.121}$$

Using the orthogonality relation $\int d\Omega_{\mathbf{n}'}\,{}_sY_{\ell m}(\mathbf{n})\,{}_sY_{\ell' m'}(\mathbf{n}) = \delta_{\ell\ell'}\,\delta_{mm'}$ we obtain

$$\int d\Omega_{\mathbf{n}'}\,P_m(\mathbf{n},\mathbf{n}')\mathcal{V}(\mathbf{n}') = \begin{pmatrix} \mathcal{M}_2^{(m)}(\mathbf{k})\,{}_0G_{2m}(\mathbf{n}) - \sqrt{6}\mathcal{E}_2^{(m)}(\mathbf{k})\,{}_0G_{2m}(\mathbf{n}) \\ -\sqrt{6}\mathcal{M}_2^{(m)}(\mathbf{k})\,{}_2G_{2m}(\mathbf{n}) + 3\mathcal{E}_2^{(m)}(\mathbf{k})\,{}_2G_{2m}(\mathbf{n}) \\ -\sqrt{6}\mathcal{M}_2^{(m)}(\mathbf{k})\,{}_{-2}G_{2m}(\mathbf{n}) + 3\mathcal{E}_2^{(m)}(\mathbf{k})\,{}_{-2}G_{2m}(\mathbf{n}) \end{pmatrix}. \tag{5.122}$$

Hence, Thomson scattering does not depend on B-mode polarization. Finally, we also need the gravitational scalar, vector and tensor terms which enter the Boltzmann equation for the temperature anisotropy. They do not directly couple to the polarization since there is no zeroth-order polarization. We obtain exactly the same terms as in Chapter 4 which we now write in terms of the basis functions $Y_{\ell m}$. We get

$$i\mu k(\Psi+\Phi) = ik\sqrt{\frac{4\pi}{3}}(\Psi+\Phi)Y_{10} = -i\sqrt{\frac{4\pi}{3}}Y_{10}S_1^{(0)}\,, \tag{5.123}$$

$$\begin{aligned} -\frac{ik}{\sqrt{2}}\mu\left[(\mathbf{n}\cdot\mathbf{e}_+)\sigma_+ + (\mathbf{n}\cdot\mathbf{e}_-)\sigma_-\right] &= ik\sqrt{\frac{4\pi}{15}}\left[\sigma_+Y_{21}+\sigma_-Y_{2-1}\right] \\ &= -\sqrt{\frac{4\pi}{5}}\left[Y_{21}S_2^{(1)}+Y_{2-1}S_2^{(-1)}\right], \end{aligned} \tag{5.124}$$

$$\begin{aligned} &-(1-\mu^2)\left[\dot{H}_+\cos(2\varphi)+\dot{H}_\times\sin(2\varphi)\right] \\ &\quad= -\frac{1}{2}(1-\mu^2)\left[(\dot{H}_+ - i\dot{H}_\times)e^{2i\varphi} + (\dot{H}_+ + i\dot{H}_\times)e^{-2i\varphi}\right] \\ &\quad= -\sqrt{\frac{4\pi}{15}}\left[\dot{H}_2Y_{22}+\dot{H}_{-2}Y_{2-2}\right] = -\sqrt{\frac{4\pi}{5}}\left[Y_{22}S_2^{(2)}+Y_{2-2}S_2^{(-2)}\right], \end{aligned} \tag{5.125}$$

where we have set $H_{\pm2} = \sqrt{2}(H_+ \pm iH_\times)$. The source terms $S_\ell^{(m)}$ are defined by these equations. In addition, we must take into account the Doppler term which is of the form $i\mu kV^{(b)} = ik\sqrt{4\pi/3}\left[V_b^{(0)}Y_{10} + V_b^{(1)}Y_{11} + V_b^{(-1)}Y_{1-1}\right]$. Here, $V_b^{(0)}$

denotes the scalar part of the baryon velocity field and $V_b^{(\pm1)}$ are the vector perturbations with helicity ±1. With all this and taking care of the normalization of the mode function ${}_0G_{\ell m}$, the Boltzmann equation for the temperature anisotropies, $(\partial_t + \mathbf{n}\cdot\nabla)\mathcal{M} = S + \dot\kappa C[\mathcal{M}]$ turns into the mode equations

$$\dot{\mathcal{M}}_\ell^{(m)} + k\left[\frac{\sqrt{(\ell+1)^2-m^2}}{(2\ell+3)}\mathcal{M}_{\ell+1}^{(m)} - \frac{\sqrt{\ell^2-m^2}}{(2\ell+1)}\mathcal{M}_{\ell-1}^{(m)}\right] = S_\ell^{(m)} + \dot\kappa\left[P_\ell^{(m)} - \mathcal{M}_\ell^{(m)}\right]. \tag{5.126}$$

with

$$S_\ell^{(0)} = -\delta_{\ell1}k(\Psi+\Phi)\,, \tag{5.127}$$

$$S_\ell^{(\pm1)} = -\delta_{\ell2}\frac{i}{\sqrt{3}}k\sigma_\pm\,, \tag{5.128}$$

$$S_\ell^{(\pm2)} = \delta_{\ell2}\frac{1}{\sqrt{3}}k\dot H_{\pm2}\,, \tag{5.129}$$

$$P_\ell^{(0)} = \delta_{\ell0}\mathcal{M}_0^{(0)} + V_b^{(0)}\delta_{\ell1} + \delta_{\ell2}\frac{1}{10}\left[\mathcal{M}_2^{(0)} - \sqrt{6}\mathcal{E}_2^{(0)}\right], \tag{5.130}$$

$$P_\ell^{(\pm1)} = V_b^{(\pm1)}\delta_{\ell1} + \delta_{\ell2}\frac{1}{10}\left[\mathcal{M}_2^{(\pm1)} - \sqrt{6}\mathcal{E}_2^{(\pm1)}\right], \tag{5.131}$$

$$P_\ell^{(\pm2)} = \delta_{\ell2}\frac{1}{10}\left[\mathcal{M}_2^{(\pm2)} - \sqrt{6}\mathcal{E}_2^{(\pm2)}\right]. \tag{5.132}$$

For the left-hand side of Eq. (5.126) we used Eq. (5.119) and have isolated terms proportional to $Y_{\ell m}$ in the expansion (5.33) for fixed $\mathbf{k}$. The terms $P_\ell^{(m)}$ come from the collision integral (5.122). Apart from the E-polarization contribution they agree with the result found in Chapter 4.

For the evolution of the polarizations, we only need to take into account free streaming and the collision term. As mentioned above, the coupling of polarization to gravity is a second-order effect which is taken into account when discussing lensing in Chapter 7. Isolating terms proportional to ${}_{\pm2}Y_{\ell m}$, in Eq. (5.118), taking the sum and the difference, ${}_2Y_{\ell m} \pm {}_{-2}Y_{\ell m}$, leads to the left-hand side of the Boltzmann equation for E- and B-mode polarization. The right-hand side is obtained from (5.122). Putting it all together we find

$$\dot{\mathcal{E}}_\ell^{(m)} + k\left[\frac{\sqrt{[(\ell+1)^2-4][(\ell+1)^2-m^2]}}{(2\ell+1)(2\ell+3)}\mathcal{E}_{\ell+1}^{(m)} - \frac{2m}{\ell(\ell+1)}\mathcal{B}_\ell^{(m)} - \frac{\sqrt{(\ell^2-4)(\ell^2-m^2)}}{(2\ell+1)(2\ell+3)}\mathcal{E}_{\ell-1}^{(m)}\right] = -\dot\kappa\left[\mathcal{E}_\ell^{(m)} + \sqrt{6}P_\ell^{(m)}\right], \tag{5.133}$$

$$\mathcal{B}_\ell^{(m)} + k\left[\frac{\sqrt{[(\ell+1)^2-4][(\ell+1)^2-m^2]}}{(2\ell+1)(2\ell+3)}\mathcal{B}_{\ell+1}^{(m)} + \frac{2m}{\ell(\ell+1)}\mathcal{E}_\ell^{(m)} - \frac{\sqrt{(\ell^2-4)(\ell^2-m^2)}}{(2\ell+1)(2\ell+3)}\mathcal{B}_{\ell-1}^{(m)}\right] = -\dot{\kappa}\mathcal{B}_\ell^{(m)} \,. \tag{5.134}$$

Eqs. (5.126)–(5.134) represent the full Boltzmann hierarchy which has to be truncated at some value $\ell_{\max}$ and can then be solved, using the relations given in Section 4.7 which determine the gravitational source terms. As in Chapter 4 it is numerically very costly to solve the hierarchy until some large value $\ell_{\max} \sim 2000$, which determines the fluctuations on angular scales larger than about 5 arc minutes. One therefore solves it only up to $\ell \sim 10$ and uses this result to determine the source terms which depend only on the multipoles $\ell = 0,\ 1$ and 2. The higher multipoles are then again calculated with the help of an integral solution which, for a given source term is obtained by simple quadrature. We now derive this integral solution.

5.5.1 Integral solution

To find the integral solution, we prefer, as in Chapter 4, to consider the sums of the harmonic expansions. We define for $m = 0,\ 1$ and 2 (scalar, vector and tensor perturbations)

$$\mathcal{M}^{(m)}(t,\mathbf{n},\mathbf{k}) = \sum_\ell \mathcal{M}_\ell^{(m)}(t,\mathbf{k})(-i)^\ell\sqrt{\frac{4\pi}{2\ell+1}}Y_{\ell m} \,, \tag{5.135}$$

$$\mathcal{E}^{(m)}(t,\mathbf{n},\mathbf{k}) + i\mathcal{B}^{(m)}(t,\mathbf{n},\mathbf{k}) = \sum_\ell (\mathcal{E}_\ell^{(m)}(t,\mathbf{k}) + i\mathcal{B}_\ell^{(m)}(t,\mathbf{k}))(-i)^\ell\sqrt{\frac{4\pi}{2\ell+1}}{}_2Y_{\ell m} \,, \tag{5.136}$$

$$\mathcal{E}^{(m)}(t,\mathbf{n},\mathbf{k}) - i\mathcal{B}^{(m)}(t,\mathbf{n},\mathbf{k}) = \sum_\ell (\mathcal{E}_\ell^{(m)}(t,\mathbf{k}) - i\mathcal{B}_\ell^{(m)}(t,\mathbf{k}))(-i)^\ell\sqrt{\frac{4\pi}{2\ell+1}}{}_{-2}Y_{\ell m} \,. \tag{5.137}$$

For each of these variables, the Boltzmann equation is of the form

$$(\partial_t - i\mu k + \dot{\kappa})X = S \,, \tag{5.138}$$

For $X = \mathcal{M}^{(m)}$ the source term is

$$S^{(m)}(t,\mathbf{n}) = \sum_{\ell=0}^{2}\left(S_\ell^{(m)} + \dot{\kappa}P_\ell^{(m)}\right)(-i)^\ell\sqrt{\frac{4\pi}{2\ell+1}}Y_{\ell m} \,, \tag{5.139}$$

while for $X = \mathcal{E}^{(m)} \pm \mathcal{B}^{(m)}$

$$S^{(m)}(t,\mathbf{n}) = -\sqrt{6}\frac{\dot\kappa}{10}\left(\mathcal{M}_2^{(m)} - \sqrt{6}E_2^{(m)}\right)\sqrt{\frac{4\pi}{5}}\,{}_{\pm2}Y_{2m}\,. \tag{5.140}$$

The general solution to Eq. (5.138) with initial condition $X(t_{\rm in})$ is simply

$$X(t) = X(t_{\rm in})\,e^{-ik\mu(t-t_{\rm in})-\kappa(t,t_{\rm in})} + \int_{t_{\rm in}}^{t} dt' e^{-ik\mu(t-t')-\kappa(t,t')}\dot\kappa(t')S(t')\,, \tag{5.141}$$

where $\kappa(t,t_1) = \int_{t_1}^{t} dt'\,\dot\kappa(t')$ and $e^{\kappa(t_0,t')}\dot\kappa(t') = g(t')$ is the visibility function defined in Chapter 4 which is strongly peaked at the last scattering surface, where the collision terms induce the higher moments and polarization due to scattering. At much earlier times the photons behave like a perfect fluid and at much later times, collisions are very rare and all the evolution is determined by free streaming.

We are interested in the solution at t_0 and choose the initial time early enough so that we can neglect the initial value for all modes we are interested in. We then obtain the integral solution

$$\begin{aligned}&\mathcal{M}^{(m)}(t_0,\mathbf{n})\\ &\quad= \sum_{\ell=0}^{2}(-i)^\ell\sqrt{\frac{4\pi}{2\ell+1}}\int_{t_{\rm in}}^{t_0} dt\left(S_\ell^{(m)}(t) + \dot\kappa\,P_\ell^{(m)}(t)\right)Y_{\ell m}e^{-ik\mu(t_0-t)-\kappa(t_0,t)}\,,\end{aligned} \tag{5.142}$$

$$\begin{aligned}&\mathcal{E}^{(m)}(t,\mathbf{n}) \pm i\mathcal{B}^{(m)}(t,\mathbf{n})\\ &\quad= -\sqrt{\frac{24\pi}{5}}\int_{t_{\rm in}}^{t_0} dt\frac{\dot\kappa}{10}\left(\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)}\right){}_{\pm2}Y_{2m}\,e^{-ik\mu(t_0-t)-\kappa(t_0,t)}\,.\end{aligned} \tag{5.143}$$

To expand this solution in spherical harmonics we use that

$$(-i)^\ell\sqrt{\frac{4\pi}{2\ell+1}}\,{}_sY_{\ell m}(\mathbf{n})e^{-ik\mu(t_0-t)} = {}_sG_{\ell m}(-\mathbf{n}(t_0-t),\mathbf{n})\,.$$

Furthermore, the total angular momentum expansion of ${}_sG_{\ell m}$ gives, see Eqs. (5.44) and (5.45),

$${}_0G_{\ell m}(-\mathbf{n}r,\mathbf{n}) = \sum_L (-i)^L\sqrt{4\pi(2L+1)}\alpha_L^{(\ell m)}(kr)Y_{Lm}\,, \tag{5.144}$$

as well as

$${}_{\pm2}G_{\ell m}(-\mathbf{n}r,\mathbf{n}) = \sum_L (-i)^L\sqrt{4\pi(2L+1)}\left(\epsilon_L^{(\ell m)}(kr) \pm i\beta_L^{(\ell m)}(kr)\right)Y_{Lm}(\mathbf{n})\,. \tag{5.145}$$

Introducing Eq. (5.144) in Eq. (5.142) and making use of the explicit form of the

source term given in Eqs. (5.127)–(5.132) yields with $x \equiv k(t_0 - t)$

$$\mathcal{M}^{(0)}(t_0, \mathbf{n}) = \sum_\ell (-i)^\ell \sqrt{4\pi(2\ell+1)} Y_{\ell m}(\mathbf{n}) \int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)} \times \left[ik(\Psi + \Phi + \dot\kappa V^{(b)}) \alpha_\ell^{(10)}(x) + \dot\kappa \left(\mathcal{M}_0^{(0)} \alpha_\ell^{(00)}(x) + \frac{1}{10}\left[\mathcal{M}_2^{(0)} - \sqrt{6}\mathcal{E}_2^{(0)}\right] \alpha_\ell^{(20)}(x) \right) \right], \tag{5.146}$$

$$\frac{\mathcal{M}_\ell^{(0)}(t_0)}{2\ell+1} = \int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)} \left[ik(\Psi + \Phi + \dot\kappa V^{(b)}) \alpha_\ell^{(10)}(x) + \dot\kappa \left(\mathcal{M}_0^{(0)} \alpha_\ell^{(00)}(x) + \frac{1}{10}\left[\mathcal{M}_2^{(0)} - \sqrt{6}\mathcal{E}_2^{(0)}\right] \alpha_\ell^{(20)}(x) \right) \right], \tag{5.147}$$

$$\mathcal{M}^{(\pm1)}(t_0, \mathbf{n}) = \sum_\ell (-i)^\ell \sqrt{4\pi(2\ell+1)} Y_{\ell m}(\mathbf{n}) \int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)} \times \left[-\frac{ik}{\sqrt{3}} \sigma_\pm \alpha_\ell^{(2\,\pm1)}(x) + \dot\kappa V_\pm^{(b)} \alpha_\ell^{(1\,\pm1)}(x) + \frac{\dot\kappa}{10}\left[\mathcal{M}_2^{(\pm1)} - \sqrt{6}\mathcal{E}_2^{(\pm1)}\right] \alpha_\ell^{(2\,\pm1)}(x) \right], \tag{5.148}$$

$$\frac{\mathcal{M}_\ell^{(\pm1)}(t_0)}{2\ell+1} = \int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)} \left[-\frac{ik}{\sqrt{3}} \sigma_\pm \alpha_\ell^{(2\,\pm1)}(x) + \dot\kappa V_\pm^{(b)}\, \alpha_\ell^{(1\,\pm1)}(x) + \frac{\dot\kappa}{10}\left[\mathcal{M}_2^{(\pm1)} - \sqrt{6}\mathcal{E}_2^{(\pm1)}\right] \alpha_\ell^{(2\,\pm1)}(x) \right], \tag{5.149}$$

$$\mathcal{M}^{(\pm2)}(t_0, \mathbf{n}) = \sum_\ell (-i)^\ell \sqrt{4\pi(2\ell+1)} Y_{\ell m}(\mathbf{n}) \int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)} \times \left[\frac{1}{\sqrt{3}} k \dot H_{\pm2} + \frac{\dot\kappa}{10}\left[\mathcal{M}_2^{(\pm2)} - \sqrt{6}\mathcal{E}_2^{(\pm2)}\right] \right] \alpha_\ell^{(2\,\pm2)}(x), \tag{5.150}$$

$$\frac{\mathcal{M}_\ell^{(\pm2)}(t_0)}{2\ell+1} = \int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)} \left[\frac{1}{\sqrt{3}} k \dot H_{\pm2} + \frac{\dot\kappa}{10}\left[\mathcal{M}_2^{(\pm2)} - \sqrt{6}\mathcal{E}_2^{(\pm2)}\right] \right] \alpha_\ell^{2\,\pm2}(x)\,. \tag{5.151}$$

Equivalently, introducing Eq. (5.145) in Eq. (5.143) we obtain

$$\mathcal{E}^{(m)}(t_0, \mathbf{n}) \pm i\mathcal{B}^{(m)}(t_0, \mathbf{n}) = -\sqrt{6} \sum_\ell (-i)^\ell \sqrt{4\pi(2\ell+1)}\, {}_{\pm2}Y_{\ell m}(\mathbf{n}) \times \int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)} \frac{\dot\kappa}{10} \left(\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)}\right) \left(\epsilon_\ell^{(2m)}(x) \pm i\beta_\ell^{(2m)}(x)\right), \tag{5.152}$$

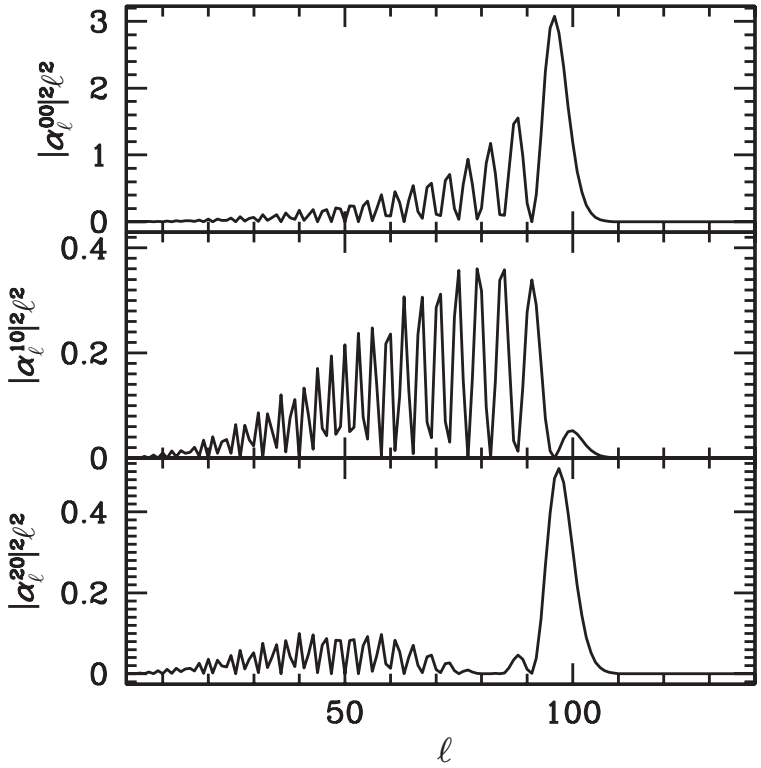


Fig. 5.4. The functions $\ell^2|\alpha_\ell^{(00)}|^2$ (top), $\ell^2|\alpha_\ell^{(10)}|^2$ (middle) and $\ell^2|\alpha_\ell^{(20)}|^2$ (bottom) are shown as function of ℓ for fixed $x = 100$. These are the kernels relevant for the scalar temperature anisotropies. Their amplitude and shape determine how strongly the corresponding source terms influence the final anisotropy spectrum.

$$\frac{\mathcal{E}_\ell^{(m)}(t_0,\mathbf{n}) \pm i\mathcal{B}_\ell^{(m)}(t_0,\mathbf{n})}{2\ell+1} = -\sqrt{6}\int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)}\frac{\dot\kappa}{10}(\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)})\left(\epsilon_\ell^{(2m)}(x) \pm i\beta_\ell^{(2m)}(x)\right). \quad (5.153)$$

Taking the sum and the difference of the last equation we obtain

$$\frac{\mathcal{E}_\ell^{(m)}(t_0,\mathbf{n})}{2\ell+1} = -\sqrt{6}\int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)}\frac{\dot\kappa}{10}(\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)})\epsilon_\ell^{(2m)}(x)\,, \quad (5.154)$$

$$\frac{\mathcal{B}_\ell^{(m)}(t_0,\mathbf{n})}{2\ell+1} = -\sqrt{6}\int_{t_{\rm in}}^{t_0} dt\, e^{-\kappa(t_0,t)}\frac{\dot\kappa}{10}(\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)})\beta_\ell^{(2m)}(x)\,. \quad (5.155)$$

The fact that the scalar B-mode, $\mathcal{B}_\ell^{(0)}$ vanishes, is now a consequence of $\beta^{(20)} = 0$.

To have some insight into the kernels $\alpha_\ell^{(ij)}$, $\epsilon_\ell^{(ij)}$ and $\beta_\ell^{(ij)}$, we plot them in Figs. 5.4–5.7 as functions of ℓ for fixed $x = k(t_0 - t) = 100$. They are all peaked

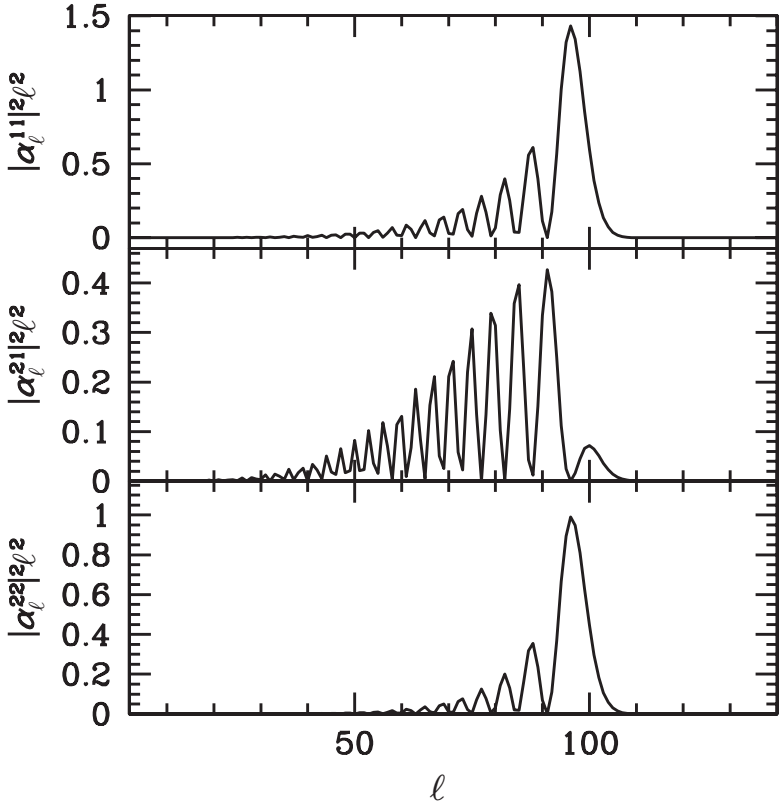


Fig. 5.5. The functions $\ell^2|\alpha_\ell^{(11)}|^2$ (top), $\ell^2|\alpha_\ell^{(21)}|^2$ (middle) and $\ell^2|\alpha_\ell^{(22)}|^2$ (bottom) are shown as functions of ℓ for fixed $x = 100$. These are the kernels relevant for vector, $\alpha_\ell^{(11)}$ and $\alpha_\ell^{(21)}$, and tensor, $\alpha_\ell^{(22)}$, temperature anisotropies.

at $\ell \simeq x$. For temperature anisotropies this peak is strongest for the tensor kernel $\alpha^{(22)}$. The kernel which dominates scalar temperature anisotropies by a factor of nearly 10 is $\alpha^{(00)}$, which comes from the free streaming of density fluctuations on the last scattering surface and therefore is responsible for the acoustic peaks. The kernel $\alpha^{(10)}$ which multiplies the ordinary and integrated Sachs–Wolfe terms and the Doppler term is significantly lower and somewhat less strongly peaked. Finally, the kernel $\alpha^{(20)}$ which couples to polarization has a narrow peak at $\ell \simeq x$ and a lower, broader one around $\ell \simeq x/2$. The decay of all kernels for $\ell > x$ is very rapid.

The kernel $\alpha^{(21)}$ which couples vector perturbations to polarization and to the gravitational vector modes is smaller and less strongly peaked than $\alpha^{(11)}$ which couples to the vector-type Doppler term. Finally, tensor temperature anisotropies

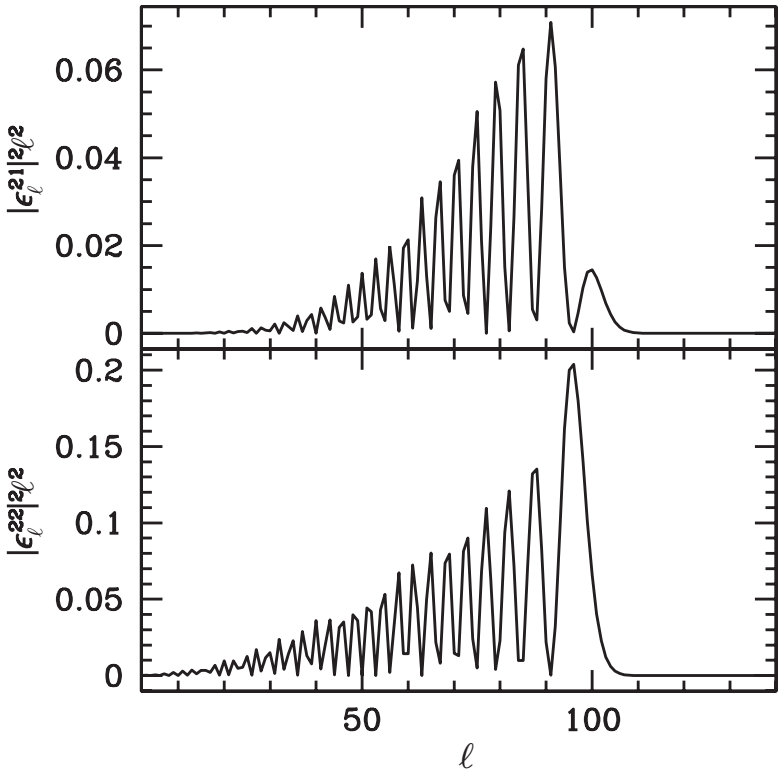


Fig. 5.6. The functions $\ell^2|\epsilon_\ell^{(21)}|^2$ (top) and $\ell^2|\epsilon_\ell^{(22)}|^2$ (bottom) are shown as functions of ℓ for fixed $x = 100$. These are the kernels relevant for E-polarization of vector and tensor modes respectively. Since $\ell^2|\epsilon_\ell^{(20)}|^2 = \ell^2|\alpha_\ell^{(22)}|^2$ this kernel for scalar E-polarization is not replotted. Note that the vector E-polarization kernel is very small and the scalar kernel is still about a factor of 5 larger than the tensor kernel.

have only one kernel, $\alpha^{(22)}$, for their coupling to both, the gravitational term and polarization.

Looking at the polarization kernels $\epsilon_\ell^{(ij)}$ and $\beta_\ell^{(ij)}$, it is interesting to note that the vector B-kernel, $\beta_\ell^{(21)}$ is nearly 8 times larger than the tensor one. For E-polarization, the situation is reversed. Hence, vector perturbations would be very effective in generating B-polarization, while tensor perturbations generate somewhat more E- than B-polarization. Summing up the relevant contributions one finds for $x = k(t_0 - t) \gg 1$,

$$\frac{\sum_\ell \ell^2|\beta_\ell^{(2\,m)}|^2}{\sum_\ell \ell^2|\epsilon_\ell^{(2\,m)}|^2} \simeq \begin{cases} 6 & \text{for} \quad m = \pm 1 \\ \frac{8}{13} & \text{for} \quad m = \pm 2\,. \end{cases} \tag{5.156}$$

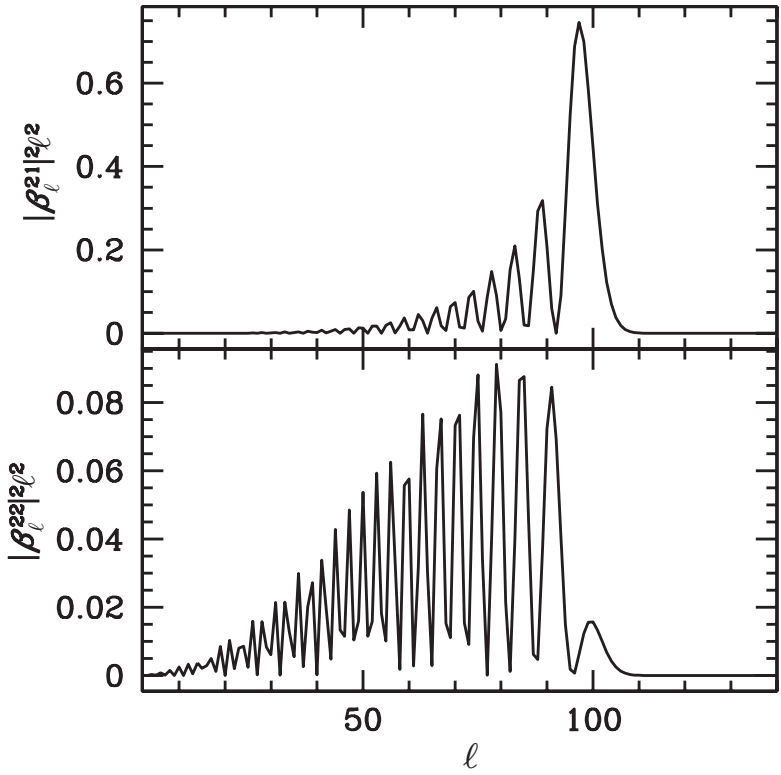


Fig. 5.7. The functions $\ell^2|\beta_\ell^{(21)}|^2$ (top) and $\ell^2|\beta_\ell^{(22)}|^2$ (bottom) are shown as functions of ℓ for fixed $x = 100$. These are the kernels relevant for B-polarization of vector and tensor modes respectively. Note that the vector B-polarization kernel is much larger than the tensor one. This is the opposite of what we find for E-polarization.

The scalar polarization kernel, $\epsilon^{(20)} = \alpha^{(22)}$ is the highest of all polarization kernels. As we have seen, scalar perturbations generate no B-polarization at all.

Exercises

Ex. 5.1 Relation to Chapter 4

Using the expressions for spherical harmonics from Appendix A4.2.3 and our definitions of $\mathcal{M}_\ell^{(T)\pm}$ and $\mathcal{M}_\ell^{(V)\pm}$ given in Eqs. (4.147) and (4.131) and (4.132) in Chapter 4 show that

$$\mathcal{M}_\ell^{(0)} = (2\ell+1)\mathcal{M}_\ell^{(5)}$$
$$\mathcal{M}_\ell^{(\pm1)} = i\sqrt{\ell(\ell+1)}\left[\mathcal{M}_{\ell+1}^{(V\pm)} + \mathcal{M}_{\ell-1}^{(V\pm)}\right], \tag{5.157}$$
$$\mathcal{M}_\ell^{(\pm2)} = \ldots\mathcal{M}^{(T\pm)}\,. \tag{5.158}$$

Hint: For Eq. (5.157) use the recurrence relation (A4.20) to relate the Legendre functions $P_{\ell 1}$ to Legendre polynomials.

Ex. 5.2 ***E*- and *B*-polarization**

For an electric field normal to $\mathbf{n}$ given by $E_i = \nabla_i f(\mathbf{n}) + \epsilon_{ij}\nabla_j g(\mathbf{n})$ determine the $\mathcal{Q}$ and $\mathcal{U}$ polarization. Here f and g are real functions on the sphere. Calculate $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{B}}$. Here ϵ_{ij} is the totally anti-symmetric tensor in two dimensions; $\epsilon_{ij} = \pm\det\gamma$ if $(i, j) = (1, 2)$ and $(i, j) = (2, 1)$ respectively and $\epsilon_{ij} = 0$ if two indices are equal. Here γ is the two-dimensional metric.

Ex. 5.3 ***E*- and *B*-polarization in real space**

Using $\ell_x + i\ell_y = \ell e^{i\varphi_\ell}$ derive the following relation between $\mathcal{E}$, $\mathcal{B}$ and $\mathcal{Q}, \mathcal{U}$ in real space:

$$\mathcal{E}(\mathbf{x}) = \nabla^{-2}(\partial_x^2 - \partial_y^2)\mathcal{Q}(\mathbf{x}) + \nabla^{-2}2\partial_x\partial_y\mathcal{U}(\mathbf{x})\,, \tag{5.159}$$

$$\mathcal{B}(\mathbf{x}) = \nabla^{-2}(\partial_x^2 - \partial_y^2)\mathcal{U}(\mathbf{x}) - \nabla^{-2}2\partial_x\partial_y\mathcal{Q}(\mathbf{x})\,. \tag{5.160}$$

6

Cosmological parameter estimation

6.1 Introduction

In the previous chapters we have calculated the CMB anisotropies and polarization. Generically the resulting spectrum shows a series of acoustic oscillations which present a snapshot of the CMB sky at the moment when photons decouple from electrons. The details of these spectra depend on the one hand on the initial fluctuations and on the other hand on the background cosmological parameters which determine the evolution of fluctuations.

If we make no hypothesis on the initial fluctuations, a given observed spectrum can be obtained by a nearly arbitrary choice of cosmological parameters. For a given initial power spectrum $P_m(k)$ of scalar ($m = 0$), vector ($m = \pm 1$) and tensor ($m = \pm 2$) perturbations, under the assumption of statistical homogeneity and isotropy, the resulting CMB power spectrum is generically of the form

$$C_\ell = \sum_{m=-2}^{2} \int dk\, T_m(\ell, k) P_m(k)\,, \tag{6.1}$$

where T_m is the CMB 'transfer function' which depends on the cosmological parameters. Therefore, for a nearly arbitrary transfer function $T_m(\ell, k)$ and arbitrary C_ℓs one can find initial power spectra P_m such that Eq. (6.1) holds. Since vector perturbations decay, the vector transfer function $T_{\pm 1}$ is very small. If the initial perturbations are sufficiently small for linear perturbation theory to hold, vector perturbations will not show up in the CMB spectrum. We shall therefore neglect them in our discussion.

'Sources' form an exception to this rule. A source is an inhomogeneous and anisotropic component of the energy–momentum tensor which is too small to contribute to the background, but which sources perturbations in all fluids. We shall discuss this case in Section 6.8. If sources are relevant, some parts of the perturbations

are generated at late time by a source term. In this case, vector perturbations can also be important.

We have to keep in mind that when determining cosmological parameters with the CMB, we are not really 'measuring' them, but we are 'estimating' them under certain, usually well motivated but very restrictive assumptions on the initial power spectrum. On the other hand, whenever we make a physical measurement we are using prior knowledge, e.g. that our apparatus obeys Maxwell's equation. However, the apparatus has usually been tested by some other measurements, while only the CMB and very few other data sets contain experimental information about the initial conditions of the fluctuations in the Universe. Therefore, in the 'ideal' world we might want to measure the cosmological parameters by other means and then with the well known transfer functions at hand, use the CMB to determine the initial fluctuations which help us to understand the physics of inflation, the physics at very high energies, probably close to the Planck scale, energies which are not available in any laboratory on Earth.

However, the real world is not ideal and since CMB fluctuations are the most accurate and theoretically the best understood dataset, we use them for both, to determine the parameters of the background cosmology and the initial fluctuations. For this to work, we have to assume that the initial power spectrum depends only on a few parameters, for example $k^3 P_0 = A_S (k/H_0)^{n_S - 1}$ and $P_{\pm 2} = 0$ with only two parameters, A_S and n_S. The better the data, the more parameters can be fitted.

Naively one might think that the knowledge of 1000 C_ℓs to a few per cent accuracy allows us to determine as many parameters with similar accuracy. But this is not so, since the CMB power spectrum can usually be well fitted with a function of only a few parameters. Also, it is only sensitive to a certain combination of cosmological parameters. This leads to *degeneracies* which we shall discuss in Section 6.6.

But before we enter the technical details of parameter estimation, we briefly want to discuss the physics of their influence on the CMB spectrum. This helps us to develop a good intuition for which parameters can be estimated with the CMB to high accuracy and which cannot.

6.2 The physics of parameter dependence

6.2.1 The acoustic peaks

The first acoustic peak corresponds to the comoving wavelength $\lambda_1 = \pi / k_1$ which has undergone exactly one compression since entering the horizon and whose fluctuations are at a maximum at decoupling. As we have discussed in Chapter 2, this scale is determined by $c_s k_1 t_{\rm dec} = \pi$ for adiabatic perturbations ($c_s k_1 t_{\rm dec} = 3\pi/2$ for iso-curvature perturbations). Subsequent peaks are at $c_s k_n t_{\rm dec} = n\pi$ for adiabatic

perturbations (and $c_s k_n t_{\rm dec} = (2n+1)\pi/2$ for iso-curvature perturbations). The angle onto which these peaks are projected in the sky is

$$\theta_n \simeq \frac{\lambda_n}{D_A(z_{\rm dec})}\,, \quad \ell_n \simeq \pi/\theta_n, \tag{6.2}$$

where $D_A(z)$ denotes the angular diameter distance to an event at time $t(z)$ and $\ell(\theta) = \pi/\theta$ is the harmonics, which corresponds roughly to the angle θ. The redshift of decoupling, $z_{\rm dec}$, and also $t_{\rm dec}$ depend mainly on the baryon density of the Universe, while the angular diameter distance depends strongly on curvature, but also on the cosmological constant and the dark matter density.

In a universe containing radiation, matter and a cosmological constant we have (see Chapter 1)

$$a_0 c_s t_{\rm dec} = \frac{1}{H_0}\int_{z_{\rm dec}}^{\infty} \frac{c_s(z)dz}{\left[\Omega_r(z+1)^4+\Omega_m(z+1)^3+\Omega_\Lambda+\Omega_K(z+1)^2\right]^{1/2}}\,, \tag{6.3}$$

$$D_A(z_{\rm dec}) = \frac{a_0}{z_{\rm dec}+1} \times \chi_K\left(\frac{1}{H_0 a_0}\int_0^{z_{\rm dec}} \frac{dz}{\left[\Omega_r(z+1)^4+\Omega_m(z+1)^3+\Omega_\Lambda+\Omega_K(z+1)^2\right]^{1/2}}\right), \tag{6.4}$$

$$\ell_n \simeq \frac{\pi D_A(z_{\rm dec})}{\lambda_n} = \frac{n\pi\,\chi_K\left(\frac{1}{H_0 a_0}\int_0^{z_{\rm dec}} \frac{dz}{\left[\Omega_r(z+1)^4+\Omega_m(z+1)^3+\Omega_\Lambda+\Omega_K(z+1)^2\right]^{1/2}}\right)}{\frac{1}{z_{\rm dec}+1}\frac{1}{H_0}\int_{z_{\rm dec}}^{\infty} \frac{c_s(z)dz}{\left[\Omega_r(z+1)^4+\Omega_m(z+1)^3+\Omega_\Lambda+\Omega_K(z+1)^2\right]^{1/2}}}\,. \tag{6.5}$$

Here

$$\chi_K(r) = \begin{cases} r & \text{if} \quad K=0 \\ \frac{1}{\sqrt{K}}\sin\left(\sqrt{K}r\right) & \text{if} \quad K>0 \\ \frac{1}{\sqrt{|K|}}\sinh\left(\sqrt{|K|}r\right) & \text{if} \quad K<0\,, \end{cases} \tag{6.6}$$

and we have used

$$\lambda_n = \frac{a_0}{n c_s t_{\rm dec}}\,.$$

It is evident, that, via χ_K, the position ℓ_n strongly depends on curvature. Since curvature and the cosmological constant are irrelevant at high redshift, the denominator of Eq. (6.5) is nearly independent of them. It depends mainly on Ω_r, Ω_m and on the baryon density via the sound speed c_s,

$$c_s^2 = \frac{1}{3}\frac{4\Omega_\gamma h^2}{4\Omega_\gamma h^2 + 3\Omega_b h^2}\,, \tag{6.7}$$

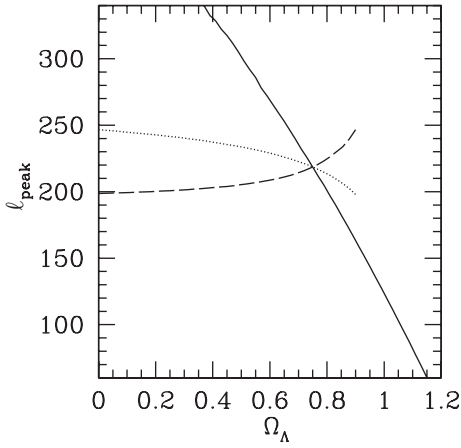


Fig. 6.1. We show the position of the first peak as a function of Ω_Λ. In the solid line we vary Ω_K, leaving all other parameters fixed, for the dashed line we vary Ω_Λ at fixed h, and for the dotted line we vary Ω_Λ at fixed $\Omega_m h^2$. The fixed parameters are $\Omega_K = 0$, $h = 0.72$, $\Omega_b h^2 = 0.022$, $\Omega_m = 0.25$, $n_s = 0.96$, $\tau_{ri} = 0.085$. Therefore, all the curves cross at $\Omega_\Lambda = 0.75$. Notice the strong dependence of $\ell_{\rm peak}$ on curvature.

where $h = H_0/100\,\mathrm{km\,s^{-1}Mpc^{-1}}$. Since $\Omega_\gamma h^2$ which is proportional to the present photon energy density, hence to T_0^4 is very well known, the sound speed provides a measure of $\Omega_b h^2$. Note that $\Omega_\gamma h^2$ is much better known than $\Omega_\gamma = 8\pi G a_{SB} T_0^4/3H_0^2$. The latter contains considerable uncertainty in the Hubble constant.

In Fig. 6.1 we show the dependence of ℓ_1 on different cosmological parameters. There one also sees that when varying a parameter, like e.g. Ω_Λ, the resulting spectrum strongly depends on what has been kept fixed during the variation. We can fix h but not Ω_m, since increasing Ω_Λ just increases the dimensionless angular diameter distance, $H_0 D_A$ due to the decrease in Ω_m, while $H_0\lambda_1$ is not affected. The opposite is true if we let h, but not $\Omega_m h^2$ vary with Ω_Λ. Then the peak position is reduced with growing Ω_Λ. This comes from the fact that $H_0\lambda_1$ now increases with increasing Hubble parameter due to the decrease in Ω_r. This effect more than compensates the increase in $H_0 D_A$ due to the decrease in Ω_m. Note that $\Omega_r h^2$, like $\Omega_\gamma h^2$, is determined by the CMB temperature T_0 and therefore always remains fixed. (We neglect neutrino masses.)

As we have seen in Chapter 2, dark matter fluctuations grow only logarithmically during the radiation dominated era, Once the Universe becomes matter dominated, they start growing like the scale factor. Therefore, the amplitude of the gravitational potential which is mainly determined by the dark matter density, depends on $\Omega_m h^2$. Especially, the ratio of the height of the first acoustic peak and the Sachs–Wolfe plateau is sensitive to this parameter.

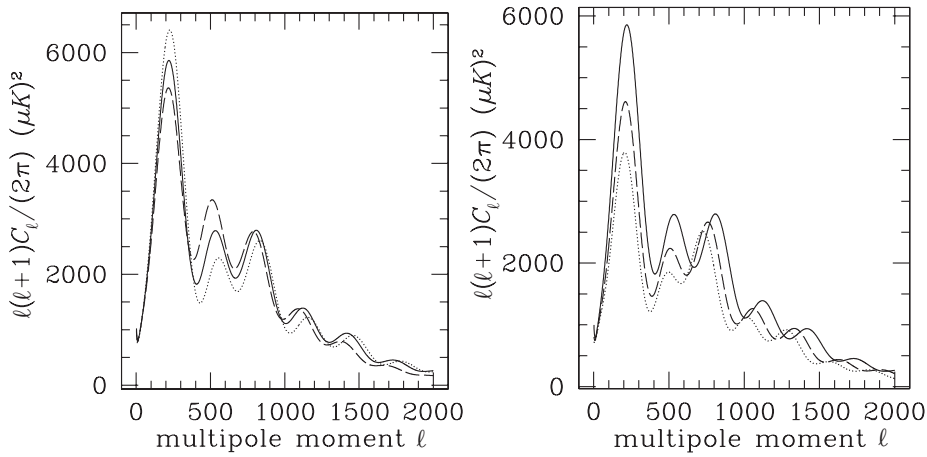


Fig. 6.2. In the left-hand panel we show the asymmetry of even and odd peaks and its dependence on $\Omega_b h^2$. The temperature anisotropy spectrum is plotted for $\Omega_b h^2 = 0.02$ (solid line), $\Omega_b h^2 = 0.03$ (dotted) and $\Omega_b h^2 = 0.01$ (dashed). On the right-hand side $\Omega_b h^2 = 0.02$ is fixed and three different values for the matter density are chosen, $\Omega_m h^2 = 0.12$ (solid), $\Omega_m h^2 = 0.2$ (dashed) and $\Omega_m h^2 = 0.3$ (dotted). Note that higher values of $\Omega_m h^2$ also lead to a stronger peak asymmetry. In addition, a smaller value of $\Omega_m h^2$ boosts the height especially of the first peak due to the stronger contribution from the early integrated Sachs–Wolfe effect. The peaks are also somewhat shifted since D_A depends on Ω_m.

The baryon density also enters CMB physics via the asymmetry of even and odd acoustic peaks. As we have said before, the first peak at scale ℓ_1 is a contraction peak, an over-density. Correspondingly, the second peak is an expansion peak, an under-density. If the oscillating fluid consists solely of massless photons it would undergo perfectly harmonic oscillations and the amplitudes of contraction and expansion peaks would be equal. However, the massive baryons re-enforce contraction via their self-gravity and their reaction to the gravitational potential of dark matter. Correspondingly they reduce expansion (see Fig. 6.2).

On small scales, the fluctuation amplitudes decay due to Silk damping. Again the strength of the damping depends on the baryon density, hence on $\Omega_b h^2$.

6.2.2 Neutrinos

As we saw in Chapter 1, neutrinos decouple when the Universe has a temperature of about $T_\nu \sim 1.4\,\mathrm{MeV}$, which corresponds to a redshift of $z_\nu \sim 0.6 \times 10^{10}$. After that, weak interactions are too weak to keep them in thermal equilibrium with the other constituents, electrons, baryons, photons and also cold dark matter. (If dark matter interactions with neutrinos were stronger than weak interactions, we would probably have detected dark matter in the laboratory.)

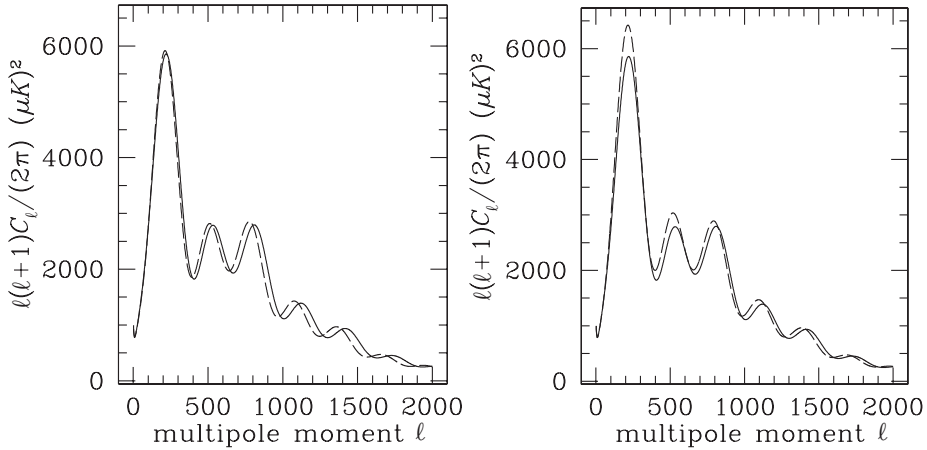


Fig. 6.3. We show the CMB power spectrum for massless neutrinos and neutrinos with mass $m_\nu = 2$ eV (dashed lines). In the left-hand panel $\Omega_{cdm}h^2 = 0.12$ is fixed while on the right-hand side $\Omega_m h^2 = 0.144$ is fixed. In all curves $\Omega_{\rm tot} = 1$, $\Omega_b h^2 = 0.022$ and $h = 0.7$. Keeping $\Omega_m h^2$ fixed, adding neutrinos acts a bit like a lower-matter density, since the neutrinos are not yet fully non-relativistic at decoupling.

After that, neutrinos propagate freely, described by the Liouville equation (see Section 4.7). As long as their masses can be neglected, they build up anisotropic stresses by free streaming and their energy density dilutes like that of radiation. As soon as their masses become relevant, their pressure and anisotropic stresses decay and they behave like dark matter. This changes their effects on CMB anisotropies and thereby leads to a way of measuring their mass with the CMB. The significance is not very high, since neutrino anisotropic stresses contribute only about 5% to the CMB fluctuations and since massive neutrinos have a similar signature like cold dark matter in the CMB. In Fig. 6.3 we compare the CMB spectrum for three sorts of degenerate neutrinos, all with mass $m_\nu = 2$ eV, with the spectrum of massless neutrinos, once by fixing the total matter density and once by fixing the cold dark matter density.

6.2.3 Gravitational waves

The CMB anisotropies from gravitational waves are significant on scales that are super-Hubble before decoupling, which corresponds to $\ell \lesssim 80$. On smaller scales they decay. If it is small, such a signal is difficult to disentangle from a slightly red ($n_s < 1$) spectrum of scalar fluctuations which simply has somewhat more power on the Sachs–Wolfe plateau than a Harrison–Zel'dovich spectrum with $n_s = 1$.

However, as we saw in Chapter 5, scalar perturbations do not generate B-mode polarization. Therefore, the detection of B-mode polarization would be a finger

print of gravitational waves from inflation. In Chapter 7 we will, however, see that second-order effects, especially lensing can generate a B-mode signal from scalar perturbations.

6.3 Reionization

Even though we have overwhelming evidence that the Universe recombined and became neutral at a redshift of about $z_{\rm rec} \simeq 1400$, no considerable fraction of neutral hydrogen can be found in the intergalactic medium. At present the intergalactic gas is fully reionized.

This conclusion is drawn from the absence of the so-called 'Gunn–Peterson trough' in quasar spectra. Quasars (or 'quasi-stellar objects') are very active galactic centres which are so luminous that they can be observed up to redshifts close to $z \simeq 7$.

Gunn and Peterson (1965) calculated that even a modest density of neutral hydrogen would lead to a significant absorption trough in the part of the quasar spectra which is bluer than Lyman-α at emission and redder at absorption. These photons have, at some moment during their propagation from the quasar to us, exactly Lyman-α frequency and are then resonantly absorbed by neutral hydrogen. Inserting numbers one finds (Peacock, 1999) that the neutral hydrogen density in the intergalactic medium amounts to less than $\Omega_H h \lesssim 10^{-8}$.

There are, however, so-called Lyman-α clouds, i.e. clouds of neutral hydrogen which intervene the lines of sight of quasars and lead to a 'forest' of absorption lines in quasars, the Lyman-α forest, see Section 6.7.5. But even integrating the total optical depth of the Lyman-α forest one infers a neutral hydrogen density of only $\Omega_H h \simeq 10^{-5}$. It is very unlikely that galaxy formation has been so efficient as to sweep up 99.9% of all the hydrogen in the Universe. We are therefore led to the conclusion that the present intergalactic hydrogen is ionized.

In recent years the Lyman-α and Lyman-β troughs have been found in very high-redshift quasars with $z > 6$. This confirms that at these high redshifts, some neutral hydrogen has been present which has been reionized later on, probably by UV-light from the first burst or star formation.

Reionization was terminated at redshift $z_{\rm ri} \simeq 6$, but it is not clear when the process started and whether it was very fast or slow. The unknown reionization history of the Universe affects CMB anisotropies and polarization. Once the Universe is reionized, CMB photons can, in principle scatter again with the free electrons. Since the electron density is significantly lower than at decoupling, the scattering probability or optical depth is rather low and the effect is probably on the level of 5–10%.

Rescattering of electrons leads to additional polarization on a scale which corresponds to the sound horizon at reionization, $\lambda_{\rm ri} \simeq c_s(z_{\rm ri})t(z_{\rm ri})$. In addition, due to

the much larger free-streaming scale of electrons after reionization, there is some additional Silk damping on scales up to $\lambda_{\rm ri}$.

Usually, reionization is parametrized simply by a reionization redshift $z_{\rm ri}$ or by the optical depth $\tau_{\rm ri}$ to reionization,

$$\tau_{\rm ri}(z_{\rm ri}) = \sigma_T \int_{t(z_{\rm ri})}^{t_0} a(t) n_e(t)\, dt \ . \tag{6.8}$$

Of course, the reionization history can be more complicated than this, e.g. it does make a difference whether reionization is instantaneous or slow. However, present data are not sufficient to determine more than the optical depth.

6.4 CMB data

So far, we have mainly discussed the theoretical aspects of the CMB. Of course these are very interesting, mainly since we have good quality data to compare with our theoretical models. On the other hand, good quality CMB data are so valuable, since, for a given cosmological model and specified initial fluctuations, we can calculate the CMB anisotropies and polarization with high accuracy.

This interplay of theory and data, which makes physics so fascinating, works in a most beautiful way in CMB physics: observing the largest structures in the Universe, the anisotropy patterns in the CMB, we learn not only a lot about the parameters of the Universe but also about physics at the highest energies corresponding to the smallest scales. The largest pattern in the cosmos turns out to be an imprint of quantum physics!

In the previous chapters we learned how to calculate the CMB anisotropy and polarization spectrum for a given model. A theoretical model does not predict the CMB anisotropy or polarization amplitude in a given position (θ_0, φ_0) in the sky. However, this is what an experiment measures.[1]

Let us assume that we are given a temperature fluctuation map $\Delta T_s(\mathbf{n}) = T_s(\mathbf{n}) - \bar{T}_s$ from an experiment. Here $\bar{T}_s$ is the mean temperature and the suffix $_s$ stands for 'signal'. The correlation function $\langle \Delta T_s(\mathbf{n}_1) \Delta T_s(\mathbf{n}_2) \rangle$ is a measure for the mean temperature difference,

$$\langle (T_s(\mathbf{n}_1) - T_s(\mathbf{n}_2))^2 \rangle = 2 \left(\langle \Delta T_s^2 \rangle - \langle \Delta T_s(\mathbf{n}_1) \Delta T_s(\mathbf{n}_2) \rangle \right) . \tag{6.9}$$

When we put brackets around observed quantities like $\langle \Delta T_s(\mathbf{n}_1) \Delta T_s(\mathbf{n}_2) \rangle$, we understand an averaging over directions $\mathbf{n}_1$ and $\mathbf{n}_2$ with fixed opening angle $\cos\theta = \mathbf{n}_1 \cdot \mathbf{n}_2$. In this paragraph we simply equate such averages to theoretical

[1] The experiment actually measures voltage differences as a function of time. We shall not enter into the rather involved process of how an optimal map $T(\theta, \varphi)$ is obtained from these time ordered data streams (for an introduction, see Dodelson (2003)).

ensemble averages. This is an implicit assumption of statistical isotropy. In the next paragraph we shall discuss additional limitations of this procedure which go under the name of 'cosmic variance'.

The measured temperature $T_s(\mathbf{n})$ is obtained from the true sky temperature by convolution with a beam profile $B(\mathbf{n}, \mathbf{n}')$ centred at $\mathbf{n}$,

$$T_s(\mathbf{n}) = \int B(\mathbf{n}, \mathbf{n}')T(\mathbf{n}')d\Omega' \,. \tag{6.10}$$

We can relate the correlation function of the measured temperature fluctuations between two directions $\mathbf{n}_1$ and $\mathbf{n}_2$ to the power spectrum by

$$\begin{aligned}\left\langle \frac{\Delta T_s(\mathbf{n}_1)\Delta T_s(\mathbf{n}_2)}{\bar{T}^2} \right\rangle &= \frac{1}{\bar{T}^2} \int B(\mathbf{n}_1, \mathbf{n}_1')B(\mathbf{n}_2, \mathbf{n}_2')\langle \Delta T(\mathbf{n}_1')T(\mathbf{n}_2')\rangle \, d\Omega_1' \, d\Omega_2' \\ &= \sum_{\ell,m,\ell',m'} \langle a_{\ell m} a^*_{\ell' m'} \rangle \int B(\mathbf{n}_1, \mathbf{n}_1')B(\mathbf{n}_2, \mathbf{n}_2')Y_{\ell m}(\mathbf{n}_1')Y^*_{\ell' m'}(\mathbf{n}_2') \, d\Omega_1' \, d\Omega_2' \\ &= \sum_{\ell} \frac{2\ell+1}{4\pi} C_\ell W_\ell(\mathbf{n}_1, \mathbf{n}_2) \,, \end{aligned} \tag{6.11}$$

where we have inserted $\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \, \delta_{\ell\ell'} \, \delta_{mm'}$. We then made use of the addition theorem for spherical harmonics, and we have defined the window function $W_\ell(\mathbf{n}_1, \mathbf{n}_2)$,

$$W_\ell(\mathbf{n}_1, \mathbf{n}_2) = \int B(\mathbf{n}_1, \mathbf{n}_1')B(\mathbf{n}_2, \mathbf{n}_2')P_\ell(\mathbf{n}_1' \cdot \mathbf{n}_2') \, d\Omega_1' \, d\Omega_2' \,. \tag{6.12}$$

Beam patterns are usually translation invariant so that $B(\mathbf{n}, \mathbf{n}')$ only depends on the angle between $\mathbf{n}$ and $\mathbf{n}'$ and the window function only depends on $\cos\theta = \mathbf{n}_1 \cdot \mathbf{n}_2$. Also the mean temperature difference and $\langle T_s(\mathbf{n}_1)T_s(\mathbf{n}_2)\rangle$ only depend on θ. This simply reflects statistical isotropy. Actually, since we determine the expectation value $\langle \bullet \rangle$ by averaging over directions which include the same angle, we obtain a result which only depends on this angle by construction.

6.4.1 Example window functions

As an illustration we calculate the window function for two examples of beam patterns. For this we observe that the beam is usually very narrow, a few degrees or less, so that, for the beam pattern $B(\mathbf{n}, \mathbf{n}')$ we can approximate the sphere by a plane orthogonal to $\mathbf{n}$ in the regime where the beam is non-vanishing. The planar vectors $\mathbf{x}_i$ then correspond to the angle between $\mathbf{n}$ and $\mathbf{n}_i'$. Setting $\mathbf{n}_i' = \frac{\mathbf{n}_i + \mathbf{x}_i}{\sqrt{1+x_i^2}}$

$$\begin{aligned}\mathbf{n}_1' \cdot \mathbf{n}_2' &= (\mathbf{n}_1 \cdot \mathbf{n}_2 + \mathbf{n}_1 \cdot \mathbf{x}_2 + \mathbf{n}_2 \cdot \mathbf{x}_1 + \mathbf{x}_1 \cdot \mathbf{x}_2)/\sqrt{\left(1+x_1^2\right)\left(1+x_2^2\right)} \\ &\simeq \mathbf{n}_1 \cdot \mathbf{n}_2 \left(1 - \frac{1}{2}\left(x_1^2 + x_2^2\right)\right) + \mathbf{n}_1 \cdot \mathbf{x}_2 + \mathbf{n}_2 \cdot \mathbf{x}_1 + \mathbf{x}_1 \cdot \mathbf{x}_2 \,. \end{aligned}$$

For the approximation we have used $x_i \equiv |\mathbf{x}_i| \ll 1$. The beam function B is negligibly small if this is not satisfied. The generic expression for the window function then becomes

$$W_\ell(\mathbf{n}_1 \cdot \mathbf{n}_2) = \int B(\mathbf{n}_1, \mathbf{x}_1) B(\mathbf{n}_2, \mathbf{x}_2) P_\ell(\mathbf{n}_1' \cdot \mathbf{n}_2')\, dx_1^2\, dx_2^2 \,. \tag{6.13}$$

We first simplify this formula for $\mathbf{n}_1 = \mathbf{n}_2 = \mathbf{n}$. Then, since $\mathbf{x}_1$ and $\mathbf{x}_2$ are normal to $\mathbf{n}$, the scalar product $\mathbf{n}_1' \cdot \mathbf{n}_2'$ becomes $\mathbf{n}_1' \cdot \mathbf{n}_2' \simeq 1 - \frac{1}{2}(x_1^2 + x_2^2) + \mathbf{x}_1 \cdot \mathbf{x}_2 \simeq \cos(|\mathbf{x}_1 - \mathbf{x}_2|)$. Furthermore, for small values of $|\mathbf{x}_1 - \mathbf{x}_2|/\ell$ and sufficiently large ℓ, we can approximate (see Appendix A4.1)

$$\begin{aligned} P_\ell\left(\cos(|\mathbf{x}_1 - \mathbf{x}_2|)\right) &\to J_0(\ell |\mathbf{x}_1 - \mathbf{x}_2|) \\ &= \frac{1}{\pi} \int_0^\pi d\phi\, \exp[-i\ell |\mathbf{x}_1 - \mathbf{x}_2| \cos\phi] \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\phi\, \exp[-i\ell |\mathbf{x}_1 - \mathbf{x}_2| \cos\phi] \,, \end{aligned} \tag{6.14}$$

where we have used Eq. (A4.90) for the first equality. Let us define the planar vector $\boldsymbol{\ell}$ as the vector with length ℓ which points at an angle ϕ from $\mathbf{x}_1 - \mathbf{x}_2$ and

$$\tilde{B}(\boldsymbol{\ell}) = \int B(\mathbf{x}) e^{-i\boldsymbol{\ell}\mathbf{x}}\, d^2x \,, \tag{6.15}$$

the two-dimensional Fourier transform of the beam pattern. Equation (6.13) together with Eq. (6.14) then yields

$$W_\ell(1) = \frac{1}{2\pi} \int_0^{2\pi} d\phi\, |\tilde{B}(\boldsymbol{\ell})|^2 \,. \tag{6.16}$$

The window function for $\mathbf{n}_1 = \mathbf{n}_2$ is the angular average of the square of the Fourier transformed beam pattern.

To find an expression for the window function for $\mathbf{n}_2 \neq \mathbf{n}_1$, let us expand $P_\ell(\mathbf{n}_1' \cdot \mathbf{n}_2')$ to second order in $\mathbf{x}_1$ and $\mathbf{x}_2$ also if $\mathbf{n}_1 \neq \pm\mathbf{n}_2$. Setting

$$\mathbf{n}_1' \cdot \mathbf{n}_2' = \mathbf{n}_1 \cdot \mathbf{n}_2 + \epsilon$$

up to order x_i^2 we have

$$\epsilon = -\frac{1}{2}\mathbf{n}_1 \cdot \mathbf{n}_2 \left(x_1^2 + x_2^2\right) + \mathbf{n}_1 \cdot \mathbf{x}_2 + \mathbf{n}_2 \cdot \mathbf{x}_1 + \mathbf{x}_1 \cdot \mathbf{x}_2 \,,$$

so that

$$P_\ell(\mathbf{n}_1' \cdot \mathbf{n}_2') \simeq P_\ell(\mathbf{n}_1 \cdot \mathbf{n}_2) + \epsilon P_\ell'(\mathbf{n}_1 \cdot \mathbf{n}_2) + \frac{1}{2}\epsilon^2 P_\ell''(\mathbf{n}_1 \cdot \mathbf{n}_2) \,. \tag{6.17}$$

This approximation is sufficient if P_ℓ does not vary too much in an interval ϵ, hence if $\ell\epsilon < 1$. Inserting it in Eq. (6.13) and keeping only terms up to second order in

x_i, we find

$$W_\ell(z) = P_\ell(\mathbf{n}_1 \cdot \mathbf{n}_2) + \frac{1}{2}\int d^2x_1\, d^2x_2\, B(\mathbf{n}_1, \mathbf{x}_1)B(\mathbf{n}_2, \mathbf{x}_2) \times \left[-2z\left(x_1^2 + x_2^2\right)P'_\ell(z) + ((\mathbf{n}_1 \cdot \mathbf{x}_2)^2 + (\mathbf{n}_2 \cdot \mathbf{x}_1)^2)P''_\ell(z)\right] , \quad (6.18)$$

with $z = \mathbf{n}_1 \cdot \mathbf{n}_2 = \cos\theta$. We have assumed that the beam is spherically symmetric around its centre and we have dropped all terms which were linear in the vectors $\mathbf{x}_i$ and therefore integrate to zero. Decomposing the vectors $\mathbf{x}_i$ into a component which lies in the $(\mathbf{n}_1, \mathbf{n}_2)$-plane $(\mathbf{e}_i)$ and a component orthogonal to it $(\mathbf{m})$, $\mathbf{x}_i = \mathbf{e}_i \cos\varphi_i + \mathbf{m}\sin\varphi$ we find,

$$\mathbf{n}_1\mathbf{x}_2 = \sin\theta\cos\varphi_2 , \quad \mathbf{n}_2\mathbf{x}_1 = \sin\theta\cos\varphi_1 .$$

Inserting this above, integration over angles gives

$$W_\ell(z) = P_\ell(z) + \frac{(2\pi)^2}{4}\int_0^\infty dx_1\, dx_2\, x_1x_2B(\mathbf{n}_1, \mathbf{x}_1)B(\mathbf{n}_2, \mathbf{x}_2) \times \left[-2z\left(x_1^2 + x_2^2\right)P'_\ell(z) + \left(x_2^2 + x_1^2\right)\sin^2\theta\, P''_\ell(z)\right] . \quad (6.19)$$

We now use the fact that $(1 - z^2)P''_\ell(z) - 2zP'_\ell(z) = -\ell(\ell + 1)P_\ell(z)$ (see Appendix A4.1) and the normalization of the beam, $2\pi\int_0^\infty dx\, xB(\mathbf{x}) = 1$. Furthermore, we define the width of the beam

$$\sigma^2 \equiv \pi\int_0^\infty dx\, x^3B(\mathbf{x}) . \quad (6.20)$$

With this, the window function simply becomes

$$W_\ell(z) = P_\ell(z)\left(1 - \sigma^2\ell(\ell + 1)\right) , \quad \text{for} \quad \sigma\ell \ll 1 . \quad (6.21)$$

This last condition is necessary to assure that $P_\ell(z + \epsilon)$ is well approximated by Eq. (6.17) for all $|\epsilon| \lesssim \sigma$. A reasonable approximation for all values of ℓ might therefore be $W_\ell(z) = P_\ell(z)\exp(-\sigma^2\ell(\ell + 1))$, which is (nearly) the result of the Gaussian beam as we see below.

Gaussian beam

For a Gaussian beam,

$$B(\mathbf{x}) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right) , \quad (6.22)$$

with Fourier transform

$$B(\ell) = \exp\left(-\ell^2\sigma^2/2\right) , \quad (6.23)$$

hence

$$W_\ell(1) = e^{-\ell^2\sigma^2} \,. \tag{6.24}$$

For $z \neq 0$ and $\sigma\ell \ll 1$ we reproduce Eq. (6.21) and

$$W_\ell(z) = P_\ell(z) \exp\left(-\sigma^2 \ell(\ell+1)\right) \tag{6.25}$$

is an excellent approximation to the numerical result for all values of z and ℓ as long as $\sigma \ll 1$. The only difference in $W_\ell(1)$ is that $\ell(\ell+1)$ becomes ℓ^2 which comes from the fact that we have approximated the sphere by a plane normal to $\mathbf{n}$. This is an irrelevant difference for sufficiently large values of ℓ.

Differencing beam

The disadvantage of the Gaussian beam is the fact that we have to subtract the mean temperature to relate it to the theoretical power spectrum of the CMB anisotropies. We subtract two large numbers to obtain a small result. A notoriously dangerous procedure to perform on noisy data. Therefore, instead of the Gaussian beam one usually utilizes a beam pattern with mean zero,

$$\int d^2\mathbf{x}\, B(\mathbf{x}) = 0 \,.$$

Often one simply adds the signals coming from different directions with weights which add up to 0. Let us analyse the simplest case of a single subtraction. We choose the line connecting the two beam centres to be the x axis in the $\mathbf{x} = (x, y)$ plane and define

$$B(x, y) = \frac{1}{2\pi\sigma^2}\left[\exp\left[-\frac{(x-x_0)^2+y^2}{2\sigma^2}\right] - \exp\left[-\frac{(x+x_0)^2+y^2}{2\sigma^2}\right]\right]. \tag{6.26}$$

This is the difference of two Gaussian beams separated by a 'throw' of $2x_0$. A similar calculation to the one for the Gaussian beam leads to

$$W_\ell(1) = e^{-\sigma^2\ell^2}\left(1 - P_\ell(\cos(2x_0))\right) \,. \tag{6.27}$$

In Fig. 6.4 we plot the window function for $\sigma = 1^\circ = \pi/180$ for two different values of the throw. As for the Gaussian beam, we cannot measure fluctuations with $\ell\sigma \gg 1$. But, what is new for the differencing beam, we are also not sensitive to fluctuations on scales much larger than the throw of the beam pattern since on these scales the beam averages to zero.

With the known window function, we can now relate the measured temperature fluctuations to the theoretical C_ℓs. As always when doing an experiment, we want to know the best estimate for C_ℓ and its error, or better, its probability distribution.

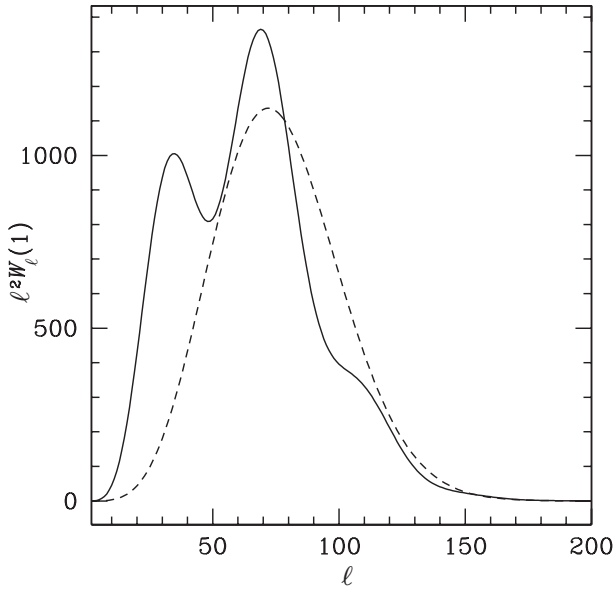


Fig. 6.4. We show the window function $\ell^2 W_\ell(1)$ for a differencing beam with width $\sigma = 1^\circ$ for two different values of the throw, $x_0 = 4^\circ$ (solid) and $x_0 = 1^\circ$ (dashed). The peak at $\ell \simeq 70$ is due to the beam size. If throw and beam size differ, a second peak appears on the left for the larger value of the throw.

Before exploring statistical methods to analyse CMB data, we discuss an error which is always present in cosmological experiments.

6.4.2 Cosmic variance

Only one CMB sky is at our disposition for observation. Therefore, when we measure the mean fluctuation in large angular patches, not many statistically independent patches are available in the sky and we expect relatively large statistical fluctuations.

Let us calculate this fluctuation under the assumption that the initial fluctuations are Gaussian. Then, the coefficients $a_{\ell m}$ are Gaussian variables and, in the optimal case when our data cover all sky, we can determine $2\ell + 1$ statistically independent $a_{\ell m}$s for a given value of ℓ. We want to determine the variance

$$\sigma_\ell = \sqrt{\frac{\langle (C_\ell^o - C_\ell)^2 \rangle}{C_\ell^2}} \, .$$

Here $C_\ell^o = (2\ell + 1)^{-1} \sum_m |a_{\ell m}|^2$ is the 'random' variable which we obtain when averaging over the $2\ell + 1$ measured $a_{\ell m}$s, and $C_\ell = \langle |a_{\ell m}|^2 \rangle$ is the statistical

expectation value. The square variance, σ^2 of $2\ell+1$ independent Gaussian variables is $1/(2\ell+1)$. For the squares of these variables, we expect σ^2 to double, by simple error propagation. This is exactly what we will find now with a more thorough calculation,

$$\begin{aligned}\langle (C_\ell^o - C_\ell)^2\rangle &= \frac{1}{(2\ell+1)^2}\left\langle \left(\sum_m \left[|a_{\ell m}|^2 - \langle |a_{\ell m}|^2\rangle\right]\right)^2\right\rangle \\ &= \frac{1}{(2\ell+1)^2}\sum_{m,m'}\left(\langle |a_{\ell m}|^2|a_{\ell m'}|^2\rangle - \langle |a_{\ell m}|^2\rangle\langle |a_{\ell m'}|^2\rangle\right) .\end{aligned}$$

The second term in the above sum is simply C_ℓ^2. For the first term we apply Wick's theorem which states that for a set of Gaussian variables, the $2n$-point correlation function is given by the sum of all the possible 2-point correlation functions that can be formed from it (see Appendix 7). Hence

$$\begin{aligned}\langle |a_{\ell m}|^2|a_{\ell m'}|^2\rangle &= \langle a_{\ell m}a^*_{\ell m}a_{\ell m'}a^*_{\ell m'}\rangle \\ &= \langle a_{\ell m}a^*_{\ell m}\rangle\langle a_{\ell m'}a^*_{\ell m'}\rangle + \langle a_{\ell m}a^*_{\ell m'}\rangle\langle a_{\ell m'}a^*_{\ell m}\rangle + \langle a_{\ell m}a_{\ell m'}\rangle\langle a^*_{\ell m'}a^*_{\ell m}\rangle \\ &= C_\ell^2 + \delta_{m,m'}C_\ell^2 + \delta_{m,-m'}C_\ell^2 .\end{aligned}$$

For the last equals sign we have used the fact that $a_{\ell m}$ and $a^*_{\ell m'} = a_{\ell,-m'}$ are independent random variables if $m \neq m'$. Summation over m and m' gives now

$$\langle (C_\ell^o - C_\ell)^2\rangle = \frac{2}{2\ell+1}C_\ell^2 \quad \text{so that}$$

$$\sigma_\ell = \sqrt{\frac{\langle (C_\ell^o - C_\ell)^2\rangle}{C_\ell^2}} = \sqrt{\frac{2}{2\ell+1}} .$$

This is the absolutely minimal error which can be achieved from one sky. It is a principle causality limit that cannot be escaped. To it we have to add instrumental noise, foregrounds, atmospheric noise, etc.

Even if there were a far away civilization at a cosmological distance which would undertake similar measurements and then send us their results, this would not really help. By the time it takes for their information to arrive on Earth, the CMB sky has grown by so much, that the region they have observed is now also inside our Hubble horizon and we can observe it in our CMB experiments. On the other hand, if they have sent us the data long ago, the sky they could observe at this time, is now also inside our Hubble horizon. The problem is, of course, causality. If the Universe is not inflating, we can by no means obtain any information about a region outside our Hubble horizon which corresponds roughly to the CMB sky.

If we observe a fraction $f < 1$ of the sky, the error increases since we now cannot determine all the $a_{\ell m}$s, and $2\ell+1$ is replaced by $(2\ell+1)f$. This is roughly the

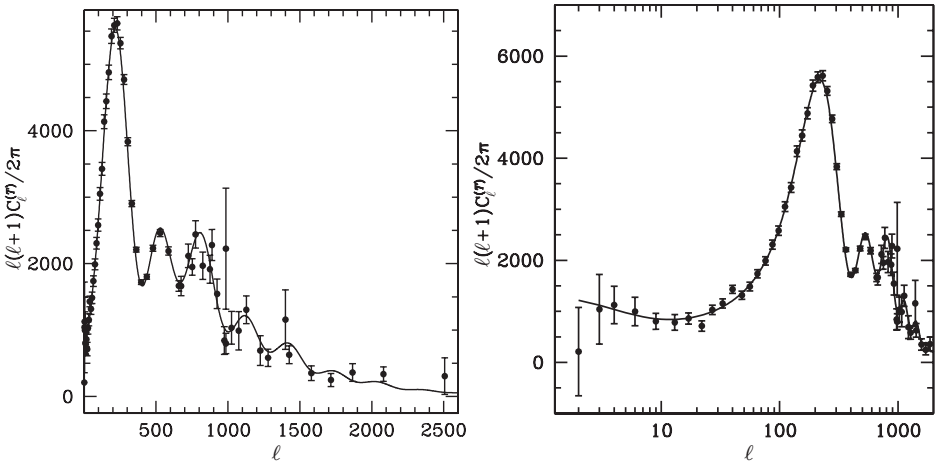


Fig. 6.5. The observed CMB anisotropy spectrum from WMAP (Hinshaw *et al.*, 2007) extended by BOOMERanG (Jones *et al.*, 2006) and Acbar (Kuo *et al.*, 2006) in linear (left) and log (right) scale. The line draws the best-fitting ΛCDM model.

number of independent $a_{\ell m}$s which can be measured in a fraction f of the sky. The variance then increases to

$$\sigma_\ell = \sqrt{\frac{2}{(2\ell+1)f}}\,.$$

Finally, we note that in real full sky CMB experiments, the data close to the galactic plane are strongly contaminated by foregrounds and it is safest not to use them at all. One therefore often cuts out a region of about 20° around the galactic plane. In this cut sky, the spherical harmonics are no longer an orthonormal basis of function and one has to conceive a new method to define such a basis. This can, in principle, be done by applying a Cauchy–Schwartz orthogonalization procedure on the old basis, see Gorski (1994).

In Figs. 6.5–6.7 we show the recently (July 2007) available CMB data in terms of the anisotropy, polarization and cross correlation spectra.

6.5 Statistical methods

To extract the optimal information from data it is most useful to apply the best statistical method. However, as a rule of thumb, results that strongly depend on the statistical method applied, are not to be trusted.

Some elementary statistical tools that are used in this chapter are presented in Appendix 7.

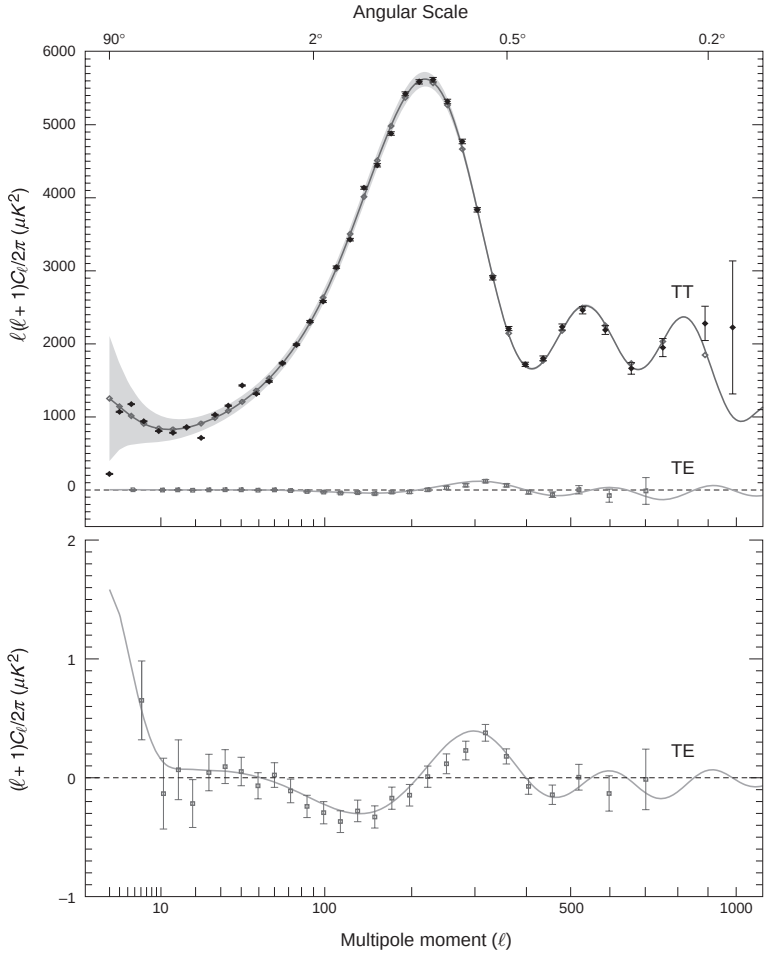


Fig. 6.6. The CMB anisotropy spectrum and the temperature–polarization cross-correlation obtained from the WMAP 3-yr data (figure from Hinshaw *et al.* (2007)).

6.5.1 Bayes' theorem and the likelihood function

To estimate cosmological parameters one uses a simple result from probability theory which goes under the name of Bayes' theorem. In a probability space, consider two sets A and B and their intersection $A \cap B$. The probability that a given event of which we know that it is in A is also in B is

$$\frac{P[A \cap B]}{P[A]} =: P[B|A]\,. \tag{6.28}$$

$P[B|A]$ denotes the conditional probability of B given A. Exchanging A and B we obtain the conditional probability of A given B. Hence

$$P[A \cap B] = P[B|A]P[A] = P[A|B]P[B]\,. \tag{6.29}$$

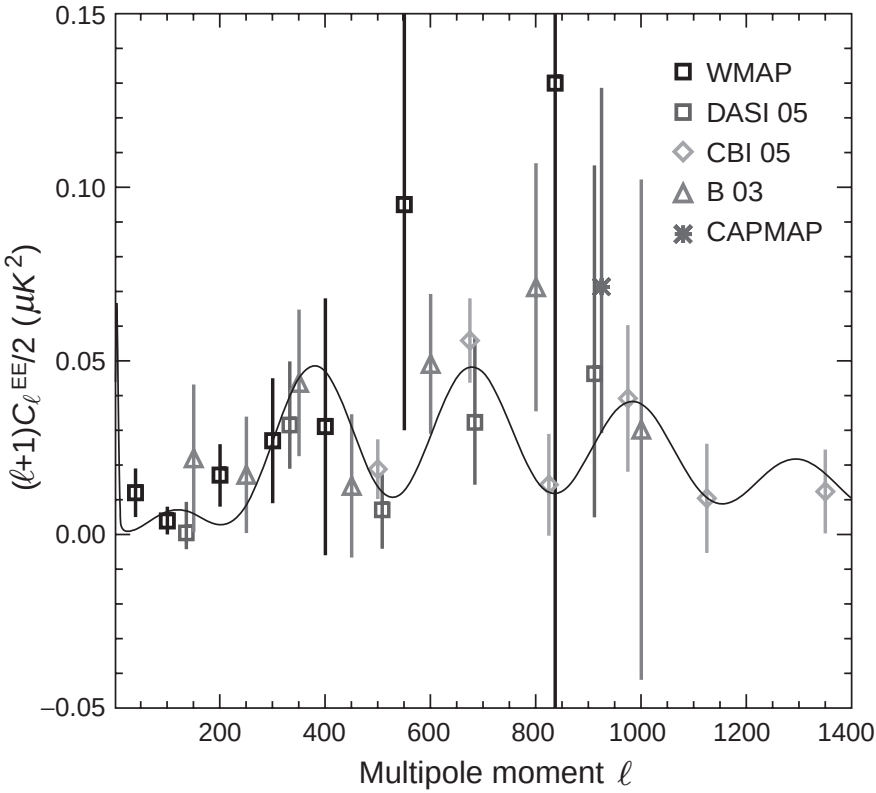


Fig. 6.7. The measured EE polarization spectrum from WMAP 3-yr data, BOOMERanG and others. For more details see Page *et al.* (2007) from where this figure is taken.

The last equation which is here written for the probabilities of sets is, of course, also true for the corresponding probability densities. What is the relevance of this simple statement for parameter estimation? Let us assume that a cosmological model is described by a set of parameters $(p_1, \ldots, p_M) = m$. Our experiment has made a series of measurements and has come up with data $(d_1, \ldots, d_N)$ with errors $(\sigma_1, \ldots, \sigma_N)$, for example the CMB temperature in different directions $\mathbf{n}_i$. For a given model m we can calculate the predicted outcome of the experiment in terms of expectation value and variance $(d_i(m), \sigma_i(m))$ for each of the data points. Here we assume the situation, as it is in cosmology, that the model is of a statistical nature and predicts expectation values for the measurements $d_i(m)$ and their variances, $\sigma_i(m)$. For example, if the d_is are coefficients $a_{\ell m}$ of the expansion of the temperature fluctuations, then their expectation values vanish and the variances are the C_ℓs. However, if your data are the C_ℓs then their variance is given by cosmic variance. It is sometimes not clear whether the errors $\sigma_i(m)$ should be added to the data

or to the model. But we just need to require that they are independent of (other) measurement errors and therefore can be added in quadrature.

If the distribution of the data points d_i is Gaussian (as we assume the CMB temperature fluctuations to be), the probability of measuring d_i in model m taking into account the measurement uncertainty σ_i is given by (see Appendix 7)

$$P[d_i|m] = \frac{1}{\sqrt{2\pi(\sigma_i(m)^2 + \sigma_i^2)}} \exp\left(-\frac{(d_i - d_i(m))^2}{2(\sigma_i(m)^2 + \sigma_i^2)}\right) . \tag{6.30}$$

Note that the theoretical uncertainty, $\sigma_i(m)$ and the measurement error, σ_i have been added in quadrature. If the measurements d_i are independent, e.g. if the directions $\mathbf{n}_i$ are much further apart than the beam width, the joint probability of measuring $(d_1, \ldots, d_N)$ with errors $(\sigma_1, \ldots, \sigma_N)$ is the product of the individual probabilities,

$$P[\{d_i, \sigma_i\}|m] \equiv \mathcal{L}(\{d_i, \sigma_i\}, m) = \Pi_{i=1}^{N} \left[\frac{\exp\left(-\frac{(d_i - d_i(m))^2}{2(\sigma_i(m)^2 + \sigma_i^2)}\right)}{\sqrt{2\pi(\sigma_i(m)^2 + \sigma_i^2)}}\right] . \tag{6.31}$$

In the more general case, when the measurements are not independent but Gaussian with correlation function,

$$\langle d_i d_j \rangle = \mathcal{C}_{ij} , \tag{6.32}$$

the above expression becomes

$$\mathcal{L}(\{d_i, \sigma_i\}, m) = \frac{1}{\sqrt{\det \mathcal{C}(2\pi)^N}} \exp\left(-\frac{d_i \mathcal{C}_{ij}^{-1} d_j}{2}\right), \tag{6.33}$$

(see Appendix 7). This expression is called the likelihood function. It gives us the likelihood of the data $\{d_i, \sigma_i\}$ given the model m. But since it is the data that we know and the model that we would like to know, we would be more interested in the probability of a model m given the data. And here, Bayes' theorem comes to our rescue. According to Eq. (6.29)

$$P[m|\{d_i, \sigma_i\}] = P[\{d_i, \sigma_i\}|m] \frac{P[m]}{P[\{d_i, \sigma_i\}]} . \tag{6.34}$$

Here $P[m]$ is called 'the prior' and the denominator is called 'the evidence'. The right-hand side is the 'posterior distribution' or simply the 'posterior'. Ideally it can be used as 'prior' for the next experiment. The evidence $P[\{d_i, \sigma_i\}]$ is unimportant for parameter estimation since it does not depend on the model parameters, hence the ratio $P[m_1|\{d_i, \sigma_i\}]/P[m_2|\{d_i, \sigma_i\}]$ is independent of the evidence. Furthermore, it can be eliminated noting that when integrating over the entire space of model

parameters, the left-hand side must be normalized to 1,

$$P[\{d_i, \sigma_i\}] = \int dm\ P[m]\mathcal{L}(\{d_i, \sigma_i\}, m)\ .$$

The prior $P[m]$ depends on our prior knowledge of the model space. For example, if previous experiments have shown us that curvature is not huge, it makes sense to choose $P[m] = 0$ for all models with $|\Omega_K| > \frac{1}{2}$. But how shall we choose $P[m]$ for $|\Omega_K| < \frac{1}{2}$? We may opt for a 'flat' prior, i.e. the same probability for all values of Ω_K. But this is as much a special case as choosing a flat prior for $\exp(\Omega_K)$. This is the weakest point of Bayesian likelihood: the probability of a model m for given data $\{d_i, \sigma_i\}$ depends on our prior.

Things are not as hopeless as this might seem. First of all, physics often tells us to some extent what the distribution of the prior should be. As a rule of thumb, parameters which add to the data should have a flat prior while parameters which multiply the data (scaling parameters) more naturally have a logarithmic prior i.e. a flat prior in $\log(\lambda)$.

If the data have sufficiently small error bars, most priors are relatively flat and the resulting maximum likelihood is nearly independent of the prior. However, if the data are relatively weak and errors are large, the Bayesian likelihood $P[m|\{d_i, \sigma_i\}]$ may well be strongly prior dependent. In this case, the data are simply not good enough to choose a model.

The model with the highest probability to be correct is given by the maximum of $P[m|\{d_i, \sigma_i\}]$. If the data are sufficiently good to choose a model, i.e., if the prior $P[m]$ is sufficiently flat around the maximum of $P[m|\{d_i, \sigma_i\}]$, then the latter is close to the maximum of the likelihood function, $\mathcal{L}(\{d_i, \sigma_i\}, m)$. When estimating parameters, we therefore simply search for a maximum of the likelihood function.

6.5.2 Fisher matrix and parameter estimation

Best-fitting parameters

We consider a set of parameters $\lambda = (\lambda_1, \ldots, \lambda_M)$. The best-fitting model with parameters $\bar{\lambda}$ corresponds to the maximum of the likelihood function,

$$\left.\frac{d\mathcal{L}(\lambda)}{d\lambda}\right|_{\bar{\lambda}} = 0\ . \tag{6.35}$$

It is most easily determined with a root-finder method applied to $d\mathcal{L}(\lambda)/d\lambda$. However, for a Gaussian distribution the likelihood function is an exponential. Therefore, its logarithm is usually better suited to a numerical root finder. One starts at some

best guess value $\lambda^{(0)}$ and then approximates the derivative $d\ln\mathcal{L}/d\lambda$ to first order,

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda_i}(\bar{\lambda}) \simeq \frac{\partial \ln \mathcal{L}}{\partial \lambda_i}(\lambda^{(0)}) + (\bar{\lambda}_j - \lambda_j^{(0)})\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i \partial \lambda_j}(\lambda^{(0)}) \, .$$

Setting $\frac{\partial \ln \mathcal{L}}{\partial \lambda}(\bar{\lambda}) = 0$, we obtain to first order

$$(\bar{\lambda}_j - \lambda_j^{(0)}) \simeq -\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_j \partial \lambda_i}\right) \frac{\partial \ln \mathcal{L}}{\partial \lambda_i}(\lambda_0) \equiv \delta\lambda_j \, . \tag{6.36}$$

For the next step we can replace $\lambda^{(0)}$ by $\lambda^{(1)} = \lambda^{(0)} + \delta\lambda$ and iterate the procedure until it converges. In directions in which the likelihood function is very flat, we make large steps while in directions along which it is steep, the steps are small.

However, this is not how it is usually done. There is a simplification which can be made without a great loss of accuracy. For the CMB, we expect the likelihood function to be of the form (6.33) so that

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda_i} = -\frac{1}{2}\frac{\partial}{\partial \lambda_i}\left[\ln(\det \mathcal{C}) + d^T \mathcal{C}^{-1} d\right] \, . \tag{6.37}$$

For notational simplicity we now denote $\partial/\partial\lambda_i \equiv \partial_i$. With $\ln(\det \mathcal{C}) = \mathrm{Trace}(\ln(\mathcal{C}))$ and $\partial_i \mathcal{C}^{-1} = -\mathcal{C}^{-1}(\partial_i \mathcal{C})\mathcal{C}^{-1}$ we obtain

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda_i} = \frac{1}{2}\left[d^T \mathcal{C}^{-1}(\partial_i \mathcal{C})\mathcal{C}^{-1} d - \mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_i \mathcal{C})\right)\right] \, . \tag{6.38}$$

For the second derivative we find after similar manipulations

$$\begin{aligned}\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i \partial \lambda_j} = -\frac{1}{2}\Big[& d^T \mathcal{C}^{-1}(\partial_i \mathcal{C})\mathcal{C}^{-1}(\partial_j \mathcal{C})\mathcal{C}^{-1} d \\ & + d^T \mathcal{C}^{-1}(\partial_j \mathcal{C})\mathcal{C}^{-1}(\partial_i \mathcal{C})\mathcal{C}^{-1} d - \mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_j \mathcal{C})\mathcal{C}^{-1}(\partial_i \mathcal{C})\right) \\ & - d^T \mathcal{C}^{-1}(\partial^2_{ij} \mathcal{C})\mathcal{C}^{-1} d + \mathrm{Trace}\left(\mathcal{C}^{-1}(\partial^2_{ij} \mathcal{C})\right)\Big] \, .\end{aligned} \tag{6.39}$$

The Fisher matrix is defined as the expectation value

$$F_{ij} \equiv \left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i \partial \lambda_j} \right\rangle \, . \tag{6.40}$$

To determine it we use $\langle d_i d_j \rangle = \mathcal{C}_{ij}$ so that

$$\begin{aligned}\langle d^T \mathcal{C}^{-1}(\partial_i \mathcal{C})\mathcal{C}^{-1}(\partial_j \mathcal{C})\mathcal{C}^{-1} d \rangle &= \langle d_m \mathcal{C}^{-1}_{mn}(\partial_i \mathcal{C})_{np}\mathcal{C}^{-1}_{pq}(\partial_j \mathcal{C})_{qr}\mathcal{C}^{-1}_{rs} d_s \rangle \\ &= \mathcal{C}^{-1}_{mn}(\partial_i \mathcal{C})_{np}\mathcal{C}^{-1}_{pq}(\partial_j \mathcal{C})_{qr}\mathcal{C}^{-1}_{rs}\mathcal{C}_{sm} \\ &= \mathrm{Trace}\left((\partial_i \mathcal{C})\mathcal{C}^{-1}(\partial_j \mathcal{C})\mathcal{C}^{-1}\right) \\ &= \mathrm{Trace}\left((\partial_j \mathcal{C})\mathcal{C}^{-1}(\partial_i \mathcal{C})\mathcal{C}^{-1}\right) \\ &= \mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_j \mathcal{C})\mathcal{C}^{-1}(\partial_i \mathcal{C})\right) \, .\end{aligned}$$

For the last two equals signs we use the fact that the trace is invariant under cyclic permutations. Equivalently

$$\langle d^T \mathcal{C}^{-1}(\partial_{ij}^2 \mathcal{C})\mathcal{C}^{-1} d\rangle = \text{Trace}\left(\mathcal{C}^{-1}(\partial_{ij}^2 \mathcal{C})\right) \ .$$

Inserting these results to calculate the expectation value of (6.39) we obtain

$$F_{ij} = \frac{1}{2}\text{Trace}\left(\mathcal{C}^{-1}(\partial_j \mathcal{C})\mathcal{C}^{-1}(\partial_i \mathcal{C})\right) \ . \tag{6.41}$$

We assume the Fisher matrix to be non-singular. Otherwise not all the parameters λ_i can be determined by the data since then the Fisher matrix has a vanishing eigenvalue for some eigenvector $\mu \neq 0$. The covariance matrix and hence the likelihood function are then independent of the linear combination $\mu_i \lambda_i$ which therefore cannot be estimated by the experiment. Clearly, F is also symmetric and $F_{ij}\lambda^i\lambda^j = \frac{1}{2}\text{Trace}\left((\mathcal{C}^{-1}(\lambda_j \partial_j \mathcal{C}))^2\right) > 0$ for all $\lambda \neq 0$. The Fisher matrix is a positive-definite symmetric matrix.

Instead of dividing by the true curvature of the likelihood function in Eq. (6.36) to determine the next estimator for the parameters, one divides by the corresponding element of the Fisher matrix,

$$\lambda_i^{(1)} = \lambda_i^{(0)} + \frac{1}{2}F_{ij}^{-1}\left[d^T\mathcal{C}^{-1}(\partial_j\mathcal{C})\mathcal{C}^{-1}d - \text{Trace}\left(\mathcal{C}^{-1}(\partial_j\mathcal{C})\right)\right](\lambda^{(0)}) \ . \tag{6.42}$$

The advantage of this method is that for a given starting parameter the Fisher matrix $F_{ij}(\lambda^{(0)})$ is readily calculated from the covariance matrix alone, without having to know the data. As we shall see below, this can be used to forecast the precision with which a given experimental setup (hence given covariance matrix) is able to estimate parameters.

The above estimated parameters $\lambda^{(1)}$ depend quadratically on the data vector d. For this reason they are called a 'quadratic estimator'. The Fisher matrix actually is a local Gaussian approximation to the distribution of the parameters in the vicinity of the parameter values λ. Dividing by the true curvature of the likelihood function and not its expectation value, would not have provided a quadratic estimator. It can be shown that the estimator described here is actually an optimal quadratic estimator (see e.g. Dodelson, 2003).

In practice, relatively few iterations are needed to achieve convergence. This is very important since we often search in a parameter space of 10 or more dimensions. A modest grid of 10 points per side, would already lead to 10^{10} evaluations of the likelihood function. Each of these requires one run of a fast CMB code which in an optimized code takes a few seconds. One evaluation of the likelihood function then requires a computational time of roughly 1 s. Hence 10^{10} evaluations take in the optimal case 10^{10} s $\simeq$ 300 yr. Not a time span we comfortably wait for the output

of a computation. Therefore, it is imperative that we use an iterative procedure and not a grid to estimate parameters.

The method described so far still has several problems, some of which we briefly want to address.

- *What if we end up in a shallow local maximum of the likelihood function and are stuck there?*
  To avoid this problem, one usually adds a small random fluctuation to the obtained $\delta\lambda$, a 'temperature', so that one can leave a shallow local maximum.
- *What if there are several local maxima, some of them quite steep and separated by deep ridges?*
  To check this, one performs not only one but many iteration chains with different starting points. One can then compare the height of the different maxima. A procedure along these lines is the Markov chain Monte Carlo method (MCMC) discussed below. It is presently the method of choice for CMB analysis. A publicly available MCMC code and more details of the method can be found in the paper by Lewis & Bridle (2002).
- *What if the maximum is somewhere at the border of parameter space?*
  The border of the parameter space is given by the prior. If the data are best fitted by parameter values lying at the border of what is allowed by the prior, this hints that either the prior is wrong or the model is incorrect. This is one of the most important drawbacks of the Fisher matrix technique. It provides relatively rapidly the best-fitting model under consideration, but it works independently of whether this model is actually a good fit to the data or not. For this an evaluation of the likelihood function at the best-fitting parameter values has to be performed. If the likelihood function is very small, this is either a sign that the model is wrong or that the real errors are much larger than those assumed.

Estimating errors

So far we have only studied the problem of how to find the best fit. But we also want error bars on the estimated parameters. A good first estimate for 1σ error bars are the diagonal elements of the Fisher matrix at the maximum $\bar{\lambda}$. To see this, we use the expression (6.42) for the deviation $\delta\lambda$

$$\begin{aligned}\langle\delta\lambda_i\delta\lambda_j\rangle &= \frac{1}{4}F_{im}^{-1}F_{jn}^{-1}\left\langle\left[d^T\mathcal{C}^{-1}(\partial_m\mathcal{C})\mathcal{C}^{-1}d - \mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_m\mathcal{C})\right)\right]\right.\\ &\quad\left.\times\left[d^T\mathcal{C}^{-1}(\partial_n\mathcal{C})\mathcal{C}^{-1}d - \mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_n\mathcal{C})\right)\right]\right\rangle\\ &= \frac{1}{4}F_{im}^{-1}F_{jn}^{-1}\left[\langle d_k d_l d_p d_q\rangle\mathcal{C}_{kr}^{-1}(\partial_m\mathcal{C})_{rs}\mathcal{C}_{sl}^{-1}\mathcal{C}_{pv}^{-1}(\partial_n\mathcal{C})_{vw}\mathcal{C}_{wq}^{-1}\right.\\ &\quad\left.-\mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_m\mathcal{C})\right)\mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_n\mathcal{C})\right)\right]\,.\end{aligned}$$

For the second term we made use of $\langle d_p d_q\rangle = \mathcal{C}_{pq}$. With this, the mixed terms become equal to the pure trace term,

$$\langle d^T\mathcal{C}^{-1}(\partial_m\mathcal{C})\mathcal{C}^{-1}d\rangle\mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_n\mathcal{C})\right) = \mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_m\mathcal{C})\right)\mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_n\mathcal{C})\right)\,.$$

For the first term we apply Wick's theorem (see Appendix 7),

$$\langle d_k d_l d_p d_q \rangle = \mathcal{C}_{kl}\mathcal{C}_{pq} + \mathcal{C}_{kp}\mathcal{C}_{lq} + \mathcal{C}_{kq}\mathcal{C}_{lp} \ .$$

Inserting this above, the first term results again in Trace $\left(\mathcal{C}^{-1}(\partial_m \mathcal{C})\right) \times$ Trace $\left(\mathcal{C}^{-1}(\partial_n \mathcal{C})\right)$ while the second and third terms give rise to twice the Fisher matrix F_{mn}. We therefore finally obtain

$$\langle \delta\lambda_i \delta\lambda_j \rangle = F_{im}^{-1} F_{jn}^{-1} F_{mn} = F_{ij}^{-1} \ . \tag{6.43}$$

Therefore, the diagonal elements of the Fisher matrix usually give reasonable errors for the parameters. Of course, these are the true 1σ errors only if the distribution is Gaussian *in the parameters* λ which usually it is not. But since the log of the likelihood function peaks at $\bar{\lambda}$ it will *locally* be of the form of a Gaussian,

$$\mathcal{L}(\bar{\lambda} + \delta\lambda) \simeq \mathcal{L}(\bar{\lambda}) \exp\left(-\frac{1}{2}\delta\lambda^T F(\bar{\lambda})\, \delta\lambda\right) \ ,$$

for small enough $\delta\lambda$. The Fisher matrix defines an ellipse around $\bar{\lambda}$ in the parameter space via the equation,

$$\delta\lambda^T F(\bar{\lambda}) \delta\lambda = 1 \ . \tag{6.44}$$

The principal directions of this error ellipse are parallel to the eigenvectors of F and their half length is given by the square root of the eigenvalues of F^{-1}.

According to Eq. (6.44), the total width of the error ellipse in a given direction λ_j at the centre $\bar{\lambda}$ is $2/\sqrt{F_{jj}}$. Therefore, $1/\sqrt{F_{jj}}$ is the error of the parameter λ_j if all other parameters are known and are equal to $\bar{\lambda}$. However, realistically, we do not know the other parameters any better than λ_j. Therefore, the true error in λ_j is given by the size of the projection of the error matrix onto the j axis, see Fig. 6.8. These are the so-called marginalized errors, which we obtain when integrating over all the other parameters.

We now show that the marginalized error of λ_i is given by $\sqrt{F_{ii}^{-1}}$, i.e., the diagonal element of the inverse of the Fisher matrix. For this we solve the quadratic equation Eq. (6.44) for $\delta\lambda_i$,

$$\delta\lambda_i = \frac{1}{F_{ii}}\left(-\sum_{j\neq i} F_{ji}\delta\lambda_j \pm \sqrt{\left(\sum_{j\neq i} F_{ji}\delta\lambda_j\right)^2 - F_{ii}\left(\sum_{jk\neq i} F_{jk}\delta\lambda_j\delta\lambda_k - 1\right)}\right) \ .$$

This equation expresses $\delta\lambda_i$ as a function of all the other parameters $\delta\lambda_j$. To determine *the maximum* of $\delta\lambda_i$, we set the gradient of this function to zero. From the

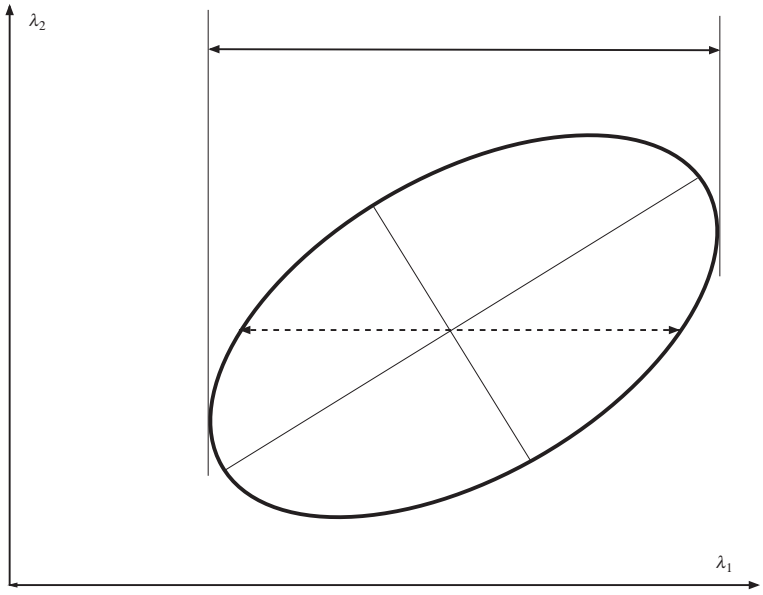


Fig. 6.8. The error ellipse is shown in a two-dimensional example. The widths $2/\sqrt{F_{11}}$ (dashed double arrow) and $2\sqrt{F_{11}^{-1}}$ (solid double arrow) are indicated.

derivative w.r.t. λ_j we obtain

$$0 = \frac{1}{F_{ii}} \left(-F_{ji} \pm \frac{F_{ji} \sum_{k \neq i} F_{ki} \delta\lambda_k - F_{ii} \sum_{k \neq i} F_{jk} \delta\lambda_k}{\sqrt{\cdots}} \right) ,$$

where the square root $\sqrt{\cdots}$ is the same as in the previous equation. Multiplying by $\mp\sqrt{\cdots} = F_{ii}\delta\lambda_i + \sum_{k \neq i} F_{ki}\delta\lambda_k$ we find

$$0 = F_{ji}\delta\lambda_i + \sum_{k \neq i} F_{jk}\delta\lambda_k \, . \tag{6.45}$$

This equation must hold for all $j \neq i$. Inserting $\delta\lambda_k = a F_{ki}^{-1}$ and $\delta\lambda_i = a F_{ii}^{-1}$ we obtain $0 = a\delta_{ij}$ which is certainly true since $j \neq i$. Since the Fisher matrix is non-singular the above solution is the only possibility. Inserting it in Eq. (6.44) determines $a = 1/\sqrt{F_{ii}^{-1}}$, so that we arrive at the important result

$$\delta\lambda_i^{(\mathrm{marg})} = \sqrt{F_{ii}^{-1}} \, . \tag{6.46}$$

These error ellipses are useful for forecasting if the errors are roughly Gaussian. However, if the true error contours have a very different shape, e.g. 'bananas' as in

Fig. 6.17, the Fisher matrix approximation usually underestimates the parameter errors.

The Fisher matrix for CMB anisotropy experiments and forecasting

For a given cosmological model we can calculate the C_ℓs. We now consider them as our model parameters λ_ℓ and want to determine their best-fitting values and errors. We determine the Fisher matrix from the correlation matrix $\mathcal{C}_{\ell m,\ell' m'} = \langle a_{\ell m} \bar{a}_{\ell' m'} \rangle$. (Take care not to mix up the CMB power spectrum C_ℓs with the correlation function $\mathcal{C}$s.)

We have already seen in Eq. (6.11) that a finite beam size σ leads to a correlation function of the form

$$\mathcal{C}_{\ell m,\ell' m'} = \delta_{\ell\,\ell'}\delta_{m\,m'} C_\ell W_\ell \quad \text{with, e.g.,}$$
$$W_\ell = \begin{cases} e^{-\ell^2\sigma^2} & \text{for a single beam,} \\ e^{-\ell^2\sigma^2}\,[1 - P_\ell(\cos(2x_0))] & \text{for a differencing beam with throw } x_0\,. \end{cases}$$

To this we have to add the correlation function of the noise. For simplicity we assume isotropic noise of amplitude σ_n in each pixel and a pixel size $\Delta\Omega$ in radians. In other words, a noise correlation function of the form

$$\frac{1}{T^2}\langle \Delta T^{(n)}(\mathbf{n}) \Delta T^{(n)}(\mathbf{n}')\rangle = \begin{cases} \sigma_n^2 & \text{if } \mathbf{n} \text{ and } \mathbf{n}' \text{ are in the same pixel} \\ 0 & \text{else}\,, \end{cases}$$
$$= \frac{1}{4\pi}\sum_\ell (2\ell+1) C_\ell^{(n)} P_\ell(\mathbf{n}\cdot\mathbf{n}')\,. \tag{6.47}$$

We have already taken into account that the noise is isotropic and therefore $\left\langle a_{\ell m}^{(n)} \bar{a}_{\ell' m'}^{(n)} \right\rangle = \delta_{\ell\,\ell'}\delta_{m\,m'} C_\ell^{(n)}$. To isolate $C_{\ell_1}^{(n)}$ we set $\mu = \mathbf{n}\cdot\mathbf{n}'$, multiply the above equation with $P_{\ell_1}(\mu)$ and integrate over μ. Defining

$$f(\mu) = \begin{cases} \sigma_n^2 & \text{if } 1-\mu < \Delta\mu \\ 0 & \text{else,} \end{cases}$$

we have

$$\int_{-1}^{1} d\mu\; P_{\ell_1}(\mu) f(\mu) \simeq -\Delta\mu\sigma_n^2 \simeq \Delta\vartheta\,\sin(\Delta\vartheta/2)\sigma_n^2\,,\ \ell < \frac{1}{\Delta\mu}.$$

Here we have used the fact that $P_\ell(\mu) \simeq 1$ for $\mu \simeq 1$ and $-\Delta\mu = -\Delta\cos\vartheta \simeq \Delta\vartheta\,\sin(\Delta\vartheta/2)$. If $\ell \gtrsim 1/\Delta\mu$ this approximation breaks down. But it is clear that with an experiment of pixel size corresponding to $\mathbf{n}\cdot\mathbf{n}' \leq 1-\Delta\mu$ we cannot measure C_ℓs with $\ell > 1/\Delta\mu$. We are therefore not considering these values. In other words, we are only determining the C_ℓs for values of ℓ with $\ell < \ell_{\max} \simeq 1/(2\Delta\mu)$. Also

multiplying the l.h.s. of Eq. (6.47) with P_{ℓ_1} and using $\int P_\ell P_{\ell_1} = \delta_{\ell\ell_1}\, 2/(2\ell+1)$ we obtain

$$C_{\ell_1}^{(n)} = 2\pi\,\Delta\vartheta\,\sin(\Delta\vartheta/2)\sigma_n^2 = \Delta\Omega\sigma_n^2\,. \tag{6.48}$$

It is reasonable to assume that the signal and the noise are uncorrelated so that we can simply add their correlation functions and arrive at

$$\left\langle a_{\ell m}^{(n)} \bar{a}_{\ell' m'}^{(n)} \right\rangle = \delta_{\ell\,\ell'}\,\delta_{m\,m'}\left[C_\ell W_\ell + w^{-1}\right]\,, \tag{6.49}$$

where we have introduced the width $w = (\Delta\Omega\sigma_n^2)^{-1}$. With this correlation function at hand, we can now calculate the Fisher matrix. Denoting the derivative with respect to C_ℓ by ∂_ℓ, Eq. (6.41) yields

$$\begin{aligned}
F_{\ell\ell'} &= \frac{1}{2}\mathrm{Trace}\left(\mathcal{C}^{-1}(\partial_\ell\mathcal{C})\mathcal{C}^{-1}(\partial_{\ell'}\mathcal{C})\right) \\
&= \frac{1}{2}\mathcal{C}^{-1}_{\ell_1 m_1,\ell_2 m_2}(\partial_\ell\mathcal{C})_{\ell_2 m_2,\ell_3 m_3}\mathcal{C}^{-1}_{\ell_3 m_3,\ell_4 m_4}(\partial_{\ell'}\mathcal{C})_{\ell_4 m_4,\ell_1 m_1} \\
&= \frac{\frac{1}{2}\delta_{\ell_1\ell_2}\delta_{m_1 m_2}}{\left[C_{\ell_1}W_{\ell_1}+w^{-1}\right]}\,\delta_{\ell_2\ell_3}\,\delta_{m_2 m_3}W_{\ell_2}\,\delta_{\ell_2\ell}\,\frac{\delta_{\ell_3\ell_4}\delta_{m_3 m_4}}{\left[C_{\ell_3}W_{\ell_3}+w^{-1}\right]}\delta_{\ell_4\ell_1}\delta_{m_4 m_1}W_{\ell'}\,\delta_{\ell_4\ell'} \\
&= \delta_{\ell\ell'}\,\frac{(2\ell+1)W_\ell^2}{2[C_\ell W_\ell + w^{-1}]^2} = \delta_{\ell\ell'}\,\frac{(2\ell+1)}{2[C_\ell + (wW_\ell)^{-1}]^2}\,.
\end{aligned} \tag{6.50}$$

The factor $2\ell+1$ comes from the summation over the ms,

$$\sum_{m_1 m_2 m_3 m_4}\delta_{m_1 m_2}\,\delta_{m_2 m_3}\,\delta_{m_3 m_4}\,\delta_{m_4 m_1} = \sum_{m_1}\delta_{m_1 m_1} = 2\ell_1+1\,,$$

while the summation over the ℓs requires $\ell_1=\ell_2=\ell=\ell_3=\ell_4=\ell'$ and therefore leads to $\delta_{\ell\ell'}$.

Using the C_ℓs as our 'parameters' has the big advantage that the Fisher matrix is diagonal and the marginalized errors are simply given by $1/\sqrt{F_{\ell\ell}} = \sqrt{F_{\ell\ell}^{-1}}$,

$$\delta C_\ell = \sqrt{\frac{2}{2\ell+1}}[C_\ell + (wW_\ell)^{-1}]\,. \tag{6.51}$$

For a given window function, pixel size and pixel noise, this gives a good error estimate for the accuracy with which the C_ℓs can be determined in a full sky experiment. If the experiment covers only a fraction $f_{\rm sky}$ of the sky, a good approximation for the error is again

$$\delta C_\ell = \sqrt{\frac{2}{(2\ell+1)f_{\rm sky}}}[C_\ell + (wW_\ell)^{-1}]\,. \tag{6.52}$$

Of course, the model parameters that we really want to estimate are not the C_ℓs, but rather cosmological parameters like $\Omega_m h^2$, Ω_Λ and the curvature. For this we can now start the process over again, considering the C_ℓs as our 'data' with correlation matrix $F_{\ell\ell'}^{-1}$ and calculate the Fisher matrix of the cosmological parameters, $\tilde{F}(\lambda)$. The only disadvantage here is that contrary to the $a_{\ell m}$s, the C_ℓs do not obey a Gaussian distribution. But for high ℓs the distribution is nearly Gaussian and for low ℓs the error is relatively large due to cosmic variance, so that treating the distribution as Gaussian does not usually induce large errors. Of course $\tilde{F}(\lambda)$ is by no means diagonal and the errors in the cosmological parameters alone are not Gaussian at all. If we want not only to estimate errors of the parameters, which can be obtained from the inverse of the Fisher matrix $\tilde{F}^{-1}$, but the full marginalized probability distribution, we have to use a more sophisticated method, like MCMC discussed below. The forecasted probability distributions calculated with a MCMC method for the Planck satellite experiment to be launched in 2008 are shown in Fig. 6.10. The probability distributions in present real data are shown in Fig. 6.9. There, WMAP 3-yr data alone and combined with small-scale CMB anisotropy data, supernovae and galaxy surveys are used to determine the marginalized distribution of $\Omega_{cdm}h^2$, $\Omega_b h^2$, n_s, the amplitude A_s and the optical depth τ for a flat cosmology with purely scalar perturbations. This is the simplest model (only five parameters) which currently seems to fit all the data satisfactorily.

6.5.3 Model comparison

Sometimes, we would like to compare two models with different parameter sets and decide which one of them is more probable given the data. At first, we certainly want to look at the likelihood functions and compare those. Models with a larger maximum likelihood fit the data better. But on the other hand, if model m_2 has many more parameters than model m_1, we are not surprised if m_2 fits the data better than m_1 and may still decide in favour of m_1 with the argument that m_1 is 'more physical' and more 'predictive' or, certainly, more economical. The latter criterion is often called 'Occam's razor': we should explain the data by the simplest possible model. Can these seemingly subjective criteria be made objective? We shall now see that under certain assumptions the answer is yes.

To be more specific let us consider two models which we want to compare; m_1 with parameters $\lambda^{(1)} = (\lambda_1^{(1)}, \ldots, \lambda_{M_1}^{(1)})$ and m_2 with parameters $\lambda^{(2)} = (\lambda_1^{(2)}, \ldots, \lambda_{M_2}^{(2)})$. We assume that the priors are fixed in both cases. (In Bayesian statistics the priors are just part of the game and cannot be ignored!) Bayes' theorem then gives us the probability of some set of parameters for given data D

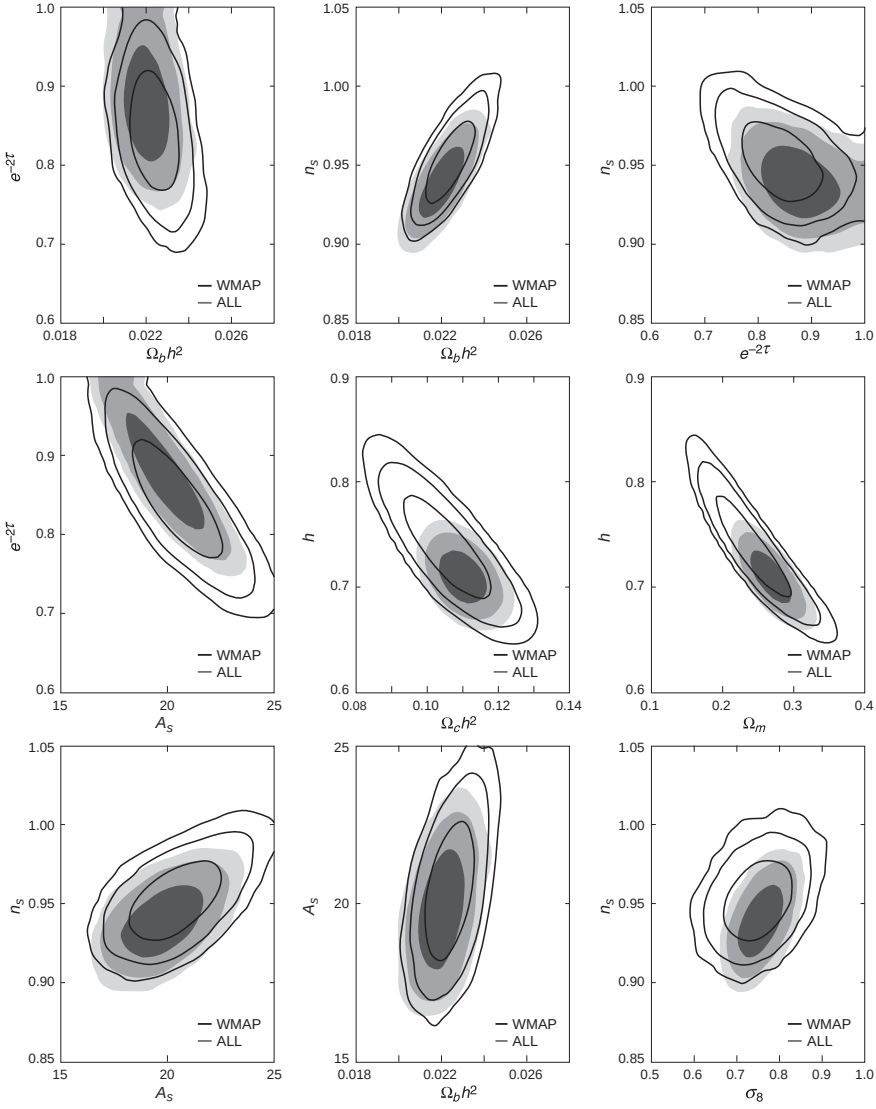


Fig. 6.9. The marginalized 2-parameter distributions from the WMAP experiment 3-yr data (black contour lines) and for WMAP combined with other CMB data, supernovae and large-scale structure data, for a flat cosmological model with purely scalar perturbations and no running. A_s is related to the amplitude of scalar metric perturbations at $k = 0.002\,\mathrm{Mpc}^{-1}$, $A_s = 10^{10}|\Psi(k = 0.002\,\mathrm{Mpc}^{-1})|^2 k^3$ and τ is the optical depth. Figure from Spergel *et al.* (2007).

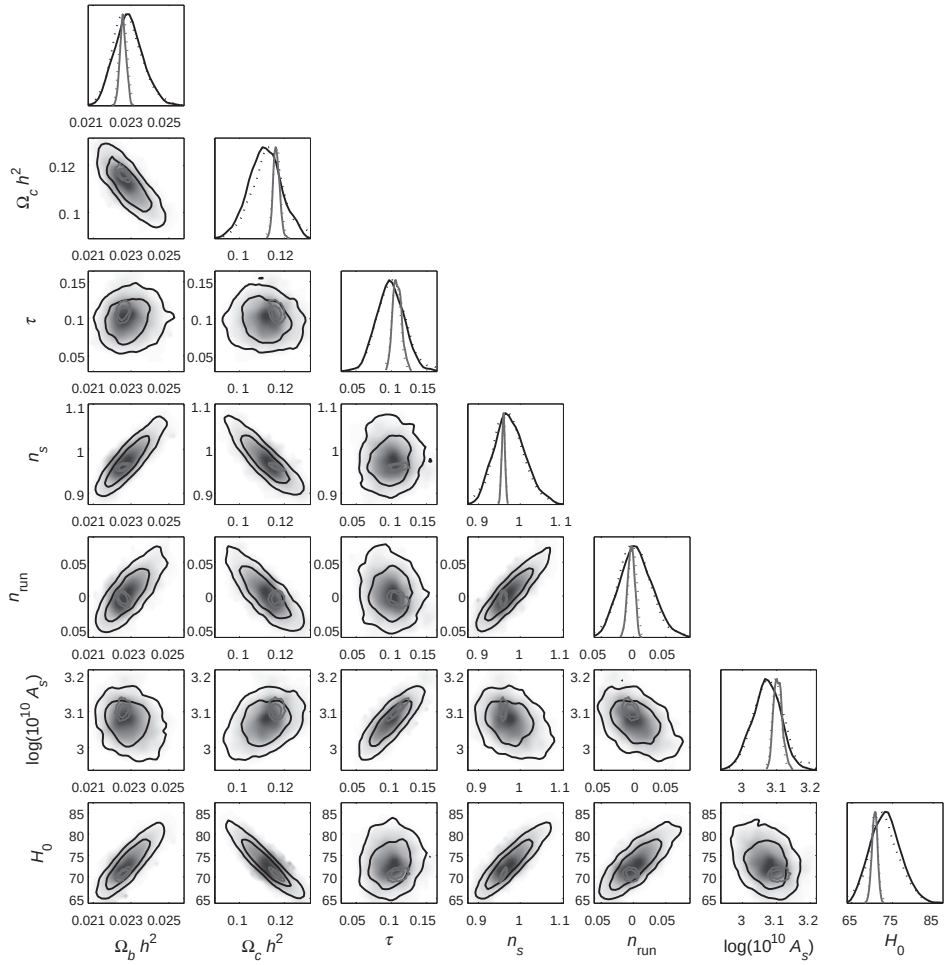


Fig. 6.10. The marginalized 2- and 1-parameter distribution forecasted for the Planck experiment, for a flat cosmological model with purely scalar perturbations and running are shown. From left to right the parameters are $\Omega_b h^2$, $\Omega_{cdm} h^2$, τ, n_s, $n_{\rm run}$, $\log(10^{10} A)$ and H_0. Figure from ESA (2005).

by

$$P[\lambda^{(1)}|D, m_1] = \frac{P[D|m_1, \lambda^{(1)}] P[\lambda^{(1)}|m_1]}{P[D|m_1]}\,, \tag{6.53}$$

$$P[\lambda^{(2)}|D, m_2] = \frac{P[D|m_2, \lambda^{(2)}] P[\lambda^{(2)}|m_2]}{P[D|m_2]}\,. \tag{6.54}$$

Our present notation indicates that our parameter choice $\lambda^{(1)}$ or $\lambda^{(2)}$ assumes the model m_1 or m_2 and, more importantly, that the evidence depends on the model under consideration. The prior is now of the form of a probability for the parameters $\lambda^{(1)}$ and $\lambda^{(2)}$ respectively, given the model m_1 and m_2 respectively. From Bayes'

theorem we also obtain

$$P[m_i|D] \propto P[D|m_i]P[m_i] \,. \tag{6.55}$$

Here, $P[m_i]$ is the total prior we assign to model m_i while $P[D|m_i]$ is simply the evidence from above. If we have no idea which model should be preferred, we can simply set $P[m_i] = \frac{1}{2}$. The probability of a model is then proportional to its evidence, and more importantly, their *ratio* is equal to the ratio of the evidence. But the evidence is given by the integral

$$P[D|m_i] = \int P[D|m_i, \lambda^{(i)}]P[\lambda^{(i)}|m_i]\, d^{M_i}\lambda^{(i)} \,. \tag{6.56}$$

For many problems the posterior $P[\lambda^{(i)}|D, m_i] = P[D|m_i, \lambda^{(i)}]P[\lambda^{(i)}|m_i]$ is strongly peaked around some best-fitting value $\bar{\lambda}^{(i)}$ with widths $\sigma^{(i)}(D) = (\sigma_1^{(i)}, \ldots, \sigma_{M_i}^{(i)})$ (let us assume, for simplicity, that the parameters are uncorrelated). We may then evaluate the integral (6.56) by multiplying the peak height with the width

$$\Sigma^{(i)}(D) = \Pi_{j=1}^{M_i} \sigma_j^{(i)} \,.$$

The evidence then becomes

$$P[D|m_i] \simeq P[D|m_i, \bar{\lambda}^{(i)}]P[\bar{\lambda}^{(i)}|m_i]\Sigma^{(i)}(D) \,. \tag{6.57}$$

Furthermore, let us assume that the prior of model i has some (large) total width $\Sigma^{(i)}$ and that $\bar{\lambda}^{(i)}$ is nicely inside the prior distribution. Then, due to normalization, $P[\bar{\lambda}^{(i)}|m_i] \simeq 1/\Sigma^{(i)}$ and the evidence for model m_i becomes

$$P[D|m_i] \simeq P[D|m_i, \bar{\lambda}^{(i)}]\frac{\Sigma^{(i)}(D)}{\Sigma^{(i)}} \,. \tag{6.58}$$

Hence the first guess, the maximum of the likelihood, is multiplied by the so-called 'Occam factor' $\Sigma^{(i)}(D)/\Sigma^{(i)}$, i.e., the ratio of the parameter space 'occupied by the data' divided by the volume of parameter space allowed by the prior. Models that are not predictive at all are penalized by a small Occam factor, and can in this way lose against a model even if they allow for a parameter choice $\bar{\lambda}^{(i)}$ with higher likelihood. Of course if the prior is constraining, but does not have a significant overlap with the posterior from the data, this will also disfavour a model. These situations are shown in the panels (a)–(c) of Fig. 6.11. As becomes clear from this example, for models which do allow a good fit to the data, the evidence strongly depends on the prior.

It is, however, important to note that the Occam factor enters the model probability just as a power law, while the offset of the measured parameter from the model prediction reduces the model probability exponentially.

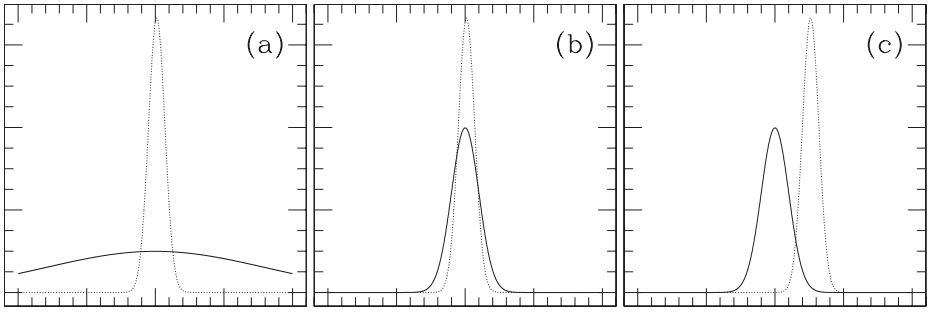


Fig. 6.11. Three 1-parameter models with their prior (solid) and posterior (dotted) distributions shown. They all have a comparable maximum likelihood. In (a) the prior distribution is much wider than the posterior. In models (b) and (c) the widths are comparable, but model (c) does not provide a good fit. The evidence for model (b) is by far the largest.

A situation which is often encountered in cosmology is the so-called 'nested models'. Two models are called 'nested' if one of them is obtained by fixing one or several parameters of the other model. For example, a model with vanishing tensor contribution is nested inside the more general models which allow for tensors, by setting the tensor to scalar ratio $r = 0$. Another example is the models which do not allow for curvature that are nested inside models with curvature. Let us concentrate on these to illustrate the significance of the prior for model selection.

As we shall see, present data yield the constraints $\Omega_K \simeq -0.01 \pm 0.03$. Does this mean that a universe with vanishing curvature is preferred? To discuss this let us approximate the marginalized posterior distribution of the parameter Ω_K by a Gaussian with width $\sigma_K = 0.03$. Cosmologist A says that the Universe might well have a curvature in the full range $\Omega_K \in [-1, 1]$ and therefore, models allowing for a curvature are penalized by an Occam factor of 0.03, hence models without curvature are strongly preferred.

Cosmologist B argues differently. She says that we know from inflation, which is in good agreement with all other observations, that curvature must be small, say $\Omega_K \in [-0.05, 0.05]$ and therefore there is no small Occam factor and models with non-vanishing but small curvature are as plausible as vanishing curvature.

This example makes it clear: for model selection the prior is crucial. It is of course a very different statement to say that since curvature is small it has little effect on all the other parameters and we therefore set it to zero in our analysis (which somewhat speeds up the CMB codes). This is a practical statement and does not address the problem of model selection.

If the posterior probability distribution is non-Gaussian, the approximations above no longer hold, and we have to resort to more complicated methods like MCMC to evaluate the posterior distribution. But the principal arguments of the

above discussion remain valid. For more details and more advanced methods of model selection see Kunz *et al.* (2006).

6.5.4 Markov chain Monte Carlo methods (MCMC)

If the posterior probability distribution of parameters is Gaussian, it is sufficient to give its mean and its width in the direction of the principal axis, i.e., the Fisher matrix, this contains all the statistical information. However, even if the distribution is Gaussian in the data, it is very often far from Gaussian in the cosmological parameters of interest to us.

Given our model with parameters λ and the data D, we can evaluate the probability density $P(\lambda) \equiv P[\lambda|D]$ up to a constant simply by Eq. (6.29),

$$P[\lambda|D] = P[D|\lambda]\frac{P[\lambda]}{P[D]}\,, \tag{6.59}$$

where usually, the distribution of the data, $D = (d_1, \ldots, d_N)$ is a Gaussian with some correlation matrix $\mathcal{C}$,

$$P[D|\lambda] = \mathcal{L}(D,\lambda)d^N d = \frac{1}{\sqrt{(2\pi)^N \det \mathcal{C}}} \exp\left(-\frac{d_i \mathcal{C}^{-1}_{ij} d_j}{2}\right) d^N d\,. \tag{6.60}$$

Therefore, once we have fixed a prior $P[\lambda]$, we can evaluate the probability $P[\lambda|D] \equiv p(\lambda)d^M\lambda$ of a given choice of parameters, $\lambda = (\lambda_1, \ldots, \lambda_M)$, up to a constant, the evidence $P[D]$. Let us define this not normalized distribution by

$$P^*(\lambda) \equiv P[\lambda|D]P[D] = P[D|\lambda]P[\lambda]\,. \tag{6.61}$$

We would like to answer the following questions.

- What is the shape of the probability density $p(\lambda)$ in the full parameter space, and what are the densities of some arbitrary subset of parameters, $\mu = (\lambda_{i_1}, \ldots, \lambda_{i_K})$, $K < M$ marginalized over all the other parameters. We are especially interested in the cases of $K = 1$ and 2, which are easy to visualize and which indicate how strongly the parameters λ_{i_1} and λ_{i_2} are correlated, see Figs. 6.9 and 6.10.
- We would also like to compute the expectation value and variance of derived parameters,

  $$\langle h \rangle = \int h(\lambda)p(\lambda)\,d^M\lambda\,,$$

  and

  $$\left\langle (h(\lambda) - \langle h \rangle)^2 \right\rangle = \int (h(\lambda) - \langle h \rangle)^2 p(\lambda)\,d^M\lambda\,,$$

  in brief, integrals of the form $\int f(\lambda)p(\lambda)\,d^M\lambda$.

If we have a *representative* (or *fair*) *sampling* of parameter space, $S = (\lambda^{(1)}, \lambda^{(2)}, \cdots, \lambda^{(R)})$, i.e., a sampling in which the number of points $\lambda \in S$ in a small volume V of parameter space is proportional to $\int_V p(\lambda)\, d^M\lambda$, we could approximate

$$\langle f \rangle = \int f(\lambda) p(\lambda)\, d^M\lambda \simeq \frac{1}{R} \sum_{i=1}^{R} f\left(\lambda^{(i)}\right) .$$

Furthermore, marginalized probability densities at $\mu = (\lambda_{i_1}, \ldots, \lambda_{i_K})$ would be proportional to the sum of all the points for which the parameters of interest are in a given infinitesimal volume around μ, i.e.,

$$p_{i_1 \cdots i_K} = \int \left(\Pi_{j \neq i_r} d\lambda_j\right) p(\lambda_1, \ldots, \lambda_M) .$$

In other words, the marginalized probability distribution of the parameters μ would just be given by the projection of the full probability distribution onto these parameters. Especially, the probability that the parameter λ_1 lies in some small interval I is proportional to the number of points which have $\lambda_1 \in I$ and all other parameters are arbitrary.

The above mentioned problems can therefore be solved, if we find a representative sampling of our parameter space. There are several methods for finding such a sampling which all have their advantages and disadvantages (see, e.g., Gamerman (1997) and MacKay (2003)). For high-dimensional problems, the Markov chain Monte Carlo methods (MCMC) are especially useful. Of these, we concentrate here on the Metropolis–Hastings algorithm which is dominantly in use for CMB analysis. We shall mention also the Hamiltonian Monte Carlo method. Full proofs that these algorithms really converge are found in the literature mentioned above.

Metropolis–Hastings algorithm

Let us start with some arbitrary point $\lambda^{(1)}$, and some 'proposal density' $Q\left(\lambda^{(2)}, \lambda^{(1)}\right)$ for a new value $\lambda^{(2)}$ which depends on $\lambda^{(1)}$. We call $\lambda^{(1)}$ the 'current point' and $\lambda^{(2)}$ the 'proposal'. For the moment, let Q be arbitrary but simple enough so that we can easily (with little numerical investment) sample it. We shall soon be more precise. To generate our sampling S we start at some arbitrary point $\lambda^{(1)}$. With probability $Q(\lambda, \lambda^{(1)})$ we now determine a proposal λ. To decide whether to accept this point as the next element of our sampling (which now becomes a chain, since it is ordered), we compute

$$r = \frac{P^*(\lambda) Q\left(\lambda^{(1)}, \lambda\right)}{P^*\left(\lambda^{(1)}\right) Q\left(\lambda, \lambda^{(1)}\right)} . \tag{6.62}$$

If $r \geq 1$, we accept λ as the next member of our chain, $\lambda^{(2)} = \lambda$, if $r < 1$ we assign $\lambda^{(2)} = \lambda$ with probability r and $\lambda^{(2)} = \lambda^{(1)}$ with probability $1 - r$. And so on, we generate our Markov chain, $S = (\lambda^{(1)}, \ldots, \lambda^{(R)})$.

If the proposal density Q is symmetric in its arguments, as for example a Gaussian centred on the current point, the factor $Q\left(\lambda^{(1)}, \lambda\right) / Q\left(\lambda, \lambda^{(1)}\right)$ drops out and we only have to calculate $P^*(\lambda)/P^*\left(\lambda^{(1)}\right)$. We now concentrate on this case which is often simply called the 'Metropolis algorithm'. The proposal density Q is then only needed to suggest the next point, but is not involved in the decision whether it is accepted or not.

It can be shown that for strictly positive Q, the distribution of the points in S always tends to the posterior distribution $P(\lambda) = P^*(\lambda)/P[D]$ for $R \to \infty$. How long will this take? Or in other words, how large do we have to choose R so that S becomes a fair sample of P? This is a difficult question, but it is relatively easy to find a lowest estimate for R. In order to have a reasonable acceptance rate r we do not want to choose a too large step size $\epsilon = |\lambda - \lambda^{(i)}|$. Hence the proposal density Q has to be sufficiently narrow. A reasonable step size is probably of the order of the smallest widths σ_s of the 1-parameter distributions (of course strictly speaking we do not yet know these widths, but in practice we can make an educated guess and revise it if necessary). Now, the distance the chain has to travel is at least equal to the width of the largest 1-parameter distribution, σ_l. Since our chain performs a random walk in parameter space, the number of steps it needs to travel a given distance is $R_{\min} \propto (\sigma_l/\sigma_s)^2$. The proportionality factor will be roughly the inverse of the mean acceptance probability, since, if the next point is not accepted, the chain does not move forward at all.

To obtain a reasonably fair sample, one certainly has to cover the high-probability part of the parameter volume several times and choose chain lengths of a few times $R_{\min}$. This all works, if our probability distribution has only one high-probability region. If it has several, even though, usually a given chain will rapidly find one of them, it is very hard to cross from one of these regions into another. Therefore, instead of generating just one chain, one usually generates several (tens of) chains. On the other hand, the first few (of order 20) points of a chain do not sample the posterior distribution but depend mainly on the random initial point $\lambda^{(1)}$. This 'burn in phase' is usually discarded. Furthermore, the points in the chain are not independent. As usual for Markov chains, each point depends on its predecessor. This is not really a hindrance, as the mathematical theorem shows, but nevertheless analysts often 'thin out' their chains, i.e., they use only every tenth or so point for the posterior distribution.

Since the 'burn in phase' of the chain is useless, it is not economical to use many very short chains. On the other hand, since the probability distribution can have several high-probability islands in parameter space which are separated by

deep canyons, it is not advisable to use only one very long chain. As so often, the 'golden middle' of many reasonably long chains is usually the best. However, if the likelihood function turns out to have more than one maximum to which chains may converge, care is required. The probability of a maximum is not simply proportional to the number of chains which converge to it. This number can be large, not because the maximum is large, but simply because its 'basin of attraction' is large. In this case, one has to analyse the shape of the likelihood function in more detail.

It is important to have a relatively good guess for the proposal density to start with. A simple possibility is a Gaussian with the correlation matrix given by the inverse of the Fisher matrix.

In practice, to test whether the chains have converged one just adds 10% more steps and investigates if the results change. In order to test whether the series of chains represents a fair sampling one adds a couple more independent chains and checks the effect on the results. If the results are not affected, or only well within error bars, one is usually confident that the procedure has converged. Often, one simply compares the variance of the parameters obtained from the chain with the error bars from the data.

In Figs. 6.9 and 6.10 we show the results obtained by this analysis method with the best currently available CMB data (WMAP 3-yr data) and for the Planck forecast.

There are several other Monte Carlo methods which people begin using for cosmological data sets like the slice sampling, MacKay (2003), but the principle behind them is always the same: to obtain a Markov chain which produces a representative sampling of an underlying probability distribution proportional to $P^*(\lambda)$, we have to find a transition probability $T(\lambda, \lambda')$ which leaves P^* invariant, i.e.,

$$P^*(\lambda) = \int T(\lambda, \lambda') P^*(\lambda')\, d^M\lambda' \,. \tag{6.63}$$

Choosing, as in the Metropolis algorithm, $T(\lambda, \lambda') = V_{\rm tot}^{-1} P^*(\lambda)/P^*(\lambda')$, is clearly a transition probability which obeys Eq. (6.63). Here $V_{\rm tot}$ is the total volume of parameter space which has to be introduced in order for T to be correctly normalized.

6.5.5 Hamiltonian Monte Carlo

Finally, let us briefly mention a method, which has not yet been widely used in cosmology (for a first attempt to study its usefulness for CMB data analysis, see Hajian (2007)). This method does not only use $P^*(\lambda)$, but also its gradient. Evaluating P^* with reasonable accuracy requires one run of a fast CMB code. Evaluating the gradient with respect to M parameters requires $2M$, or for a stable numerical derivative, $3M$ evaluations. As this is the costly part of CMB analysis, its usefulness has to be checked in detail. However, as we explain now, the Hamiltonian Monte Carlo

might reduce the number of points needed for a representative sampling from $R_{\min}$ to order $\sqrt{R_{\min}}$. Therefore, if the chains have difficulties in converging, it might be useful.

The basic idea is very simple. Let us write

$$P^*(\lambda) = e^{-V(\lambda)} .$$

Maximizing the probability is then equivalent to minimizing the potential V. Therefore, it is useful to take the next step in the direction $-\nabla V$. But this is exactly what Hamiltonian dynamics does. Therefore, we introduce momentum variables $\pi = (\pi_1, \ldots, \pi_M)$ and define the Hamiltonian

$$H(\lambda, \pi) = V(\lambda) + K(\pi) ,$$

where K is the kinetic energy, for example $K = \frac{1}{2}\pi^T\pi$. We then define the non-normalized probability

$$P_H^*(\lambda, \pi) = e^{-H(\lambda,\pi)} = e^{-V(\lambda)}e^{-K(\pi)} .$$

We now sample P_H^* in the following way. We first choose some initial value $\lambda^{(1)}$ and draw a momentum $\pi^{(1)}$ from the Gaussian distribution $e^{-K(\pi)}/Z_K$. For the next, dynamical proposal, the present momentum $\pi^{(1)}$ decides the displacement of λ and the gradient of the present potential $V(\lambda^{(1)})$ decides the change in the momentum, via the canonical equations

$$\dot{\lambda} = \pi , \quad \text{and} \quad \dot{\pi} = -\nabla V(\lambda) .$$

We then advance this system for some (fixed) number of steps to arrive at the proposal (λ, π), which is then accepted or rejected according to the Metropolis rule (6.62) for P_H^* (and some symmetrical Q which is no longer needed).[2]

The big advantage of this method, is that the distance covered by the parameters $\lambda^{(i)}$ is now proportional to the computer time per step and not only to its square root.

From the representative sampling of P_H^* obtained in this way, we obtain a fair sampling of $P^*(\lambda)$ by simply ignoring the momentum variables in the chains.

6.6 Degeneracies

To explain the problem of degeneracies, let us first consider parameters that obey a Gaussian distribution with some Fisher matrix $F(\bar{\lambda})$ at the maximum likelihood parameter $\bar{\lambda}$. Its inverse is the correlation matrix of the parameters,

[2] If the simulation is perfect, the proposal is accepted every time since $H = V + K$ is a constant of motion and so r is always equal to 1. In practice however, the inaccuracy in the numerical evaluation of the gradient will lead to some rejections.

$C_{ij} = \langle(\lambda_i - \bar{\lambda}_i)(\lambda_j - \bar{\lambda}_j)\rangle = F_{ij}^{-1}(\bar{\lambda})$. If one (or several) of the eigenvalues of the Fisher matrix is (are) very small, the variance of the parameters in the corresponding direction is very large. Hence this linear combination of parameters cannot be determined accurately. In the limiting case, when the eigenvalue vanishes, the linear combination which defines the eigenvector with vanishing eigenvalue is not at all constrained by the data.

Such directions are called degenerate or nearly degenerate directions. It is very useful to identify them and to express the results of the experiment in terms of quantities that are well determined and have small errors, i.e., that are orthogonal to the degenerate directions. Usually, degenerate directions have a simple physical interpretation. They can be lifted either by improving the data (if they are only nearly degenerate) or by considering complementary data. Here we discuss the main examples.

6.6.1 Curvature

The most prominent example of a degeneracy of CMB anisotropies comes from the fact, that they are strongly dependent on $\Omega_b h^2$, $\Omega_m h^2$ and the angular diameter distance D_A. Considering a ΛCDM model, only three combinations of the parameters of interest, Ω_m, Ω_b, Ω_Λ, h are well determined by CMB anisotropies. This is especially important for the determination of the curvature $\Omega_K = 1 - \Omega_\Lambda - \Omega_m$. Changes in the curvature which keep D_A, $\Omega_m h^2$ and $\Omega_b h^2$ fixed have nearly no effect on CMB anisotropies (see Fig. 6.12).

It is therefore, actually wrong if we say we can determine the curvature from CMB anisotropies. For this we need some additional information. The simplest way to break this degeneracy is, for example, to introduce a relatively strong prior on the Hubble constant. Using the HST key project result (Freedman *et al.*, 2001), $h = 0.71 \pm 0.7$, for example, we can determine the curvature to $\Omega_K = -0.02 \pm 0.02$. On the other hand, requiring $\Omega_K = 0$, reproduces a Hubble constant in perfect accord with the HST result. Another possibility is the inclusion of large-scale structure data, galaxy surveys. As we shall see in Section 6.7.3, these are mainly sensitive to the shape parameter $\Gamma \simeq \Omega_m h$. Together with the measurement of $\Omega_m h^2$ from the CMB this also determines h and thereby lifts the degeneracy. Also inclusion of supernova data, see Section 6.7.1 which measures the luminosity distance to much smaller redshifts, lifts the degeneracy.

6.6.2 Scalar spectral index, tensor component and related degeneracies

The tensor contribution to the CMB anisotropies only significantly adds to the C_ℓs with $\ell \lesssim 60$. For $\ell \gtrsim 60$ it rapidly decays and can be neglected (see Fig. 2.4). But a slight enhancement of power on large scales can also be produced by reducing

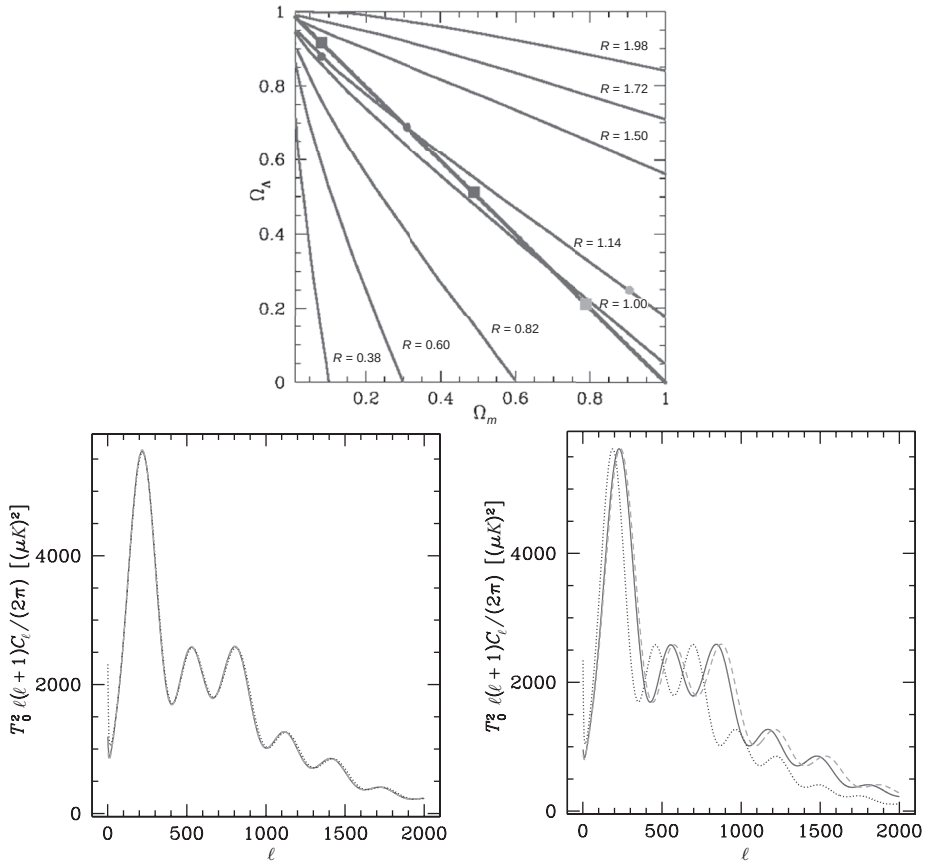


Fig. 6.12. In the upper panel lines of equal angular diameter distance are indicated. The number R is the ratio of the angular diameter distance of the model to the one of a concordance model with $\Omega_\Lambda = 0.7, \Omega_m = 0.3, h = 0.7$. The lines of constant curvature are parallel to the diagonal which is also drawn. In the lower left-hand panel we show CMB anisotropy spectra with $\Omega_K > 0$ (dashed), $\Omega_K < 0$ (dotted) and $\Omega_K = 0$ (solid), which have identical angular diameter distance, matter density and baryon density. They correspond to the dots indicated in the upper panel. The spectra overlay so precisely that we cannot distinguish them by eye. On the lower right-hand panel we show three spectra with curvature zero, identical matter density and baryon density, but with different angular diameter distances (the squares indicated in the upper panel on the $K = 0$ line). The spectra are significantly different.

somewhat the spectral index n_s. A slightly redder spectrum also has somewhat more power on large scales, see Fig. 6.13.

Another parameter that affects the power on large scales is the optical depth to the last scattering surface, τ. Enhancing it leads to more damping on larger scales which in turn can be compensated by a tensor component.

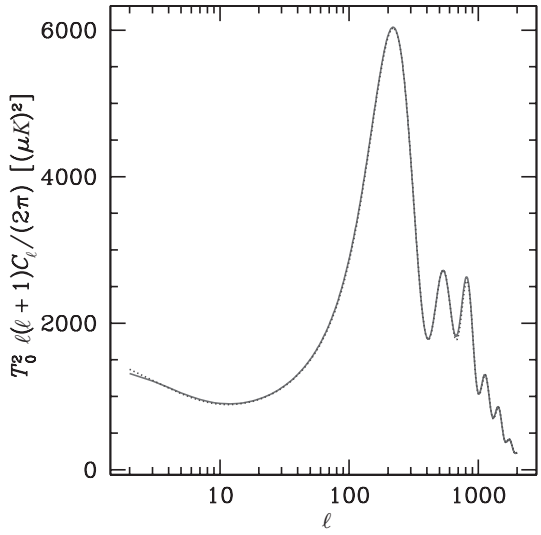


Fig. 6.13. A spectrum of purely scalar perturbations with $n_s = 0.96$ (solid) is compared to one with a tensor contribution of $r = 0.3$ (dotted). The cosmological parameters of the two models differ somewhat but are all within 1 to 2 sigma of the 'concordance values' given in Table 6.1: ($h = 0.73$, $\Omega_b h^2 = 0.0225$, $\Omega_m h^2 = 0.135$, $\tau = 0.1$, $n_s = 0.96$) for the purely scalar model and ($h = 0.8$, $\Omega_b h^2 = 0.023$, $\Omega_m h^2 = 0.118$, $\tau = 0.1$, $n_s = 1.0$, $r = 0.3$, $n_T = 0$) for the model with tensor contribution. Clearly, these two models cannot be distinguished from their temperature anisotropies.

This degeneracy can, in principle, be lifted by including high-quality polarization data. For example, reionization leads to the rescattering of photons at late times and induces a small amount of polarization on large scales which would not be seen in an ionized universe. The presently available polarization data from WMAP and other experiments, provides the best estimate of the optical depth τ.

On the other hand, a tensor component can be (nearly) unambiguously determined by a measurement of the B-mode of polarization. This is the next 'quantum leap' to be expected from CMB data: good B-polarization data which even might allow the testing of the slow roll consistency condition (3.87),

$$\frac{A_T}{A_S} = -\frac{18}{25} n_T \ .$$

6.6.3 Initial conditions

As we discussed in Chapter 3, simple inflationary models generate perturbations with adiabatic initial conditions. In the simplest case, neglecting a possible tensor component, the initial conditions are characterized by only two parameters.

However, in principle, other initial conditions are also possible. We have discussed the mixture of adiabatic and CDM iso-curvature perturbations. Also allowing for the two neutrino modes (the neutrino density and the neutrino velocity mode), one obtains four essentially different modes. Together with arbitrary correlations this gives a 4×4 symmetric, positive semi-definite matrix of initial conditions, hence ten parameters. The spectral indices of each component allow for additional parameters (see however Ex. 6.2).

Introducing these many additional parameters leads to serious additional degeneracies. Especially, the first peak of the acoustic oscillations for iso-curvature perturbations no longer determines the angular diameter distance to the last scattering surface, since it also depends strongly on the initial conditions. Even if the Hubble parameter is known, curvature can no longer be determined by CMB anisotropies alone.

However, iso-curvature perturbations, while contributing to the Sachs–Wolfe plateau, do not induce significant density fluctuations. But the normalization of CMB fluctuations on large scales gives about the right amplitude for density fluctuations in the purely adiabatic case. Therefore, combining CMB fluctuations with the galaxy power spectrum leads to stringent constraints for the iso-curvature contribution (Trotta *et al.*, 2001, 2003). An analysis (Trotta, 2006) with fixed spectral index n_S, comparing with the WMAP 3-yr data combined with small-scale CMB experiments, the HST prior for the Hubble parameter and the Sloan Digital Sky Survey (SDSS) (Tegmark *et al.*, 2004) for the matter power spectrum, infers that data are fully compatible with purely adiabatic perturbations and a possible iso-curvature fraction may not contribute more than about 20%.

Very roughly, iso-curvature modes have an effect similar to a tensor component: they contribute to the CMB anisotropies but not to the galaxy power spectrum. However, since they contribute to the CMB not only on very large scales, but also on smaller scales where the data have smaller error bars, they are better constrained than a tensor contribution.

Finally, polarization information will also help to break the degeneracies from iso-curvature modes (see Bucher *et al.*, 2001). This comes first of all from the fact that when several acoustic peaks are available, their relative position in ℓ-space, i.e., their distance, depends only on the cosmological parameters and not on the initial conditions.

Even when considering simple adiabatic perturbations, one can allow for additional features in the initial power spectrum that will influence the estimated cosmological parameters. As long as these are parametrized by a few numbers, we can test good enough data against them, but as mentioned in the beginning of this chapter, if we do not make any simplifying assumptions on the initial power spectrum, we must know the cosmological parameters which then fix the transfer

Table 6.1. *Joint likelihoods for a flat ΛCDM model with purely scalar perturbations. The WMAP 3-yr data are combined with small-scale CMB experiments (CBI + VSA, BOOMERanG) or galaxy survey data (2dFGRS). A is the amplitude of density fluctuations at $k = 0.002/Mpc$ and τ is the optical depth to the last scattering surface. The parameters σ_8 and Ω_m are derived. Table from Spergel et al. (2007).*

Parameter	WMAP only	WMAP +CBI+VSA	WMAP+ACBAR +BOOMERanG	WMAP + 2dFGRS
$100\Omega_b h^2$	$2.233^{+0.072}_{-0.091}$	$2.212^{+0.066}_{-0.084}$	$2.231^{+0.070}_{-0.088}$	$2.223^{+0.066}_{-0.083}$
$\Omega_m h^2$	$0.1268^{+0.0072}_{-0.0095}$	$0.1233^{+0.0070}_{-0.0086}$	$0.1259^{+0.0077}_{-0.0095}$	$0.1262^{+0.0045}_{-0.0062}$
h	$0.734^{+0.028}_{-0.038}$	$0.743^{+0.027}_{-0.037}$	$0.739^{+0.028}_{-0.038}$	$0.732^{+0.018}_{-0.025}$
A	$0.801^{+0.043}_{-0.054}$	$0.796^{+0.042}_{-0.052}$	$0.798^{+0.046}_{-0.054}$	$0.799^{+0.042}_{-0.051}$
τ	$0.088^{+0.028}_{-0.034}$	$0.088^{+0.027}_{-0.033}$	$0.088^{+0.030}_{-0.033}$	$0.083^{+0.027}_{-0.031}$
n_s	$0.951^{+0.015}_{-0.019}$	$0.947^{+0.014}_{-0.017}$	$0.951^{+0.015}_{-0.020}$	$0.948^{+0.014}_{-0.018}$
σ_8	$0.744^{+0.050}_{-0.060}$	$0.722^{+0.043}_{-0.053}$	$0.739^{+0.047}_{-0.059}$	$0.737^{+0.033}_{-0.045}$
Ω_m	$0.238^{+0.030}_{-0.041}$	$0.226^{+0.026}_{-0.036}$	$0.233^{+0.029}_{-0.041}$	$0.236^{+0.016}_{-0.024}$

function in order to determine the initial power spectrum. Without assumptions we cannot determine them both.

One often allows for a so-called 'running' of the spectral index. For this one fits for a 'running parameter' $\alpha = dn_s/d\ln(k)$ which is assumed to be constant. An initial spectrum, with running, is determined by three parameters. First one has to fix some 'pivot scale' k_* which is arbitrary, but has to be a scale where the power spectrum is well constrained by the data. One then sets

$$k^3 P_\Psi(k) = A_*(k/k_*)^{n_s - 1 + \alpha \ln(k/k_*)} \,. \tag{6.64}$$

The resulting amplitude A_* and spectral index n_s, in general, depend on the pivot scale. Only if $\alpha = 0$ does n_s not depend on k_*. If $\alpha = 0$ and $n_s = 1$ as well A_* is independent of k_*.

One can of course allow for more complicated features like one or several kinks in the power spectrum, i.e., sudden changes of the spectral index. There are inflationary models which lead to such predictions, for example models where inflation is driven by several scalar fields (Adams *et al.*, 1997).

A comparison of the cosmological parameters obtained by CMB data alone and CMB in conjunction with complementary data is given in Tables 6.1 and 6.2. At

Table 6.2. *As in Table 6.1 but including other data sets: galaxy surveys (SDSS), supernovae (SNLS and SN Gold) and weak lensing (CFHTLS). Table from Spergel et al. (2007).*

Parameter	WMAP+ SDSS	WMAP+ SNLS	WMAP+ SN Gold	WMAP+ CFHTLS
$100\Omega_b h^2$	$2.233^{+0.062}_{-0.086}$	$2.233^{+0.069}_{-0.088}$	$2.227^{+0.065}_{-0.082}$	$2.255^{+0.062}_{-0.083}$
$\Omega_m h^2$	$0.1329^{+0.0056}_{-0.0075}$	$0.1295^{+0.0056}_{-0.0072}$	$0.1349^{+0.0056}_{-0.0071}$	$0.1408^{+0.0034}_{-0.0050}$
h	$0.709^{+0.024}_{-0.032}$	$0.723^{+0.021}_{-0.030}$	$0.701^{+0.020}_{-0.026}$	$0.687^{+0.016}_{-0.024}$
A	$0.813^{+0.042}_{-0.052}$	$0.808^{+0.044}_{-0.051}$	$0.827^{+0.045}_{-0.053}$	$0.846^{+0.037}_{-0.047}$
τ	$0.079^{+0.029}_{-0.032}$	$0.085^{+0.028}_{-0.032}$	$0.079^{+0.028}_{-0.034}$	$0.088^{+0.026}_{-0.032}$
n_s	$0.948^{+0.015}_{-0.018}$	$0.950^{+0.015}_{-0.019}$	$0.946^{+0.015}_{-0.019}$	$0.953^{+0.015}_{-0.019}$
σ_8	$0.772^{+0.036}_{-0.048}$	$0.758^{+0.038}_{-0.052}$	$0.784^{+0.035}_{-0.049}$	$0.826^{+0.022}_{-0.035}$
Ω_m	$0.266^{+0.026}_{-0.036}$	$0.249^{+0.024}_{-0.031}$	$0.276^{+0.023}_{-0.031}$	$0.299^{+0.019}_{-0.025}$

the present level of accuracy, complementary datasets do not reduce the error bars considerably, but it is important that they are consistent.

6.7 Complementary observations

CMB observations are not the only cosmological observations at our disposal. The main reason why they are so useful is that they are relatively easy to calculate to good accuracy. On a wide range of scales, we do not expect any complicated physics to obscure the relation between data and theory. Nevertheless, it would be a waste not to also consider other available data and, especially in view of the degeneracies, we need other data to confidently interpret the CMB. Here we only briefly introduce the most important complementary observations.

6.7.1 The Hubble parameter $H(z)$

A notoriously difficult quantity to measure is not only the function $H(z)$, but more basically the value of the present Hubble parameter, $H_0 = H(0)$. The main difficulty lies in the measurement of cosmological distances. The history of astrophysics and cosmology is marked by repeated underestimations of distances. E. Hubble originally overestimated his parameter by a factor of about 7 (Hubble, 1929). Even though it is relatively straight forward to measure the redshift of a cosmological source, how can we find its distance? The main tools are standard candles or standard

rulers. If we know the intrinsic size or luminosity of a distant source, we can use this to determine its angular diameter or luminosity distance. To lowest order in the redshift z, these are simply $H_0^{-1}z$. For redshift higher than $z \simeq 0.2$, one has to take into account the full expression for the distance as derived in Chapter 1. The full function $H(z)$ determines not only the present Hubble parameter but, via the Friedmann equation, also the matter content and the curvature of the Universe.

At present, the most promising method is to determine $H(z)$ with the observation of supernovae of type Ia. These are supernovae without hydrogen lines. They are extremely luminous and can be seen out to redshift of 2 and maybe more. The idea is that they come from white dwarfs which accrete material, e.g., from a companion star, until they pass over the Chandrasekhar mass limit of about $1.4\,M_\odot$ (Chandrasekhar, 1939). At this moment they become unstable and explode. Most probably this leads to the formation of a neutron star. The intrinsic luminosity of this explosion is quite constant and the mild variation is strongly correlated with the width of the light curve. Correcting for this variation with a phenomenological formula, one can obtain very small variations in the corrected intrinsic luminosity (about 0.1 magnitude). This allows a very accurate measurement of the luminosity distance to these explosions.

At present, luminosity distances to supernovae with redshifts up to 1.7 have been determined (Astier *et al.*, 2006 and Wood-Vasey *et al.*, 2007). These are used not only to measure H_0 (performed with HST (Freedman *et al.*, 2001; Sandage *et al.*, 2006)) but especially to determine $H(z)/H_0$ which can be obtained with much better accuracy. These measurements have provided the first clear indication that the expansion of the Universe is currently dominated by a cosmological constant or some form of dark energy with strong negative pressure leading to acceleration (see Chapter 1). They are especially important in our aim to reveal the nature of dark energy. Is it simply a cosmological constant with $w = -1$ or is it dynamical e.g. a scalar field with a time-dependent equation of state, $w(z)$? Or is it even some 'phantom matter' with $w < -1$?

Unfortunately, the luminosity distance only directly measures the integral (see Chapter 1, Eqs. (1.39) and (1.51))

$$D_L(z) = \frac{(1+z)\chi}{\sqrt{|\Omega_k|}H_0}\left(\int_0^z \frac{\lambda dz}{(1+z)H(z)}\right).$$

The equation of state parameter w on the other hand enters on the level of the derivative $H'(z)$. To determine it, two derivatives from relatively noisy data have to be taken; a very difficult task. Recently, it has been shown that the dipole of the luminosity distance which is, like the CMB dipole, due to our motion with respect to the Friedmann background, allows a direct measure of $H(z)$ so that only one additional derivative is needed to arrive at the equation of state (Bonvin *et al.*, 2006).

On the other hand, to accurately determine the amplitude of the dipole, many SNIas in a given redshift bin are needed. It remains to be seen whether this new approach will bear fruit.

6.7.2 Nucleosynthesis

As we have seen in Chapter 1, by an analysis of what has happened during nucleosynthesis, i.e. at $T \simeq 0.1$ MeV and $z \simeq 2.3 \times 10^9$ we can calculate the light element abundance as a function of the baryon density, $\Omega_b h^2$, see Fig. 1.10. Comparing with the observations of these abundances yields the nucleosynthesis value of the baryon density. How to estimate the primordial abundance from the present abundance of light elements is entirely non-trivial, an art which we do not discuss here any further. An average, relatively conservative estimate gives (Particle Data Group, 2006),

$$0.017 \leq \Omega_b h^2 \leq 0.024 \quad \text{(at 95\% confidence)} \ . \tag{6.65}$$

The agreement of this result with the CMB estimate is most remarkable. Both values are based on completely different physics. Such agreements give us confidence in the standard cosmological model.

The abundance of helium generated during nucleosynthesis is very sensitive to the number of relativistic degrees of freedom at the time of nucleosynthesis, which determines the expansion rate during nucleosynthesis. Assuming the photon and three types of neutrinos (at their somewhat lower temperature) give a good fit to the observed helium abundance. The nucleosynthesis constraint on the content of relativistic particles at the time of nucleosynthesis is usually formulated as a constraint on the number N_ν of light neutrino species. The data require (Particle Data Group, 2006)

$$N_\nu = 3.24 \pm 1.2 \quad \text{(at 95\% confidence)}. \tag{6.66}$$

In very good agreement with the value from the width of Z-decay obtained at accelerators (Particle Data Group, 2006),

$$N_\nu = 2.99 \pm 0.016 \ .$$

Even though the cosmological result is older, the accelerator result has become much more accurate.

6.7.3 The galaxy distribution

When determining the scalar CMB anisotropies we also calculate the matter power spectrum of linear perturbations, $P_D(k)$. Unfortunately, we cannot observe this power spectrum directly. First of all, on small scales, $k \gtrsim 2\pi/(30\,h^{-1}\,\text{Mpc}) \simeq$

$0.2\,h$ Mpc^{-1}, the spectrum is affected by non-linear clustering. On these scales the fluctuations in the matter density are of order unity and larger. Therefore, they cannot be accurately determined by linear perturbation theory. Non-linear perturbation theory or numerical simulations are needed which are much less accurate. While linear perturbation calculations are easily made with accuracies of better than 1%, it is very difficult to achieve an accuracy of 10% with numerical N-body simulations.

Second, we cannot observe matter directly. We just observe luminous galaxies and it is unclear how these are related to the underlying matter distribution. This problem is usually 'encoded' in the so-called bias b, which is often set to a constant, but most probably depends on scale. Nevertheless, the galaxy power spectrum is usually a good measure of the so-called 'shape parameter' Γ which determines the scale k_e, at which the linear power spectrum turns from the $P_D \propto k^{n_s}$ behaviour on large scales to the $P_D \propto k^{n_s-4}$ behaviour on small scales. Even though neither of the two asymptotic behaviours has ever been observed in a galaxy power spectrum, the position of the kink is relatively well established and yields a shape parameter $\Gamma \simeq 0.2$. The shape parameter is defined as the real position of the kink k_e compared to k_{m1}, the position of the kink in a universe with $\Omega_m = h = 1$,

$$k_e = k_{m1}/\,\Gamma\ . \tag{6.67}$$

The kink position is given by the horizon scale at matter radiation equality and one easily shows that $\Gamma \simeq \Omega_m h$, see Ex. 6.3. Assuming that the bias 'function' does not have any structure in the neighbourhood of the kink, the position of the latter allows us to determine $\Omega_m h$.

The reason why the power spectrum does not decay like k^{n_s-4} on small scales (large k) is non-linear clustering. On scales smaller than the kink, the fluctuations have become sufficiently large for non-linear clustering to become relevant, before the slope no longer has its asymptotic value from linear perturbation theory.

The question of why the power spectrum is not seen to rise like k^{n_s} on large scales has different answers. The first is simply that we have not measured the power spectrum on large enough scales to see this asymptotic behaviour. An additional difficulty is that a galaxy power spectrum, coming from a distribution of points, always has a white noise component. On sufficiently large scales, this white noise, being a constant, will always win over a nearly scale-invariant spectrum, $\propto k^{n_s}$ with $n_s \simeq 1$. The relevant question is: what is the amplitude of the white noise component? There are arguments, that it may be so high that we shall never be able to observe the Harrison–Zel'dovich part, $P_D \propto k^{n_s}$ of the matter spectrum with the observation of galaxies (Durrer *et al.*, 2003).

The mean matter fluctuation inside a ball of radius R is given by

$$\langle(\Delta M/M)^2\rangle_R = \left\langle \left(\int W_R(\mathbf{x}-\mathbf{y}) D(\mathbf{y})\, d^3y \right)^2 \right\rangle . \tag{6.68}$$

Here W_R is a window function of size R and D is the density contrast. The most common shapes for window functions are a Gaussian or a 'top hat'. Using the fact that the Fourier transform of a convolution is the product of the Fourier transforms, we have

$$\int W_R(\mathbf{x}-\mathbf{y}) D(\mathbf{y})\, d^3y = \int \frac{d^3k}{(2\pi)^3}\, e^{i\mathbf{k}\cdot\mathbf{x}} W_R(k) D(k) ,$$

so that

$$\langle(\Delta M/M)^2\rangle = \int \frac{d^3k}{(2\pi)^3}\, |W_R(k)|^2\, P_D(k) . \tag{6.69}$$

If the bias is relatively close to unity, the average fluctuations inside a ball of radius R can therefore give the normalization of the matter power spectrum. One usually defines

$$\sigma_8^2 = \int \frac{d^3k}{(2\pi)^3} W^2_{8\,h^{-1}\,\mathrm{Mpc}}(k) P_D(k) , \tag{6.70}$$

the mean fluctuation inside a ball of radius $R = 8\,h^{-1}$ Mpc. Of course, this quantity can only be measured from the galaxy power spectrum if the bias is known. However, within linear perturbation theory it is readily determined by the CMB. The parameters which enter are mainly the CMB normalization, A_s, and the spectral index, n_s. Current CMB data yield $\sigma_8 \simeq 0.74 \pm 0.05$.

The matter power spectrum is also very sensitive to the presence of light massive neutrinos in the dark matter. Such a 'hot dark matter' component, containing relativistically fast particles, leads to damping by free streaming on small scales. Therefore, hot dark matter suppresses power on small scales. This suppression is very strong and can be severely limited even with only 10% accurate N-body simulations. Comparisons of data and simulations require $m_\nu < 1$ eV.

The best currently available galaxy power spectra come from the 2dF Galaxy Redshift Survey (Cole *et al.*, 2005) and the Sloan Digital Sky Survey (Tegmark *et al.*, 2004). The SDSS power spectrum is shown in Fig. 6.14.

Finally, the matter power spectrum contains acoustic oscillations which come from the baryon acoustic oscillations before decoupling. If galaxies trace matter closely enough, these oscillations are also visible in the galaxy power spectrum. This seems to be the case and first reports on the identification of these features in observed galaxy distributions have been made (Eisenstein *et al.*, 2005). Since the

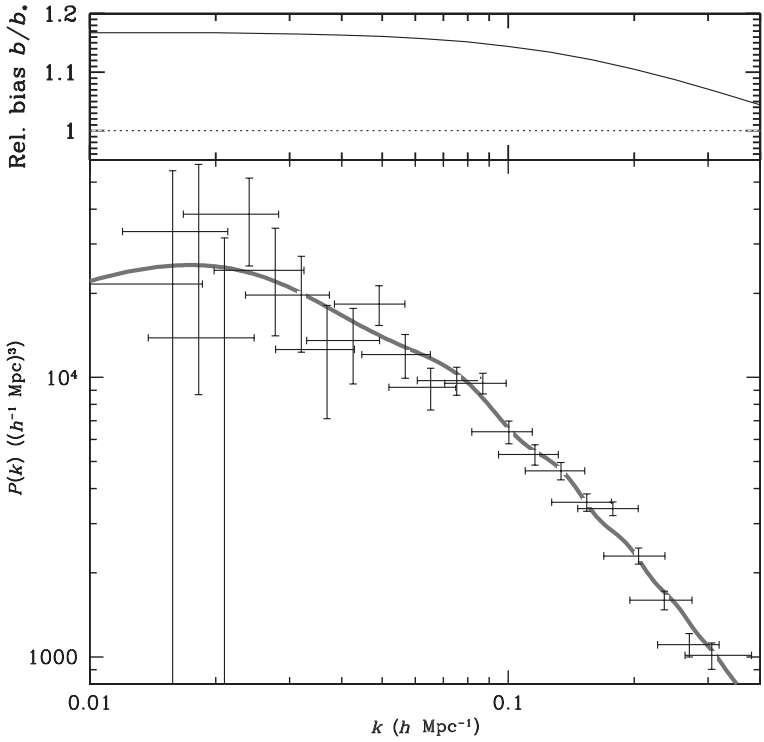


Fig. 6.14. The galaxy power spectrum from the SDSS corrected by the modelled bias (top) is shown. The solid line is the best-fitting linear ΛCDM model (from Tegmark *et al*., 2004).

scale of the oscillations can be calculated by linear perturbation theory, they provide a standard ruler which can be used to determine the angular diameter distance to the mean redshift of the sample. Once we have large galaxy catalogues with many different redshifts at our disposition, the acoustic peaks in the matter distribution can (in principle) lead to an evaluation of $D_A(z)$ and thereby $H(z)$ in a way similar to the supernova observations but completely independent of these.

6.7.4 Measuring the late integrated Sachs–Wolfe effect by correlations of CMB anisotropies with density fluctuations

The integrated Sachs–Wolfe effect is the term

$$\left(\frac{\Delta T}{T}\right)_{\rm ISW}(\mathbf{x}_0, \mathbf{n}) = 2\int_{t_{\rm dec}}^{t_0} \dot{\Psi}(x_0 - \mathbf{n}(t - t_0), t)\, dt \tag{6.71}$$

in the expression for the temperature anisotropy given in Eq. (2.231). The overdot in $\dot{\Psi}$ denotes the partial derivative with respect to conformal time t. We have set

$\Pi \simeq 0$ so that $\Psi = \Phi$. The Bardeen potential is determined by the Einstein equation (2.104). Neglecting curvature this is

$$\Psi = -\frac{4\pi G}{k^2} a^2 \rho D = -\frac{3H_0^2 \Omega_m (1+z)}{2k^2} D \,. \tag{6.72}$$

For the second equals sign we have assumed that ρ comes from pressureless matter with density parameter Ω_m, so that $\rho \propto a^{-3}$. We have set $a_0 = 1$. As we have seen in the examples treated in Chapter 2, in a purely matter dominated universe $D \propto a = (1+z)^{-1}$ and therefore $a^2 \rho D =$ constant, so that there is no integrated Sachs–Wolfe effect (ISW). This is different at relatively early times, $t \sim t_{\rm dec}$ where the radiation content cannot be neglected, and also at very late times if there is either curvature or a cosmological constant that becomes relevant.

The late ISW effect leads to a correlation between matter density fluctuations and the CMB temperature fluctuations on large scales. This is already evident from Eq. (6.72). At any fixed time, $\Psi(\mathbf{k}, t)$ and $D(\mathbf{k}, t)$ are perfectly correlated since they differ by a deterministic multiplicative function. However, most of the CMB anisotropies actually measure Ψ (and D) at $t_{\rm dec}$, a time at which we can by no means measure the matter power spectrum directly. This is different for the late ISW effect which measures Ψ at late times, when the cosmological constant becomes relevant, $z \lesssim 1$. At these times we can also observe the galaxy distribution and infer from it the matter distribution.

Let us estimate the signal obtained from the correlation of the matter fluctuations at some fixed redshift z with the CMB. The density fluctuation at z in a direction $\mathbf{n}$ from us is

$$D(\mathbf{x}_0, \mathbf{n}, z) = D(x_0 - \mathbf{n}(t(z) - t_0), t(z)) \,.$$

Here $\mathbf{x}_0$ is our position, t_0 denotes today and $t(z)$ is the conformal time at redshift z. Expressed in terms of the Fourier transform $D(\mathbf{k}, t(z))$ we obtain

$$\begin{aligned} D(\mathbf{x}_0, \mathbf{n}, z) &= \frac{1}{(2\pi)^3} \int d^3k \, e^{-i\mathbf{k}\cdot(\mathbf{x}_0 - \mathbf{n}(t(z)-t_0))} D(\mathbf{k}, t(z)) \\ &= -\frac{2}{3H_0^2 \Omega_m (1+z)(2\pi)^3} \int d^3k \, k^2 e^{-i\mathbf{k}\cdot(\mathbf{x}_0 - \mathbf{n}(t(z)-t_0))} \Psi(\mathbf{k}, t(z)). \end{aligned} \tag{6.73}$$

We want to correlate these density fluctuations with the ISW effect,

$$\left\langle D(\mathbf{x}_0, \mathbf{n}', z) \left(\frac{\Delta T}{T}\right)_{\rm ISW} (\mathbf{x}_0, \mathbf{n}) \right\rangle = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell^{(D)}(z) P_\ell(\mathbf{n}\cdot\mathbf{n}') \,. \tag{6.74}$$

For the second term we insert Eq. (6.71) with its Fourier transform

$$\left(\frac{\Delta T}{T}\right)_{\rm ISW} (\mathbf{x}_0, \mathbf{n}) = \frac{2}{(2\pi)^3} \int d^3k \int_{t_{\rm dec}}^{t_0} dt \, e^{-i\mathbf{k}\cdot(\mathbf{x}_0 - \mathbf{n}(t-t_0))} \dot{\Psi}(\mathbf{k}, t) \,.$$

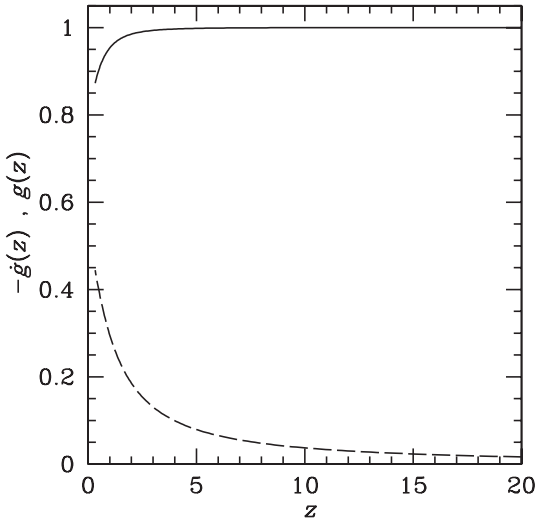


Fig. 6.15. The growth function g (solid) and its derivative, $\dot{g}$ (dashed) are shown as functions of redshift in a universe with $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$. Note that only for $z \lesssim 2$, the growth function g starts to deviate significantly from 1.

We now set

$$\Psi(\mathbf{k}, t) = g(t, k)\Psi_{\rm in}(\mathbf{k}) \,. \tag{6.75}$$

For scales which enter the horizon only in the matter dominated era, $g(t, k)$ is nearly independent of k and constant in time for $t < t_1$, where t_1 denotes the time when the cosmological constant becomes relevant, say $\Omega_\Lambda(t_1) \simeq 0.1$. After t_1, linear density fluctuations grow more slowly than the scale factor. In Fig. 6.15 we plot $g(z)$ and $\dot{g}(z)$ for a universe with $(\Omega_\Lambda, \Omega_m) = (0.7, 0.3)$.

We make use of $\dot{\Psi} = \dot{g}\Psi_{\rm in}$ and $\langle \Psi_{\rm in}(\mathbf{k})\Psi^*_{\rm in}(\mathbf{k}')\rangle = (2\pi)^3\delta(\mathbf{k}-\mathbf{k}')P_\Psi(k) = (2\pi)^3\delta(\mathbf{k}-\mathbf{k}')A_S^2(k/H_0)^{n_s-1}k^{-3}$, see Eq. (2.242). Furthermore, we rewrite $e^{i\mathbf{kn}(t_0-t)}$ in terms of spherical Bessel functions and Legendre polynomials (see Eq. (A4.101)). Applying the addition theorem for spherical harmonics, Eq. (A4.42), the integration over directions in $\mathbf{k}$-space yields

$$C_\ell^{(D)}(z) = \frac{-8g(t(z))}{3\pi\,\Omega_m H_0^2(1+z)} \int dk\, k^4 P_\Psi(k)\, j_\ell(k(t_0 - t(z))) \times \int_{t_{\rm in}}^{t_0} dt\, \dot{g}(t)\, j_\ell(k(t_0 - t)) \,. \tag{6.76}$$

Eq. (6.76) is still exact. For a given initial power spectrum $k^3 P_\Psi(k) = A^2(kt_0)^{n_s-1}$, the transfer function $g(t)$ is determined by the cosmological parameters Ω_m and Ω_Λ. If $\Omega_m = 1$ and $\Omega_\Lambda = 0$, $g =$ constant and the effect vanishes.

The kernel for the k-integration,

$$K(k,\ell,z) \equiv j_\ell(k(t_0 - t(z))) \int_{t_{\rm dec}}^{t_0} dt\, \dot{g}(t)\, j_\ell(k(t_0 - t)) \,,$$

is largest, when $j_\ell(k(t_0 - t))$ has its first peak in the interval from t_1 to t_0 during which $\dot{g}$ is considerable. Hence for values of $\ell \lesssim k(t_0 - t_1) \lesssim \ell + \pi$. In order for $j_\ell(k(t_0 - t(z)))$ to also be large in this interval, we must also require $\ell \lesssim k(t_0 - t(z)) \lesssim \ell + \pi$. Since $t_1 \sim t_0/2$, this can only be achieved if $t(z) \gtrsim t_1$. The kernel then peaks roughly at

$$k_{\rm max} \sim \frac{\ell}{t_0 - t(z)} \,, \qquad \text{with a width} \,, \quad \Delta k_{\rm max} \sim \frac{\pi}{t_0 - t(z)} \,. \tag{6.77}$$

Its value at this position can be approximated by $K(k_{\rm max}, \ell, z) \sim \ell^{-2}$. Here we use the relatively crude approximation for the maximum of spherical Bessel functions[3] $j_\ell(x_{\rm max}) \sim 1/\ell$. Inserting this in the expression (6.76), we obtain

$$\begin{aligned} C_\ell^{(D)}(z) &\sim -\frac{8}{3\pi\,\Omega_m H_0^2(1+z)} \frac{\Delta k_{\rm max} k_{\rm max}}{\ell^2} A_S^2 (k_{\rm max}/H_0)^{n_s - 1} \\ &\simeq -\frac{8(t_0 - t(z))^{-2} A_S^2}{3\Omega_m H_0^2 (1+z)\ell} \left(\frac{\ell}{H_0(t_0 - t(z))} \right)^{n_s - 1} \,. \end{aligned} \tag{6.78}$$

If z is sufficiently small, we may approximate $z(t)$ by a Λ-dominated expansion, $z + 1 \sim \exp(H_0(t_0 - t(z)))$, so that $H_0(t_0 - t(z)) \sim \ln(z+1) \sim z$. For the last approximation we used $z \ll 1$ in order for Λ-domination to be a reasonable approximation. With these approximations we can estimate

$$\ell^2 C_\ell^{(D)}(z) \sim -\frac{8\ell A_S^2 (\ell/z)^{n_s - 1}}{3\Omega_m (1+z) z^2} \,, \quad z \ll 1 \,. \tag{6.79}$$

This formula seems to diverge for $z \to 0$. This is due to the fact that $D \propto k^2 \Psi$ grows for large wave numbers. But our result is of course only valid if $k_{\rm max}$ is well within the linear regime. This means $k_{\rm max} \lesssim 0.1\, h\,{\rm Mpc}^{-1}$. Inserting Eq. (6.77) for $k_{\rm max}$ this becomes $\ell H_0/z \lesssim 0.1\, h\,{\rm Mpc}^{-1}$, or, with $H_0 = h/(3000\,{\rm Mpc})$, $\ell/300 \equiv z_{\rm min} \lesssim z$. Hence, for a nearly scale-invariant spectrum, $n_s \simeq 1$, $\ell^2 C_\ell^{(D)}(z)$ increases linearly with ℓ, but also the minimum redshift $z_{\rm min}$ for which the result is valid increases linearly with ℓ, so that $\ell^2 C_\ell^{(D)}(z_{\rm min}(\ell))$ decreases linearly with ℓ. Since the data suffer less from cosmic variance for larger ℓs, a good compromise is $z \sim 0.2$ which allows the measurement of $C_\ell^{(D)}(z)$ up to $\ell_{\rm max} \sim 0.2 \times 300 = 60$.

[3] A better approximation (Abramowitz & Stegun, 1970) would be $j_\ell(x_{\rm max}) \sim 0.84/\ell^{5/6}$, but since the subsequent uncertainties are rather large, it is not worth it to complicate our formulae by using it.

Measurements of these correlations need large-scale CMB anisotropies and density fluctuations in a relatively narrow redshift interval. Of course the result over a broader redshift interval is easily obtained by integrating Eq. (6.76) over redshift.

So far, several tentative detections at the $(2-3)\sigma$ level have been reported e.g., in Padmanabhan *et al.* (2005) and Pietrobon *et al.* (2006).

6.7.5 The Ly-α forest

The light from a high-redshift quasar, on its way from emission into our detector, not only propagates through the reionized intergalactic medium, but also crosses through clouds of hydrogen. These are regions where the matter density is relatively high and baryons are relatively cool so that at least some of them have recombined into hydrogen. These hydrogen clouds are not considered as isolated 'proto-galaxies', but simply as regions of relatively high density, where collisions that allow cooling processes can occur. When quasar light passes through them, the Lyman-α photons are absorbed by the neutral hydrogen, leading to a 'forest' of absorption lines in the quasar spectrum, see Fig. 6.16.

This Ly-α forest is related to the one-dimensional distribution of neutral hydrogen which is in turn related to the matter power spectrum. The depth of the lines is a measure of the hydrogen density. The observations are usually presented in terms of the power spectrum of the transmitted flux fraction, $P_F(k, z)$, as $F(\lambda) = \exp(\tau(\lambda))$, where $\tau(\lambda)$ is the optical depth to Ly-α averaged over the scale λ.

To relate this to the matter density power spectrum $P_D(k, z)$, we have to make assumptions about cooling and recombination, see e.g. McDonald *et al.* (2005). With such a relation at hand (which is usually non-linear, given by hydrodynamical simulations), the correlations of the lines provide, in principle, a measure of the matter power spectrum. If the underlying density fluctuations were still linear, we could relate them to the initial fluctuation by some deterministic growth factor $T(z)$ via $D(\mathbf{x}, z) = T(z)D_{\rm in}(\mathbf{x})$. The growth factor depends only on the background cosmology, i.e. on the cosmological parameters. The correlation between linear density fluctuations in some fixed direction $\mathbf{n}$ at redshifts z_1 and z_2 is simply

$$\begin{aligned}\langle D(z_1)D(z_2)\rangle &= \langle D\left(\mathbf{x}_0 - \mathbf{n}(t_0 - t(z_1)), t(z_1)\right) D\left(\mathbf{x}_0 - \mathbf{n}(t_0 - t(z_2)), t(z_2)\right)\rangle \\ &= T(z_1)T(z_2)C_{\rm in}(|t(z_1) - t(z_2)|)\,, \end{aligned} \tag{6.80}$$

where $C_{\rm in}$ is the initial correlation function, i.e. the Fourier transform of the initial power spectrum. Typically $z_1, z_2 \sim 2$–3 and $\Delta z = |z_1 - z_2| < 0.01$, so that $|t(z_1) - t(z_2)| \sim t(z)(\Delta z/2(z+1))$. Here z is the mean redshift and we have approximated $z + 1 \propto t^{-2}$, which gives the right order of magnitude for $z > 1$. These length scales

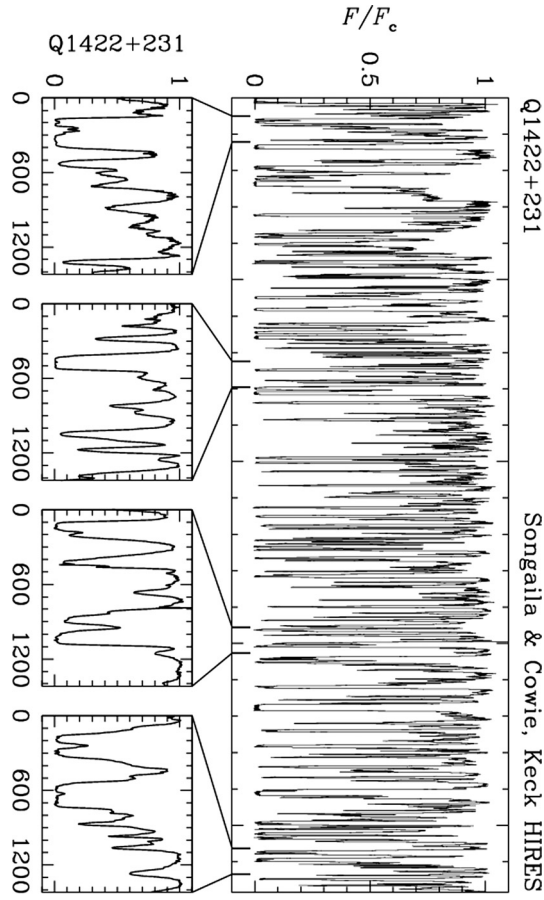


Fig. 6.16. The Ly-α forest region in the Keck HIRES spectrum of the Quasar QSO $1422 + 231$ at $z = 3.61$. From Songaila & Cowie (1996).

are small, hence the Ly-α forest explores the power spectrum at small scales. For example, for $z = 2.5$ we have $t(z) \simeq 1500\,h^{-1}$ Mpc so that we find $|t(z) - t(z + \Delta z)| \simeq 1.5\,h^{-1}$ Mpc for $\Delta z = 0.01$.

On the other hand, at $z \sim 2$–3 perturbations on these scales are already non-linear, so that time evolution and correlation do not simply factorize as in Eq. (6.80). To use the Ly-α forest for parameter estimation, we have to rely on N-body simulations. The main sensitivity of the Ly-α forest is then the amplitude of fluctuations on a small scale, which is sensitive to both, σ_8 and the spectral index n_s, as well as to a possible contribution of massive neutrinos.

Due to the complicated physics involved in the conversion from absorption lines which determine the flux power spectrum to the linear density fluctuations, it is difficult to estimate possible systematic errors in the inferred parameters. At

present, it is reassuring, that parameter estimations using the Ly-α forest are in the same bulk part as other estimates.

If the theoretical difficulties can be overcome, the Ly-α forests of quasars provide a very interesting data set. They are our most promising tool for estimating the linear power spectrum on small scales. In that sense they help enormously to extend the 'lever arm' of our knowledge on the initial power spectrum. The length of this lever arm is crucial for a precise estimation of the scalar spectral index n_s or of a possible 'running' of the spectral index, $dn/dk \neq 0$. An analysis of the Ly-α forest from quasars in the SDSS is presented in McDonald *et al.* (2005) and Seljak *et al.* (2006), there a neutrino mass limit of $m_\nu < 0.17\,\mathrm{eV}$ and no running is inferred. Even if this analysis may be somewhat optimistic, it illustrates the potential of the method.

6.7.6 Weak lensing

The deflection of light from a source behind a mass concentration can lead to the formation of multiple images. This effect is called lensing, or more precisely, 'strong lensing'. If the impact parameter of the light ray connecting the source to the observer is too large, or if the intervening mass is too small, no separate images are formed but the source is deformed by the gravitational field. For example, a spherical source behind a point mass deforms into an ellipse which has its large axis aligned with the radial direction (see Ex. 6.4). In this case we speak of 'weak lensing'. Such an alignment of the ellipticity of galaxies behind massive clusters has been observed and can be used to estimate the cluster mass (Schneider, 2007,).

But weak lensing can also be used in a statistical way, i.e., by measuring the correlation of the direction of galaxy ellipticities, to gain information about the matter power spectrum. The observations are extremely difficult because the ellipticity of galaxies from lensing is typically less than about 1% of their intrinsic ellipticity and is only detectable by a statistical analysis: weak lensing leads to ellipticities which are correlated if the galaxies are close in angular position but may be very far apart in physical space, at different redshifts. In this way, the correlation of ellipticities from lensing can in principle be separated from physical alignment of galaxies which may come from the process of galaxy formation. These galaxy ellipticities measure the shear of the gravitational field which is closely related to the matter power spectrum (Schneider, 2007,).

The advantage of weak lensing observations over galaxy catalogues is that lensing responds purely to the matter density, it does not distinguish luminous matter and dark matter. The disadvantage is that observations of weak lensing are much more difficult. One has to determine the correlation of the ellipticities of background galaxies behind some foreground over density. These correlations are on the level of a few per cent of the actual ellipticities of the galaxies.

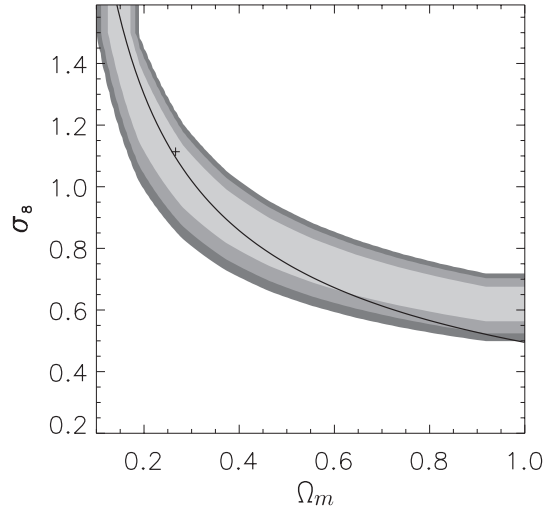


Fig. 6.17. The Ω_m–σ_8 constraints from the combination of all shear measurements available in 2001. The light, medium and dark contours enclose the 1-, 2- and 3-σ contours. The cross indicates the position of the best fit. The line shows the result from the local cluster abundance (see below). From Maoli *et al.* (2001).

In Chapter 7, we shall develop some of the beautiful theory of weak lensing. Here we simply note that lensing can be expressed in terms of the gravitational potential $\Psi \propto \Omega_m D \propto \Omega_m \sigma_8$. What really enters in weak lensing is the line-of-sight integral of the gravitational potential, which reduces the sensitivity to Ω_m to roughly $\sqrt{\Omega_m}\sigma_8$. Constraints from weak lensing lead to the typical 'banana-shaped' contours in the Ω_m–σ_8 plane, see Fig. 6.17. The best limit from Maoli *et al.* (2001) is of the form $\sigma_8 \Omega_m^{0.47} \simeq (0.59 \pm 0.03)$.

For the future, one plans to measure the entire lensing power spectrum which determines the matter power spectrum without any bias. This very powerful tool, which can, in principle, measure the dark matter distribution as a function of redshift, is theoretically nearly as simple as the CMB. If observational difficulties can be overcome, this will provide a most valuable complementary tool for parameter estimation, for a review of weak lensing see Schneider (2007,).

6.7.7 Galaxy clusters

Finally, we want to address briefly the relevance of galaxy clusters for cosmological parameters. Rich clusters with masses $M \sim 10^{14.5}\, h^{-1}\, M_\odot$ are the largest bound structures in the Universe ($M_\odot$ is the mass of the Sun). But clusters vary from groups of tens of galaxies to more than 1000 galaxies. They represent over-densities

of several and are thus non-linear. For Gaussian fluctuations the probability for measuring an over- (or under-) density $\delta\rho = \rho_m D$ on a scale λ is

$$P(D, z) = \frac{1}{\sqrt{2\pi}\sigma(\lambda, z)} \exp\left(-\frac{D^2}{2\sigma^2(\lambda, z)}\right) . \tag{6.81}$$

Here $\sigma^2(\lambda, z)$ is the variance of the density fluctuations on scale λ at redshift z. Assuming that a relative over-density D_c is needed for an object to collapse, the probability that an over-density on scale λ has collapsed into a cluster becomes

$$P(D > D_c|\lambda, z) = 1 - \mathrm{erf}\left(\frac{D_c}{\sqrt{2}\sigma(\lambda, z)}\right) , \tag{6.82}$$

where $\mathrm{erf}(x)$ denotes the error function. This is the basic ingredient of the Press–Schechter formalism (Press & Schechter, 1974).

The spherical collapse model (see e.g. Peebles (1993)) requires $D_c \simeq 1.69$. Assigning the total mass M inside a sphere of radius λ to the collapsed object, allows one to determine the number density of clusters of mass larger than M for different redshifts. This quantity is very sensitive to Ω_m which determines the redshift at which the growth of linear density fluctuations stops. It was one of the first observations indicating $\Omega_m \sim 0.3$ (see Bahcall & Cen, 1992).

Furthermore, clusters usually form at a fixed velocity dispersion. The kinetic energy has to be smaller than the gravitational potential energy for a bound structure to form. Therefore, the cluster density strongly constrains the velocity power spectrum, $P_V \propto \Omega_m^{1.2}\sigma_8^2$ (see Eq. (2.238)). Comparing observations with numerical simulations gives (Pierpaoli *et al.*, 2001) $\sigma_8\Omega_m^{0.6} = 0.495^{+0.034}_{-0.037}$. The line tracing this relation in the Ω_m–σ_8 plane is also indicated in Fig. 6.17.

Comparing lensing observations from clusters that are sensitive to the total mass to X-ray emission which depends on the baryon density in clusters, one can determine the ratio $\Omega_b/\Omega_m \sim 0.1$. Several assumptions go into this value. First of all, X-ray emission is proportional to the line-of-sight integral of ρ_b^2 and assuming $\langle\rho_b^2\rangle \simeq \langle\rho_b\rangle^2$ is not trivial at all, since baryons are strongly clustered on small scales. Furthermore, this assumes that the baryon to matter ratio in clusters is similar to its mean in the total Universe. We know, for example that this is not so in the central parts of galaxies. However, clusters seem to have a sufficiently low mean density, so that hydrodynamical processes that affect baryons, but not dark matter, do not significantly modify the ratio Ω_b/Ω_m in clusters.

Concluding, we state that clusters, being the largest bound structures in the Universe, are an interesting tracer of the mass distribution which should not be ignored. It is very reassuring that the cosmological parameters inferred from clusters fit well with the results from CMB and other observations. Of course, since they

are non-linear structures, they will never achieve the degree of accuracy which we expect from probes of linear perturbations.

6.8 Sources

So far we have assumed that small initial perturbations were generated early in the Universe during an inflationary phase and then evolved under linear perturbation theory. For a given spectrum of, e.g., scalar initial fluctuations $P_\Psi(k)$, the spectrum at some later time is then determined by a transfer function, $P_\Psi(k,t) = g^2(k,t)P_\Psi(k)$. This transfer function only depends on the background cosmology, i.e., on the cosmological parameters.

There is, however yet another possibility: an intrinsically inhomogeneous and anisotropic matter distribution, which makes up only a small perturbation, and which interacts with the cosmological matter and radiation only gravitationally. We consider the energy–momentum tensor of this component as a first-order perturbation. Within linear perturbation theory, it then evolves with the equations of motion determined by the background geometry.

Such a component is termed a **source** or 'seed'. The source's energy–momentum tensor seeds first-order perturbations in the geometry which in turn affect the evolution of matter and radiation, generating fluctuations in the matter density and in the CMB.

6.8.1 Topological defects

Topological defects which can form during symmetry breaking phase transitions are physically well motivated seeds. If the vacuum manifold (i.e., the manifold of minima of the Higgs field (or order parameter) which is responsible for the symmetry breaking is topologically non-trivial, regions where the field cannot relax to the minimum generically occur. The simplest examples are cosmic strings which form, e.g., when a $U(1)$ symmetry is broken. Below a critical temperature T_c, the temperature dependent effective potential $V(\phi, T)$ of the complex Higgs field ϕ changes from a form with a single minimum at $\phi = 0$ to a Mexican hat shape with an entire circle SS of minima, see Fig. 6.18.

When the temperature drops below T_c, the field at a given position $\mathbf{x}$ assumes some value in the new vacuum manifold SS. The field values at positions that are further apart than the Hubble horizon are uncorrelated. Therefore the configuration $\phi(s) = \phi(\mathbf{x}(s))$ along some large closed curve $\mathbf{x}(s)$ in a plane of physical space may well make one (or several) full turns in SS. If this happens, in order to remain continuous, ϕ has to leave the vacuum manifold and assume a value with higher potential energy somewhere in the interior of this curve. Continuing this argument

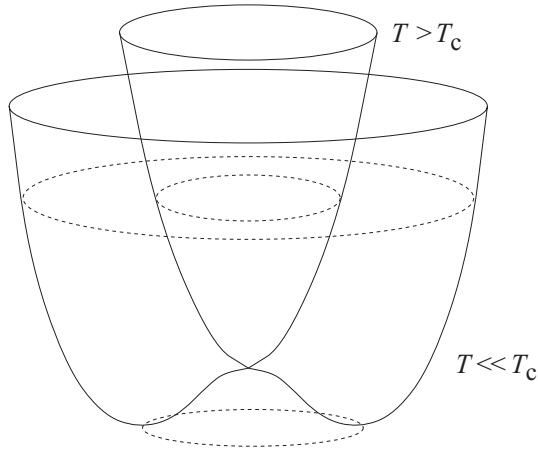


Fig. 6.18. The effective potential of a complex Higgs field for two values of the temperature, $T > T_c$ and $T < T_c$ is shown. The circle at the bottom is the vacuum manifold SS of the low-temperature phase.

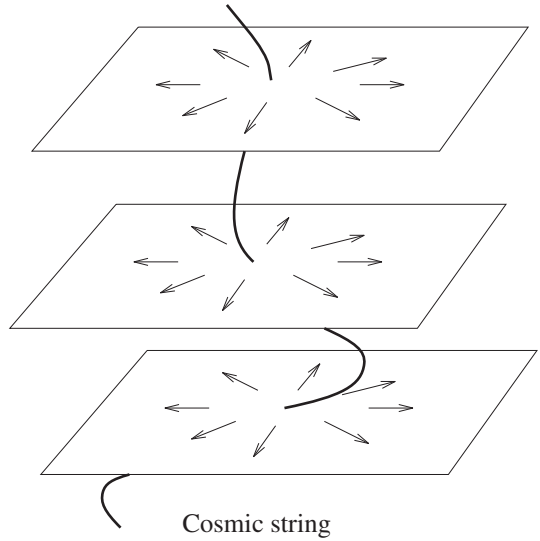


Fig. 6.19. A cosmic string in space is shown with the corresponding configuration of the complex Higgs field, indicated as arrows.

in the third dimension, one obtains a line of higher energy. These lines, which are either closed or infinite, are cosmic strings, see Fig. 6.19.

As the Universe expands, the Higgs field straightens out. Strings which intersect exchange partners and can thereby chop off loops from the network of long strings. In this way the long string network loses energy by shortening the total length of

strings. The strings from a broken gauge symmetry interact with other matter components only gravitationally. They shed energy only into a background of gravity waves which they produce. This process is slow, but sufficiently effective to lead to a mean energy density in cosmic strings which scales like the background energy density $\rho_S \propto 1/\tau^2$. If $M \simeq T_c$ is the energy scale of the phase transition, we expect $\rho_S \simeq M^2/\tau^2$ so that

$$\frac{\rho_S}{\rho} \simeq 4\pi G M^2 = 4\pi \left(\frac{M}{M_{Pl}}\right)^2 \equiv \epsilon \, . \tag{6.83}$$

The amplitude of the induced perturbations will be of the order of ϵ. Recalling that the gravitational potential responsible for the CMB anisotropies is roughly 10^{-5}, we infer that the symmetry breaking scale cannot be much smaller than $M \sim 10^{-3} M_P \sim 10^{16}$ GeV, if such a component is to play a role for CMB anisotropies. Interestingly, this is a grand unified (GUT) scale where some drastic changes of physical interactions, for example a phase transition, are expected to occur from the running of the coupling constants of gauge interactions. If cosmic strings were generated at the electroweak transition (which is not the case in the standard model), they would have far too low energy to play a role for structure formation or CMB anisotropies.

Another type of topological defects, called monopoles, occur at symmetry breaking phase transitions if the vacuum manifold of the broken phase, SS, has the topology of a sphere. More generically, monopoles form if the second homotopy group, $\pi_2(SS)$ is non-trivial. Monopoles are points of higher potential energy. If the broken symmetry is gauged, such massive monopoles cease to interact soon after the phase transition. Their energy density then scales like ordinary matter $\rho \propto a^{-3}$ and soon dominates over the radiation density of the Universe. Therefore, local monopoles are ruled out by observations.

However, if the symmetry is not gauged, the gradients of the scalar field cannot be compensated by the presence of a gauge field. In this case, long range interactions lead to very effective annihilation of monopole–anti-monopole pairs and the remaining energy density has the correct scaling, $\rho_M \propto 1/\tau^2$.

This is not just true for a symmetry breaking Higgs field: an arbitrary, un-ordered multi-component scalar field with a potential minimum at some scale $M \neq 0$, evolves in an expanding universe such that $\rho_S/\rho \simeq GM^2 =$ constant. The field orders on the Hubble scale, so that its gradient and kinetic energy are of the order M^2/τ^2. These findings have been confirmed by numerical simulations and they become very accurate for fields with three or more components (Durrer *et al.*, 2002). For fields with only two components (global strings) this scaling law seems to obtain logarithmic corrections. One-component, or real scalar fields does not scale at all. They generically lead to domain walls which soon come to dominate the

energy density of the Universe and are therefore ruled out. (Their vacuum manifold consists of isolated points, so that they have negligible gradient and kinetic energy. Their energy is dominated by potential energy.)

6.8.2 Causal scaling seeds

If the initial conditions of the scalar field are uncorrelated on scales larger than the Hubble scale, correlations evolve causally and will always vanish on scales larger than the Hubble scale during non-inflationary expansion. The correlation functions of arbitrary components of the energy–momentum tensor are therefore functions with compact support. An important mathematical theorem (Reed & Simon, 1980) states that the Fourier transform of a function with compact support is analytic. On scales which are much smaller than the Hubble scale, the field has already had sufficient time to order and will therefore not contribute much. The only scale in the problem is the Hubble scale $\mathcal{H} \simeq 1/t$. We therefore expect the power spectra to depend on scale only via the dimensionless variable $z \equiv kt$. Seeds which have this behaviour are called 'causal scaling seeds'. They are most interesting since, as we shall see below, they generically predict a scale-invariant spectrum of CMB fluctuations.

We now consider an arbitrary seed energy–momentum tensor which may or may not come from a scalar field or even from cosmic strings, but which has the above properties of scaling and of causality and therefore analyticity. Let us parametrize the correlations of its energy–momentum tensor, $\Theta_{\mu\nu}$ in the form

$$\Theta_{\mu\nu}(\mathbf{k}, t) = M^2 \theta_{\mu\nu}(\mathbf{k}, t) \,, \tag{6.84}$$

$$\langle \theta_{\mu\nu}(\mathbf{k}, t) \theta^*_{\rho\lambda}(\mathbf{k}', t) \rangle = (2\pi)^3 C_{\mu\nu\rho\lambda}(\mathbf{k}, t) \delta(\mathbf{k} - \mathbf{k}') \,. \tag{6.85}$$

The correlators $C_{\mu\nu\rho\lambda}$ are analytic functions of $\mathbf{k}$. Scaling requires that they only depend on $t\mathbf{k}$ and on t, where the t dependence is a simple power law with the power required for dimensional reasons and the dependence on $t\mathbf{k}$ is analytical. We also require $C_{\mu\nu\rho\lambda} \to 0$ for $kt \to \infty$. The dimension of $\theta_{\mu\nu}(\mathbf{x})$ is $1/(\text{length})^2$ so that $\theta_{\mu\nu}(\mathbf{k})$ has the dimension of a length. Hence $C_{\mu\nu\rho\lambda}$ must have the dimension of an inverse length and therefore be of the form $t^{-1} \times$ (an analytical function of $t\mathbf{k}$). For example

$$C_{0000} = \frac{1}{t} F_1(kt) \,, \quad \text{or} \tag{6.86}$$

$$C_{0i0j} = \frac{1}{t} \left[t^2 k_i k_j F_2(kt) + (kt)^2 \delta_{ij} F_3(kt) \right] \tag{6.87}$$

where the functions $F_n(z)$ are analytic in z^2 with the asymptotics

$$\lim_{z\to\infty} F_n(z) = 0\,.$$

It can be shown that if the energy–momentum of the source is conserved, the correlators (6.85) can all be expressed in terms of five free functions with this asymptotic behaviour.

In a given specific model, numerical simulations are usually employed to determine these functions, see Durrer *et al.* (2002).

Let us now show, that on large scales, the CMB anisotropy spectrum from causal scaling seeds is always scale invariant. Since the Laplacians of the Bardeen potentials are of the form

$$k^2\Phi, \qquad k^2\Psi \sim \epsilon\theta$$

where θ denotes some components of $\theta_{\mu\nu}$, their power spectrum must be of the form

$$\left\langle(\Psi+\Phi)(\mathbf{k},t)(\Psi+\Phi)^*(\mathbf{k}',t)\right\rangle = \epsilon^2(2\pi)^3\delta(\mathbf{k}-\mathbf{k}')\frac{F(z^2)}{k^4 t}\,, \tag{6.88}$$

$$\equiv (2\pi)^3\delta(\mathbf{k}-\mathbf{k}')P(k,t)\,, \tag{6.89}$$

where again F is an analytic function of z^2 which tends to 0 for large z. We have written the power spectrum for $\Phi+\Psi$, since this is the quantity which determines the large-scale CMB anisotropy spectrum.

$$\frac{\Delta T}{T}(\mathbf{x}_0,\mathbf{n}) = (\Psi+\Phi)(\mathbf{x}(t_{\rm dec}),t_{\rm dec}) + \int_{t_{\rm dec}}^{t_0}\partial_t(\Psi+\Phi)(\mathbf{x}(t),t)\,dt\,, \tag{6.90}$$

see Eq. (2.231). The Fourier transform of this equation yields

$$\frac{\Delta T}{T}(\mathbf{k},\mathbf{n}) = e^{i\mathbf{k}\cdot\mathbf{n}(t_0-t_{\rm dec})}(\Psi+\Phi)(\mathbf{k},t_{\rm dec}) + \int_{t_{\rm dec}}^{t_0} e^{i\mathbf{k}\cdot\mathbf{n}(t_0-t)}\,\partial_t(\Psi+\Phi)(\mathbf{k},t)\,dt\,. \tag{6.91}$$

We have $\partial_t\left[e^{i\mathbf{k}\cdot\mathbf{n}(t_0-t)}(\Psi+\Phi)\right] = e^{i\mathbf{k}\cdot\mathbf{n}(t_0-t)}[-i\mathbf{kn}(\Psi+\Phi)+\partial_t(\Psi+\Phi)]$. As long as $kt<1$, the second term in this expression dominates and we may therefore approximate the above derivative of $\Phi+\Psi$ by the time derivative of the total integrand. Since $\Psi+\Phi$ decays rapidly inside the horizon, it suffices to integrate until $t=1/k$. The integral can now be perfomed and the value at the lower boundary

simply cancels the 'ordinary Sachs–Wolfe' term. We obtain

$$\frac{\Delta T}{T}(\mathbf{k},\mathbf{n}) \simeq e^{i\mathbf{k}\cdot\mathbf{n}(t_0-1/k)}(\Psi+\Phi)(\mathbf{k},1/k)\,, \tag{6.92}$$

and

$$\left\langle \frac{\Delta T}{T}(\mathbf{k},\mathbf{n})\frac{\Delta T^*}{T}(\mathbf{k}',\mathbf{n}')\right\rangle \simeq e^{i\mathbf{k}\cdot(\mathbf{n}-\mathbf{n}')(t_0-1/k)}\frac{\epsilon^2}{k^3}F(1)(2\pi)^3\delta(\mathbf{k}-\mathbf{k}')\,. \tag{6.93}$$

Expanding $e^{i\mathbf{k}\cdot\mathbf{n}(t_0-1/k)}$ and $e^{-i\mathbf{k}\cdot\mathbf{n}'(t_0-1/k)}$ in Legendre polynomials and spherical Bessel functions, along the same steps as in Section 2.6, we arrive at

$$C_\ell \simeq \epsilon^2 F(1)\frac{2}{\pi}\int_0^\infty \frac{dk}{k}\, j_\ell^2(kt_0) = \frac{\epsilon^2 F(1)}{\pi}\frac{1}{\ell(\ell+1)}\,. \tag{6.94}$$

We have approximated $kt_0 - 1 \sim kt_0$ in the argument of the spherical Bessel function and used the integral (A4.102). As promised, we obtain a scale-invariant spectrum, $\ell(\ell+1)C_\ell =$ constant. The numerical value obtained in this way is not accurate, but the scaling is correct. Note that the main ingredient for the scaling was that the power spectrum of $\Psi+\Phi$ does not contain any other dimensionful parameter other than t and k and that it decays inside the horizon. For dimensional reasons, the spectrum P then is such that $P(k, t = 1/k) \propto 1/k^3$, and the CMB anisotropies become scale invariant.

We expect $F(1)$ to be of order unity, so that ϵ determines the amplitude of the fluctuations.

In the next subsection, we explain how to go beyond such a rough approximation and calculate the CMB anisotropies and polarization from scaling causal seeds in more detail.

6.8.3 Calculating CMB anisotropies from sources

The linear perturbation equations in the presence of sources take the form (in **k**-space)

$$\mathcal{D}X(\mathbf{k},t) = \epsilon SS(\mathbf{k},t)\,, \tag{6.95}$$

where $\mathcal{D}$ is a first-order linear differential operator in time and X is a long vector containing as its components: all the perturbation variables, e.g. D_m; all the temperature fluctuation variables, $\mathcal{M}_\ell^{(m)}$; the polarizations, $\mathcal{E}_\ell^{(m)}$, $\mathcal{B}_\ell^{(m)}$ and so on. SS is the source vector. It describes the gravitational interaction of the source with the cosmic fluid. Its elements are linear combinations of components of the source

energy–momentum tensor $\theta_{\mu\nu}$ and $\epsilon = 4\pi G M^2$ determines the gravitational coupling strength of the source.

In principle, one can simulate the source, Fourier transform it and insert the random variable $SS(\mathbf{k}, t)$ in Eq. (6.95). Averaging over directions in $\mathbf{k}$-space one can then obtain the correlation matrix $P_{nm}(k, t)(2\pi)^3\delta(\mathbf{k} - \mathbf{k}') = \langle X_n(\mathbf{k}) X_m^*(\mathbf{k}')\rangle$. For this one needs as input $SS(\mathbf{k}, t)$ depending on four variables, the three-dimensional $\mathbf{k}$-vector and time. This way has proved to be very tedious, requiring a huge $(3 + 1)$-dimensional numerical simulation with a dynamical range of several hundred only to determine the C_ℓs for $\ell \lesssim 100$. The following observation allows one to reduce the numerical complexity of the problem considerably.

To solve Eq. (6.95), we use the Green function method. If $\mathcal{G}(\mathbf{k}, t, t')$ is the Green function for the operator $\mathcal{D}$ with initial condition $\mathcal{G}(t_1, t_1) = 0$ and $\mathcal{D}\mathcal{G}(t, t_1) = \delta(t - t_1)$, the general solution of Eq. (6.95) is

$$X(\mathbf{k}, t_0) = \epsilon \int_{t_{\rm in}}^{t_0} dt\, \mathcal{G}(\mathbf{k}, t_0, t) SS(\mathbf{k}, t) + X_0(\mathbf{k}, t_0)\,, \tag{6.96}$$

where $t_{\rm in}$ denotes the time at which the source first appears, e.g. the phase transition, and $X_0(\mathbf{k}, t_0)$ is an arbitrary homogeneous solution of Eq. (6.95). A specific example of a Green function is given in Ex. 6.5.

If perturbations are a mixture of two components, one coming from inflation and one from topological defects, X_0 denotes the component from inflation. These two components can be considered as uncorrelated and the resulting perturbation spectra can just be added. We have discussed the computation of the perturbation spectra from inflation in detail in Chapters 2, 4 and 5. Here we want to concentrate on the part induced by the sources. We therefore neglect X_0, so that we obtain for the correlation matrix,

$$\begin{aligned}\langle X_i(\mathbf{k}, t_0) X_j^*(\mathbf{k}', t_0)\rangle = \epsilon^2 \int_{t_{\rm in}}^{t_0} dt'\, dt\, \mathcal{G}_{im}(\mathbf{k}, t_0, t)\mathcal{G}_{nj}^*(\mathbf{k}', t_0, t') \\ \times \langle SS_m(\mathbf{k}, t) SS_n^*(\mathbf{k}', t')\rangle\,.\end{aligned} \tag{6.97}$$

To calculate it, we need to determine the unequal time correlators of the source,

$$\left\langle SS_i(\mathbf{k}, t) SS_j^*(\mathbf{k}', t')\right\rangle = (2\pi)^3 t^p \mathcal{F}_{ij}\left(\sqrt{tt'}\mathbf{k}, r\right)\delta(\mathbf{k} - \mathbf{k}')\,. \tag{6.98}$$

Here we have introduced the ratio $r = t/t'$. The details of the correlation functions $\mathcal{F}_{ij}$ have to be determined case by case, via numerical simulations. But they are much easier to obtain on a large dynamical range than the full random variable SS. The source just consists of linear combinations of the energy–momentum tensor of the seed, which is determined by five functions, $F_n(r, z)$, $n \in \{1, 2, 3, 4, 5\}$ where now $z = \sqrt{tt'}k$. These functions are analytic in $\mathbf{k}$ and therefore in z^2. They go to

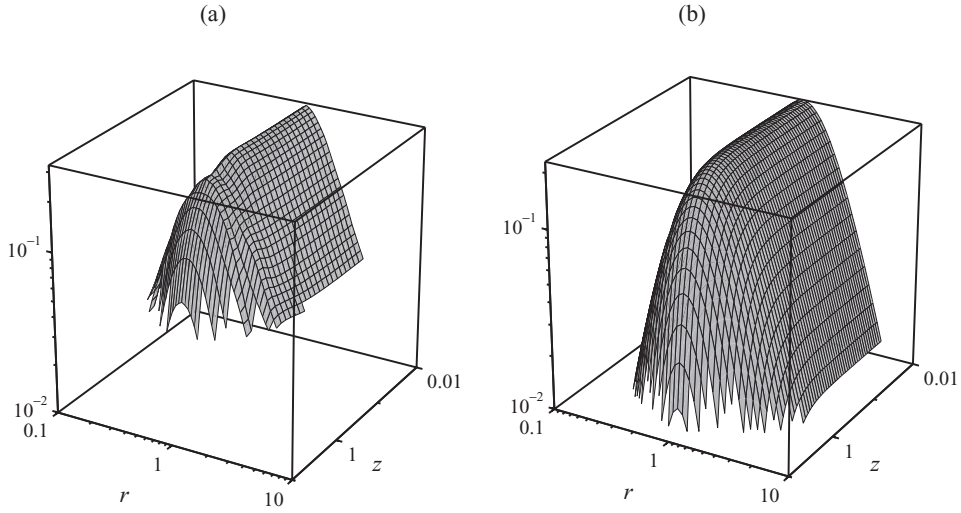


Fig. 6.20. The source correlation function $F(r, z)$ for vector perturbations from numerical simulations for a 4-component scalar field (texture) in panel (a) and for the semi-analytical result for the large N limit, panel (b). From Durrer *et al.* (2002).

zero for either $z \to \infty$ or $r \to \infty, r \to 0$; furthermore, $F_n(r, z) = F_n^*(1/r, z)$. We now only have to determine the amplitude of the functions at $r \sim 1,\ z \sim 0{-}1$ and their behaviour around these values. This is numerically very feasible and has been done with reasonable accuracy. The source functions for vector perturbations of a self-ordering scalar field are shown in Fig. 6.20.

6.8.4 Decoherence

From its definition (6.98) it is clear that the source correlation function $\mathcal{F}(t, t', \mathbf{k})$ can be interpreted as a positive symmetric operator. For a given function $V(t)$ setting $(\mathcal{F}V)(t) = \int dt'\, \mathcal{F}(t, t')V(t')$, we find (suppressing the argument $\mathbf{k}$ for simplicity)

$$\langle V, \mathcal{F}V\rangle \equiv \int dt\, dt'\, V^*(t)\mathcal{F}(t, t')V(t') \geq 0\,.$$

Discretizing it in time, $\mathcal{F}$ becomes a positive semi-definite hermitian matrix, which we can diagonalize. Let us denote its non-negative eigenvalues by $\lambda_1^2, \ldots, \lambda_n^2$, and $D = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_n^2)$ so that $\mathcal{F} = UDU^*$ for some unitary matrix U. For simplicity, we suppress the vector structure of X and present the argument for a simple scalar quantity X and $\mathcal{F}(t, t', \mathbf{k})$. According to Eqs. (6.97) and (6.98), the power

spectrum of X is given by

$$P_X = \int dt\, dt'\, \mathcal{G}(t_0, t)\bar{\mathcal{G}}(t_0, t')\mathcal{F}(t, t') \,.$$

Discretizing this integral and diagonalizing the source function $\mathcal{F}(t, t')$ we obtain

$$P_X = \sum_{ijm} (\Delta t)^2 \mathcal{G}(t_0, t_j)\bar{\mathcal{G}}(t_0, t_i) U_{jm}\bar{U}_{im}\lambda_m^2 = \sum_m \lambda_m^2 \left| \int dt\, \mathcal{G}(t_0, t) U_m(t) \right|^2 , \tag{6.99}$$

where we define $U_m(t_j) \equiv U_{jm}$ and interpolate for values between the time steps. The last equality sign shows that the spectrum P_X is the sum of the spectra with deterministic source terms $\lambda_m U_m(t)$.

In this way, the problem of a stochastic source term is reduced to the problem of many deterministic source terms. In practice, one orders the eigenvalues according to size, $\lambda_1 > \lambda_2 \cdots$ and sums the contributions of about the 20 largest eigenvalues to achieve an accuracy of a few per cent which is also typically the accuracy of the source term from numerical simulations.

With this procedure in mind, let us discuss the acoustic peak structure of the CMB generated by sources. The acoustic peaks from inflationary perturbations reflect the maxima and minima in the radiation/baryon density at the moment of decoupling. The radiation density perturbation oscillates like a cosine wave since it starts at maximum amplitude, $D(kt) \simeq A\cos(c_s kt)$. In the case of sourced perturbations, however, sources generate perturbations at different moments in time, so that $D \simeq \sum_n D_n(kt) = \sum_n A_n \cos(c_s kt - \delta_n)$, with different phases δ_n that are determined by the time at which the perturbation D_n is generated. At the time of decoupling, $t_{\rm dec}$, many different wavelengths can have their maximum or minimum in one of the contributions $D_n(kt_{\rm dec})$. Instead of a distinct peak structure we therefore rather expect a broad hump in the acoustic peak region of the CMB spectrum. This phenomenon, which is rather generic for seeds, is called 'decoherence'. It was first pointed out by Albrecht *et al.* (1996). If only very few eigenvalues dominate, i.e., if the above sum contains only a few terms, decoherence is not very effective and a peak structure can still be seen.

6.8.5 Results

In addition to decoherence, an important characteristic of scaling seeds is that they typically generate vector perturbations with an amplitude comparable to that of scalar perturbations, see Durrer *et al.* (2002) and Bevis *et al.* (2007). Tensor

perturbations are usually somewhat smaller, $C_2^{(T)}/C_2^{(S)} \sim \frac{1}{4}$. Sources are probably the only way to obtain significant vector perturbations in the CMB. Scaling seeds always generate vector perturbations at the horizon scale while vector perturbations that are generated early in the Universe simply decay and leave no traces in the CMB.

Furthermore, the amplitude of the Sachs–Wolfe part of CMB anisotropies is roughly 2Ψ as compared to $\Psi/3$ for adiabatic inflationary perturbations. Therefore, we expect the acoustic peak structure, or the acoustic hump, to be not much higher than the Sachs–Wolfe plateau.

These are the main results for CMB anisotropies seeded by topological defects. They have a scale-invariant Sachs–Wolfe plateau which determines the normalization and which contains important contributions from all, scalar, tensor and especially vector perturbations. This is followed by a very low acoustic hump or peak structure. Since the perturbations are rather of iso-curvature than of adiabatic nature (even though this classification does not strictly apply for sources), this 'hump' is around $\ell \sim 300$–500 in a flat universe. This wide range stems from the uncertainty of the scale at which perturbations are induced. This may be the horizon scale, as for global defects or somewhat less, as for cosmic strings.

In Fig. 6.21 we show the scalar, vector and tensor CMB spectra from a 4-component global scalar field (cosmic texture). The contributions from the largest eigenvalues as well as their sum (bold solid line) are shown. Even though single eigenvalue contributions do show acoustic oscillations, these are washed out in the sum. Similar results have recently been obtained for cosmic strings (Bevis *et al.*, 2007).

From these results it is clear that topological defects or similar sources cannot generate the observed CMB anisotropy spectrum. However, they might make up a small contribution in models where inflation ends with a symmetry breaking phase transition that leads to cosmic strings. It has been argued, that the formation of cosmic strings is quite generic for GUTs and can actually be used to constrain them using the CMB (Rocher & Sakellariadou, 2005).

Another question of interest is the following: as the entire class of scaling causal seeds allows for five nearly free functions of two variables, is it possible to 'manufacture' these functions such that they reproduce the observed CMB anisotropies and polarization or can this be excluded? It has been argued that the small hump at low $\ell \cong 100$ which is generated in the E-polarization spectrum from inflation during decoupling, cannot be reproduced by causal seeds (Spergel & Zaldarriaga, 1997). At decoupling, scales corresponding to $\ell \lesssim 100$ are still super-Hubble and causal seeds have no power on these scales. The problem is that this nice distinction is hard to see in the actual data, where this small hump is affected by the contributions after reionization, which are present in both inflationary models and

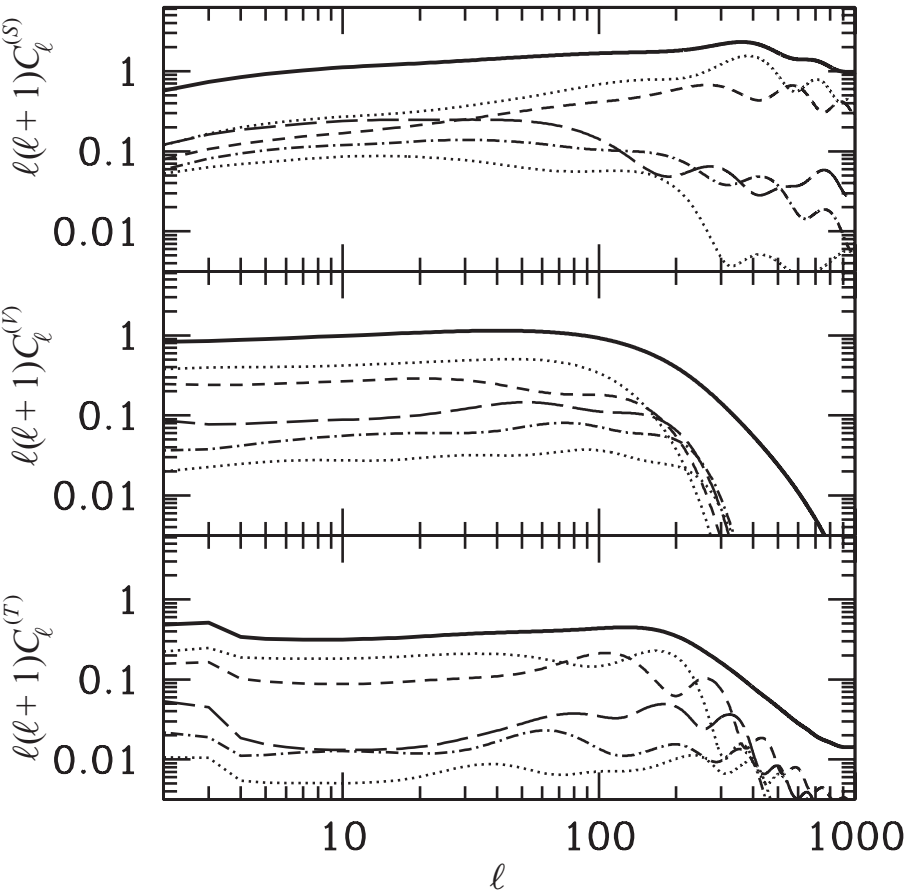


Fig. 6.21. The scalar, vector and tensor contributions for the texture model of structure formation are shown. The dashed lines show the contributions from the first few single eigenfunctions while the solid line represents the sum (over 100 eigenfunctions). Note that the single contributions to the scalar and tensor spectrum do show oscillations which are, however, washed out in the sum (vector perturbations do not obey a wave equation and thus do not show oscillations). Data courtesy of N. Bevis and M. Kunz (see Bevis *et al*., 2004).

models with seeds. The question of whether a seed model can reproduce the data is therefore not answered with real satisfaction.

On the other hand, inflationary models very naturally reproduce the data whereas simple seed models do not. Therefore, Occam's razor certainly cuts in favour of inflationary models but it cannot harm to keep in mind a possible contribution from defects, especially when estimating cosmological parameters.

Exercises

(Exercises marked with an asterisk are solved in Appendix A10.5.)

Ex. 6.1 Optical depth from reionization

Calculate the optical depth $\tau_{\rm ri}(z_{\rm ri})$ as a function of the reionization redshift, $z_{\rm ri}$, for a pure matter universe, $\Omega_{\rm tot} = \Omega_m = 1$ and for a Λ-dominated universe with $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$. Express the result as a function of $\Omega_b h^2$.

Hint: For the Λ-dominated case you may assume $z_{\rm ri} \gtrsim 6$ and neglect the influence of the cosmological constant for $z > 2$ and the contribution to $\tau_{\rm ri}$ for $z < 2$. Estimate your error. Consider two cases.

(i) The universe ionized suddenly at redshift $z_{\rm ri}$.

(ii) Ionization started at redshift $z_{\rm ri} > 6$ and was completed at $z = 6$. In the reionization interval, $6 \le z \le z_{\rm ri}$, the free electron fraction, x rises linearly with the scale factor, $1 - x = (z - 6)/(z_{\rm ri} - 6)$.

Ex. 6.2 Iso-curvature initial conditions*

Let us denote $X_1 = D_\gamma$, $X_2 = D_m$, $X_3 = D_\nu$ and $X_4 = V_\nu$. We parametrize the initial conditions by

$$C_{ij} = \langle X_i(\mathbf{k}) X_j^*(\mathbf{k}')\rangle = A_{ij}(k/H_0)^{n_{ij}}\delta(\mathbf{k} - \mathbf{k}') \,.$$

Show that C_{ij} is positive semi-definite for all values of k if and only if the matrix A_{ij} is positive semi-definite and $n_{ii} \le n_{ij} \le n_{jj}$ or $n_{jj} \le n_{ij} \le n_{ii}$ for all i, j with $A_{ij} \neq 0$.

Ex. 6.3 The shape parameter*

Show that the comoving Hubble scale at equality $H_0 t_{eq} \propto (\Omega_m h)^{-1}$.

Ex. 6.4 Weak lensing

Consider a point mass M at distance D_L in front of a circular source with radius r_s at distance D_S, the centre of which passes the lens with impact parameter b. Using the small angle and small deflection approximation calculate the shape of the image. Show that the ellipticity is parallel to the radial direction. Calculate the ellipticity for a source distance, $D_S = 30$ Mpc, lens distance, $D_L = 25$ Mpc, impact parameter $b = 0.1$ Mpc, source radius $r_s = 0.03$ Mpc and mass $M = 10^{15}\, M_\odot$.

Hint: Approximate the gravitational potential by $\Phi = GM/r$. Calculate the impact parameter of a point on the circular border of the source as a function of $D_{LS} = D_S - D_L$ (neglect the expansion of the Universe). Determine now the image position of this point.

Ex. 6.5 The Green function

Show that for

$$\mathcal{D}X = \ddot{X} + \alpha\dot{X} + \beta X$$

the Green function with initial condition $\mathcal{G}(t, t) = 0$ and $\dot{\mathcal{G}}(t, t) = 1$ is given by

$$\mathcal{G}(t_1, t) = \frac{D_1(t_1)D_2(t) - D_1(t)D_2(t_1)}{\dot{D}_1(t)D_2(t) - D_1(t)\dot{D}_2(t)} \,,$$

where D_1 and D_2 are two linearly independent solutions of the homogeneous equation $\mathcal{D}X = 0$. Show that $\mathcal{G}$ is independent of the choice of D_1 and D_2. $\dot{\mathcal{G}}|t_1, t| \cong \frac{\partial}{\partial t_1}\mathcal{G}(t_1, t)$.

Consider the case $\alpha = 0$ and $\beta = c_s^2 k^2 =$ constant. Introduce a source of the form $SS_1(t) = A_1\delta(t - t_1)$. Discuss decoherence by adding the signal of several sources of this kind.

7

Lensing and the CMB

In this chapter we discuss the most important second-order effect on CMB anisotropies and polarization. Patches of higher or lower CMB temperature are modified and polarization patterns are distorted when they propagate through an inhomogeneous gravitational field. The content of this chapter is strongly inspired by the excellent review by Lewis & Challinor (2006) on the subject.

7.1 An introduction to lensing

On their path from the last scattering surface into our antennas, the CMB photons are deflected by the perturbed gravitational field. If the CMB were perfectly isotropic, the net effect of this deflection would vanish, since, by the conservation of photon number, as many photons would be deflected out of a small solid angle as into it. On the other hand, if there is no perturbation in the gravitational field, the latter is perfectly isotropic and the effect also vanishes. Hence, gravitational lensing of the CMB is a second-order effect and we have not discussed it within linear perturbation theory.

To estimate the effect let us consider the CMB temperature in a point $\mathbf{n}$ in the sky, $T(\mathbf{n})$. If the direction $\mathbf{n}$ is deflected by a small angle $\boldsymbol{\alpha}$, we receive the temperature $T(\mathbf{n})$ from the direction $\mathbf{n}' = \mathbf{n} + \boldsymbol{\alpha}$. Note that, since $\boldsymbol{\alpha}$ is a vector normal to $\mathbf{n}$ also $\mathbf{n}'$ is a unit vector to first order in $\boldsymbol{\alpha}$. To lowest order, this induces a change $\delta T = \boldsymbol{\alpha} \cdot \nabla_{\mathbf{n}} T(\mathbf{n})$, since the angular dependence of the temperature as well as $\boldsymbol{\alpha}$ are first-order quantities this effect is second order.

The deflection angle from a gravitational potential Ψ of an isolated mass distribution is roughly given by $4\Psi_m$, where Ψ_m is the maximum of the gravitational potential along the photon trajectory (see Ex. 7.1). The mean amplitude of the cosmic gravitational potential is about $\sqrt{\langle \Psi^2 \rangle} \simeq 2 \times 10^{-5}$ so that we have $\langle |\boldsymbol{\alpha}| \rangle \sim 10^{-4}$. The typical size of a primordial 'potential well' is difficult to estimate since the potential is scale invariant, but let us approximate it by the horizon

size at equality which is roughly 300 Mpc (comoving). The distance from the last scattering surface to us is about 14 000 Mpc, so that a light ray passes through of the order of 50 such potential wells. Assuming the direction of deviation to be random this yields a total deviation of about $\sqrt{50}|\boldsymbol{\alpha}| \sim 7 \times 10^{-4} \simeq 2$ arc minutes. This corresponds to a deviation of order unity for a patch with an angular size of 2 arc minutes, i.e. for $\ell \sim 4000$. In fact, primary CMB anisotropies on these scales are severely damped by Silk damping so that lensing and other secondary effects like the Sunyaev–Zel'dovich (SZ) effect already dominate on scales larger than $\ell \sim 3000$. In the acoustic peak region which corresponds to about 1° we expect lensing to change the size of the patches by roughly half a per cent on average. Some patches are enlarged while others are reduced in size. In the C_ℓ spectrum this leads to a broadening of the peak. The peak position is somewhat less well defined.

Requiring better than 1% accuracy we have to take into account lensing for $\ell \gtrsim 400$.

7.1.1 The deflection angle

We first want to compute the deflection of a light ray in a perturbed FL universe. We consider only scalar perturbations so that the metric is of the form

$$ds^2 = a^2(t)\left(-(1+2\Psi)\,dt^2 + (1-2\Phi)\gamma_{ij}\,dx^i\,dx^j\right), \tag{7.1}$$

with, see Eqs. (1.9) and (1.12)

$$\gamma_{ij}\,dx^i\,dx^j = dr^2 + \chi(r)(d\vartheta^2 + \sin^2\vartheta\,d\varphi^2)\,. \tag{7.2}$$

Since we are only interested in deflection, we may also consider the conformally related metric

$$d\tilde{s}^2 = -(1+4\Psi_W)\,dt^2 + \gamma_{ij}\,dx^i\,dx^j\,, \tag{7.3}$$

where

$$\Psi_W = \frac{1}{2}(\Psi + \Phi), \tag{7.4}$$

is the Weyl potential. According to Eq. (A3.21) the Weyl tensor from scalar perturbations is given by $(\nabla_i\nabla_j - \frac{1}{3}\gamma_{ij}\Delta)\Psi_W$. Without loss of generality we set the observer position to $\mathbf{x} = 0$ and we consider a photon with an unperturbed trajectory radially towards the observer, $(\bar{x}^\mu) = (s\bar{n}^\mu) = s(1, \mathbf{n})$, where $\mathbf{n}$ is the radially inward photon direction fixed by two angles ϑ_0 and φ_0 and s is an affine parameter. With our choice for s we have $dt/ds = dx^0/ds = 1$, hence $s = t - t_0$ for the unperturbed trajectory. The perturbed photon velocity is given by $(n^\mu) = (1 + \delta n^0(s), \mathbf{n} + \delta\mathbf{n}(s))$. The

Christoffel symbols for $d\tilde{s}^2$ to first order in Ψ_W are easily determined as

$$\tilde{\Gamma}^0_{00} = 2\partial_t\Psi_W\,, \qquad \tilde{\Gamma}^0_{0i} = \tilde{\Gamma}^0_{i0} = 2\partial_i\Psi_W\,, \quad \tilde{\Gamma}^0_{ij} = 0\,,$$
$$\tilde{\Gamma}^i_{00} = 2\gamma^{ij}\partial_j\Psi_W\,, \quad \tilde{\Gamma}^i_{j0} = 0\,, \qquad \tilde{\Gamma}^i_{jm} = \bar{\Gamma}^i_{jm}\,,$$

where $\bar{\Gamma}^i_{jm}$ are the Christoffel symbols of the unperturbed three-dimensional metric γ_{ij}.

With this, the geodesic equation of motion, $\frac{d^2x^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 0$ leads to the following equations of motion for the perturbation of the photon velocity δn^μ

$$\frac{d}{ds}\delta n^0 = -2\partial_t\Psi_W - 4n^i\partial_i\Psi_W = -2\frac{d}{ds}\Psi_W - 2n^i\partial_i\Psi_W, \tag{7.5}$$
$$\frac{d}{ds}\delta n^i = -2\gamma^{ij}\partial_j\Psi_W - 2\delta n^j n^m\bar{\Gamma}^i_{jm}\,. \tag{7.6}$$

Here s denotes the affine parameter along the photon trajectory and we made use of $(d/ds)\Psi_W = \partial_t\Psi_W + n^i\partial_i\Psi_W$. For the rest of this paragraph we set $\dot{} = d/ds$.

The deflection is given by the ϑ- and φ-components of $\delta\mathbf{n} = \epsilon\mathbf{n} + \dot{\vartheta}\partial_\vartheta + \dot{\varphi}\partial_\varphi$. In spherical coordinates (r, ϑ, φ) we have $\mathbf{n} = (-1, 0, 0) \equiv -\partial_r$. The unperturbed Christoffel's in Eq. (7.6) are given by

$$\bar{\Gamma}^\vartheta_{r\vartheta} = \bar{\Gamma}^\varphi_{r\varphi} = \frac{\partial_r\chi}{\chi}$$

and all other $\bar{\Gamma}^j_{ri} = 0$. With this we obtain the following equations of motion for $\dot{\vartheta}$ and $\dot{\varphi}$

$$\ddot{\vartheta} = \frac{-2}{\chi^2}\partial_\vartheta\Psi_W + 2\frac{\partial_r\chi}{\chi}\dot{\vartheta}, \tag{7.7}$$
$$\ddot{\varphi} = \frac{-2}{\chi^2\sin^2\vartheta}\partial_\varphi\Psi_W + 2\frac{\partial_r\chi}{\chi}\dot{\varphi}, \quad \text{so that} \tag{7.8}$$
$$-\frac{d}{ds}\left(\chi^2\dot{\vartheta}\right) = 2\partial_\vartheta\Psi_W\,, \tag{7.9}$$
$$-\frac{d}{ds}\left(\chi^2\dot{\varphi}\right) = 2\sin^{-2}\partial_\varphi\Psi_W\,. \tag{7.10}$$

For the last two lines we have used the fact that to lowest order $\dot{\chi} = -\partial_r\chi$ for radial geodesics. Integrating these equations and using the fact that to lowest order $ds = dt$ we obtain

$$\chi^2(t_0 - t)\dot{\vartheta}(t) = 2\int_{t_0}^{t} dt'\,\partial_\vartheta\Psi_W(t', t_0 - t', \vartheta_0, \varphi_0)\,, \tag{7.11}$$
$$\chi^2(t_0 - t)\dot{\varphi}(t) = 2\int_{t_0}^{t}\frac{dt'}{\sin^2\vartheta_0}\,\partial_\varphi\Psi_W(t', t_0 - t', \vartheta_0, \varphi_0)\,. \tag{7.12}$$

Here t_0 is the time at which we receive the photon at $\mathbf{x} = 0$ and the constant of integration has been fixed by requiring $\chi^2|_{t=t_0} = 0$. Integrating this equation once more leads to

$$\vartheta(t_*) = \vartheta_0 - 2\int_{t_*}^{t_0} dt\, \frac{\chi(t-t_*)\partial_\vartheta \Psi_W(t, t_0-t, \vartheta_0, \varphi_0)}{\chi(t_0-t_*)\chi(t_0-t)}\,, \tag{7.13}$$

$$\varphi(t_*) = \varphi_0 - \frac{2}{\sin^2\vartheta_0}\int_{t_*}^{t_0} dt\, \frac{\chi(t-t_*)\partial_\varphi \Psi_W(t, t_0-t, \vartheta_0, \varphi_0)}{\chi(t_0-t_*)\chi(t_0-t)}\,. \tag{7.14}$$

The easiest way to see that these are the integrals of Eqs. (7.11) and (7.12) is to take the derivative of Eqs. (7.13) and (7.14) with respect to t_* and use $\chi'(t-t_*)\chi(t_0-t_*) - \chi(t-t_*)\chi'(t_0-t_*) = \chi(t_0-t)$ for all three functions χ given in Eq. (1.12).

The deflection angle $\boldsymbol{\alpha} = (\vartheta - \vartheta_0, \sin\vartheta_0(\varphi-\varphi_0))$ is therefore given by

$$\boldsymbol{\alpha} = -2\int_{t_*}^{t_0} dt\, \frac{\chi(t-t_*)}{\chi(t_0-t_*)\chi(t_0-t)}\nabla_\perp \Psi_W(t, t_0-t, \vartheta_0, \varphi_0)\,, \tag{7.15}$$

where $\nabla_\perp = (\partial_\vartheta, (\sin\vartheta)^{-1}\partial_\varphi)$ is the gradient on the sphere and the above expression gives the components of the deflection angle $\boldsymbol{\alpha}$ in this basis. The application $(\vartheta_0, \varphi_0) \to (\vartheta_*, \varphi_*) = (\vartheta_0, \varphi_0) + \boldsymbol{\alpha}(\vartheta_0, \varphi_0)$ is called the lens map. When investigating lensing of the CMB, we want to choose $t_* = t_{\rm dec}$.

Einstein's equation relates Ψ_W to the energy–momentum tensor. Equations (2.104) and (2.105) together with the definition (7.4) yield

$$(\Delta + 3K)\Psi_W = 4\pi G a^2(D + \Pi)\,. \tag{7.16}$$

Using the canonical basis $\mathbf{e}_1 \equiv \mathbf{e}_\vartheta = \partial_\vartheta$ and $\mathbf{e}_2 \equiv \mathbf{e}_\varphi = (\sin\vartheta)^{-1}\partial_\varphi$ we introduce the gradient of the lens map,

$$A_{ab}(\vartheta, \varphi) = \delta_{ab} - 2\int_{t_*}^{t_0} dt\, \frac{\chi(t-t_*)\nabla_a\nabla_b\Psi_W(t, t_0-t, \vartheta, \varphi)}{\chi(t_0-t_*)\chi(t_0-t)}, \tag{7.17}$$

$$\equiv \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}\,. \tag{7.18}$$

The matrix A describes the deformation of a bundle of light rays from direction (ϑ, φ). Its trace, $\mathrm{tr}A = 2(1-\kappa)$ is a measure for the amount of focusing while its traceless part is often represented as the complex number $\gamma = \gamma_1 + i\gamma_2$ represents the shear. As a double gradient of a scalar is symmetric. To first order in perturbation theory, lensing from scalar perturbations does not induce vorticity.

The surface brightness of a source $\iota(\mathbf{n}')$ becomes, after passing through the lensing potential $\iota(\mathbf{n}) = \det(A^{-1})\iota(\mathbf{n}')$. With $\det(A^{-1}) = \left[(1-\kappa)^2 - |\gamma|^2\right]^{-1} \simeq 1 + 2\kappa$, we obtain the magnification μ to first order in the gravitational potential, $\mu = 1 + 2\kappa$. Focusing not only increases the number of photons which reach us from a source (or a patch in the CMB sky), but it also enhances the solid angle

under which we see this patch exactly by the factor $\det A$, so that the number of photons per unit solid angle is conserved. Lensing conserves surface brightness. Photons are neither absorbed nor created by lensing.

The shear is very important for the weak lensing of galaxy surveys as it renders spherical sources elliptical. For CMB lensing both, the focusing κ and the shear γ are relevant.

7.2 The lensing power spectrum

Let us introduce the lensing potential

$$\psi(\mathbf{n}) = -2\int_{t_*}^{t_0} dt \frac{\chi(t-t_*)}{\chi(t_0-t_*)\chi(t_0-t)}\Psi_W(t, \mathbf{n}(t_0-t)) \,. \tag{7.19}$$

This is a function on the sphere and the deflection angle is its gradient. The deflection potential seems to be divergent because $\chi(t_0-t)\to 0$ for $t\to t_0$. But this divergence affects only the monopole term which we may set to zero since it does not affect the lens map. We consider the CMB as a single source at fixed $t_* = t_{\rm dec}$. We expand the lensing potential in spherical harmonics,

$$\psi(\mathbf{n}) = \sum_{\ell m} \psi_{\ell m} Y_{\ell m}(\mathbf{n}) \,, \tag{7.20}$$

$$\langle \psi_{\ell m} \bar{\psi}_{\ell' m'}\rangle \equiv \delta_{\ell\ell'}\delta_{mm'} C_\ell^\psi \,. \tag{7.21}$$

The expectation values C_ℓ^ψ are the lensing power spectrum, and the Kronecker deltas are a consequence of statistical isotropy like for the CMB. The same manipulations as in Chapter 2, Eq. (2.140) now give the lensing correlation function in terms of the power spectrum,

$$\langle \psi(\mathbf{n})\psi(\mathbf{n}')\rangle = \frac{1}{4\pi}\sum_\ell (2\ell+1) C_\ell^\psi P_\ell(\mathbf{n}\cdot\mathbf{n}') \,. \tag{7.22}$$

We want to relate the lensing power spectrum to the primordial power spectrum of the Weyl potential. For simplicity we restrict ourselves to the case $K=0$, with $\chi_0(r) = r$. In this case, the power spectrum of the Weyl potential is given by the Fourier transform,

$$\Psi_W(t,\mathbf{x}) = \frac{1}{(2\pi)^3}\int d^3k\, \Psi_W(t,\mathbf{k})\, e^{-i\mathbf{k}\cdot\mathbf{x}} \,, \tag{7.23}$$

$$\langle \Psi_W(t,\mathbf{k})\bar{\Psi}_W(t',\mathbf{k}')\rangle = (2\pi)^3 T(k,t)\bar{T}(k,t') P_\Psi(k)\,\delta(\mathbf{k}-\mathbf{k}') \,. \tag{7.24}$$

Here we have introduced the primordial power spectrum P_Ψ and the transfer function $T(k,t)$. For a fixed wave number k the transfer function is the solution of the evolution equation for Ψ with initial condition $T(k,t)\to 1$ for $kt\to 0$. The

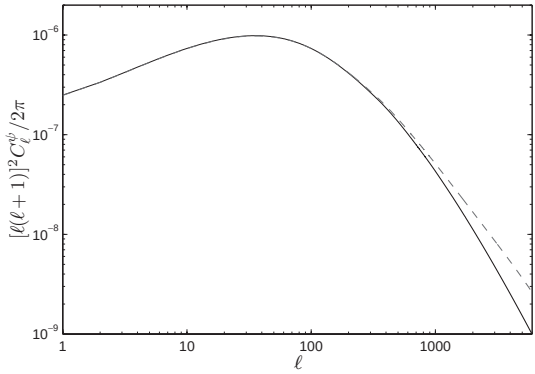


Fig. 7.1. The lensing power spectrum for a ΛCDM concordance model. The solid line is the linear approximation while in the dashed line non-linear corrections in the matter power spectrum are included. Figure from Lewis & Challinor (2006).

transfer function for a matter/radiation universe, neglecting the cosmological constant is constant during the matter era and given by Eq. (2.216), setting $A = \frac{10}{3}$.

For simplicity, we neglect the difference between Ψ and Ψ_W which is given by the anisotropic stresses and never contributes more than a few per cent. This is easily corrected for in a numerical treatment.

Inserting Eqs. (7.24) and (7.19) in Eq. (7.22) and expanding

$$e^{i\mathbf{k}\cdot\mathbf{n}(t_0-t)} = 4\pi \sum_{\ell m} i^\ell j_\ell(k(t_0-t)) Y_{\ell m}(\mathbf{n})\ ,\ \bar{Y}_{\ell m}(\hat{\mathbf{k}}),$$

we obtain

$$C_\ell^\psi = \frac{8}{\pi}\int_0^\infty dk\, k^2 P_\Psi(k) \left| \int_{t_*}^{t_0} dt\, T(k,t) j_\ell(k(t_0-t)) \frac{t-t_*}{(t_0-t_*)(t_0-t)} \right|^2 . \tag{7.25}$$

The relevant quantity for us is the spectrum of the deflection angle $\boldsymbol{\alpha}(\mathbf{n}) = \nabla_\perp \psi(\mathbf{n})$. The correlation function of ψ only depends on the angle between $\mathbf{n}$ and $\mathbf{n}'$. It is invariant under simultaneous infinitesimal variations $\mathbf{n} \to \mathbf{n} + \boldsymbol{\epsilon}$ and $\mathbf{n}' \to \mathbf{n}' + \boldsymbol{\epsilon}$ so that $\langle \psi(\mathbf{n})\psi(\mathbf{n}' + \boldsymbol{\epsilon})\rangle = \langle \psi(\mathbf{n} - \boldsymbol{\epsilon})\psi(\mathbf{n}')\rangle$. Therefore $\langle \nabla_\perp \psi(\mathbf{n}) \nabla_\perp \psi(\mathbf{n}')\rangle = -\langle \Delta\psi(\mathbf{n})\psi(\mathbf{n}')\rangle$. Since $\Delta Y_{\ell m} = -\ell(\ell+1) Y_{\ell m}$, the power spectrum of the deflection angle is simply given by $\ell(\ell+1)C_\ell^\psi$. This power spectrum, multiplied by the usual factor $\ell(\ell+1)/2\pi$, is shown in Fig. 7.1.

7.3 Lensing of the CMB temperature anisotropies

We now want to determine how lensing affects the CMB. On small scales, the lensing potential is nearly completely uncorrelated with the CMB anisotropies.

At $\ell > 60$ the correlation of the lensing and CMB spectra is less than 10% of its maximum value and at $\ell \gtrsim 600$ it drops below 0.1%. Most of the lensing power was generated relatively recently, at $z \lesssim 20$ and it therefore does not correlate with the CMB anisotropies which were generated at $z \simeq 1100$. The lensing signal correlates significantly only with the late integrated Sachs–Wolfe effect which is relevant on very large scales. But the latter has very little structure, so that lensing on large scales is negligible.

Since we are mostly interested in small scales, we approximate the sky by a flat plane as in Section 5.4. The temperature anisotropy is given by Eq. (5.86). The correlation function between two points $\mathbf{x}$ and $\mathbf{x}'$ in the sky

$$\langle \mathcal{M}(\mathbf{x})\mathcal{M}(\mathbf{x}')\rangle = \xi(|\mathbf{x}-\mathbf{x}'|) = \xi(|\mathbf{r}|)\,, \quad \mathbf{r} = \mathbf{x}-\mathbf{x}'\,,$$

is related to the power spectrum by Eq. (5.103),

$$\langle \mathcal{M}(\boldsymbol{\ell})\bar{\mathcal{M}}(\boldsymbol{\ell}')\rangle = \delta^2(\boldsymbol{\ell}-\boldsymbol{\ell}')C_\ell^{(\mathcal{M})}$$
$$C_\ell^{(\mathcal{M})} = 2\pi\int_0^\infty dr\, r J_0(r\ell)\xi(|\mathbf{r}|)\,.$$

The inverse relation is given in Eq. (5.100).

The same equations also relate the lensing potential correlation function to its power spectrum, C_ℓ^ψ.

7.3.1 Approximation for small deflection angles

We now expand the lensed temperature fluctuation in the deflection angle $\boldsymbol{\alpha} = \nabla\psi$,

$$\widetilde{\mathcal{M}}(\mathbf{x}) = \mathcal{M}(\mathbf{x}+\nabla\psi)$$
$$\simeq \mathcal{M}(\mathbf{x}) + \nabla^a\psi(\mathbf{x})\nabla_a\mathcal{M}(\mathbf{x}) + \frac{1}{2}\nabla^a\psi(\mathbf{x})\nabla^b\psi(\mathbf{x})\nabla_b\nabla_a\mathcal{M}(\mathbf{x}) + \cdots\,.$$

This is a good approximation only if the deflection angle is much smaller than the scales of interest to us. If not, we cannot truncate this expansion at second order.

Using

$$\nabla_a\psi(\mathbf{x}) = \frac{-i}{2\pi}\int d^2\boldsymbol{\ell}\,\ell_a\psi(\boldsymbol{\ell})\,e^{-i\boldsymbol{\ell}\cdot x}\,, \qquad \text{and}$$
$$\nabla_a\mathcal{M}(\mathbf{x}) = \frac{-i}{2\pi}\int d^2\boldsymbol{\ell}\,\ell_a\mathcal{M}(\boldsymbol{\ell})\,e^{-i\boldsymbol{\ell}\cdot x}\,,$$

we can obtain the Fourier components for $\widetilde{\mathcal{M}}(\boldsymbol{\ell})$. For this we use that the Fourier transform of a product is equal to the convolution of the Fourier transforms. For

example for the functions $\nabla_a\psi$ and $\nabla_a\mathcal{M}$ we obtain

$$\begin{aligned}
&\frac{1}{2\pi}\int d^2x\,\nabla^a\psi(\mathbf{x})\nabla_a\mathcal{M}(\mathbf{x})e^{i\mathbf{x}\cdot\boldsymbol{\ell}}\\
&\quad=\frac{-1}{(2\pi)^3}\int d^2x\int d^2\ell_1\int d^2\ell_2\,(\boldsymbol{\ell}_1\cdot\boldsymbol{\ell}_2)\,e^{i\mathbf{x}\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)}\psi(\boldsymbol{\ell}_1)\mathcal{M}(\boldsymbol{\ell}_2)\\
&\quad=\frac{-1}{2\pi}\int d^2\ell_2\,((\boldsymbol{\ell}-\boldsymbol{\ell}_2)\cdot\boldsymbol{\ell}_2)\,\psi(\boldsymbol{\ell}-\boldsymbol{\ell}_2)\mathcal{M}(\boldsymbol{\ell}_2)\\
&\quad=\frac{-1}{2\pi}\,(\boldsymbol{\ell}\psi\star\boldsymbol{\ell}\mathcal{M})\,(\boldsymbol{\ell})\,.
\end{aligned}$$

Here $\star$ indicates convolution and for the second equals sign we have used that

$$\frac{1}{(2\pi)^2}\int d^2x\,e^{i\mathbf{x}\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)}=\delta^2(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)\,.$$

Using the above for the second term in the Fourier transform of $\widetilde{\mathcal{M}}$ and the equivalent identity for the third term, we find

$$\begin{aligned}
\widetilde{\mathcal{M}}(\boldsymbol{\ell})\simeq\mathcal{M}(\boldsymbol{\ell})&-\int\frac{d^2\ell'}{2\pi}\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')\psi(\boldsymbol{\ell}-\boldsymbol{\ell}')\mathcal{M}(\boldsymbol{\ell}')\\
&-\frac{1}{2}\int\frac{d^2\ell_1}{2\pi}\int\frac{d^2\ell_2}{2\pi}\boldsymbol{\ell}_1\cdot(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2-\boldsymbol{\ell})\boldsymbol{\ell}_1\cdot\boldsymbol{\ell}_2\mathcal{M}(\boldsymbol{\ell}_1)\psi(\boldsymbol{\ell}_2)\bar{\psi}(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)\,.
\end{aligned}\tag{7.26}$$

To work out the lensed power spectrum we neglect correlations of $\mathcal{M}$ with ψ and use $\bar{\psi}(\boldsymbol{\ell})=\psi(-\boldsymbol{\ell})$. We take into account only terms up to first order in C_ℓ^ψ so that

$$\tilde{C}_\ell\simeq C_\ell+\int\frac{d^2\ell'}{(2\pi)^2}\left[\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')\right]^2C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|}C_{\ell'}-C_\ell\int\frac{d^2\ell'}{(2\pi)^2}(\boldsymbol{\ell}'\cdot\boldsymbol{\ell})^2C^\psi_{\ell'}\,.\tag{7.27}$$

Integrating the second term over the angle gives

$$\tilde{C}_\ell\simeq(1-R^\psi)C_\ell+\int\frac{d^2\ell'}{(2\pi)^2}\left[\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')\right]^2C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|}C_{\ell'}\,,\tag{7.28}$$

where we have introduced the mean square of the deflection angle

$$R^\psi\equiv\frac{1}{2}\langle\boldsymbol{\alpha}^2\rangle=\frac{1}{4\pi}\int_0^\infty d\ell\,\ell^3C_\ell^\psi\,.\tag{7.29}$$

The deflection power spectrum peaks at relatively large scales, $\ell\simeq50$ (see Fig. 7.1) and the bulk of the contribution of the convolution integral in Eq. (7.28) comes from $\boldsymbol{\ell}\sim\boldsymbol{\ell}'$.

We first investigate the result for a scale-invariant CMB power spectrum i.e. $\ell^2C_\ell=$ constant. For such a **scale-invariant spectrum** the above integral

becomes

$$\begin{aligned}\tilde{C}_\ell &\simeq (1-R^\psi)C_\ell + \ell^2 C_\ell \int \frac{d^2\ell'}{(2\pi)^2}\frac{\left[\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')\right]^2}{\ell'^2} C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|} \\ &= (1-R^\psi)C_\ell + \ell^2 C_\ell \int \frac{d^2\ell_1}{(2\pi)^2}\frac{\left[(\boldsymbol{\ell}_1-\boldsymbol{\ell})\cdot\boldsymbol{\ell}_1\right]^2}{(\boldsymbol{\ell}-\boldsymbol{\ell}_1)^2} C^\psi_{\ell_1} \\ &= C_\ell\left[1+\frac{\ell^2}{4\pi}\int_\ell^\infty d\ell_1\,\ell_1 C^\psi_{\ell_1}\left(\ell_1^2-\ell^2\right)\right]. \end{aligned} \tag{7.30}$$

For the last equals sign we have performed the angular integral which is derived in Ex. 7.2.

The integral in Eq. (7.30) is, in general, small, of $\mathcal{O}(10^{-3})$. Note that the spectrum at a scale ℓ is only affected by the lensing power on smaller scales. If the lensing power vanished above a certain value ℓ_0, a scale-invariant spectrum would not be modified by lensing for ℓ's larger than ℓ_0. A large-scale lensing mode magnifies and demagnifies small-scale fluctuations, which has no effect if the fluctuations are scale invarant. The effect of CMB lensing is important because of the acoustic oscillations and Silk damping on small scales which break scale invariance.

7.3.2 Arbitrary deflection angles

As we argued at the beginning of this chapter, for $\ell > 3000$, the deflection angle is comparable to the angular separations which contribute mainly to C_ℓ. A gradient expansion in the deflection angle is therefore no longer justified.

Let us first consider very small scales, $\ell \gg 3000$. On these scales the primordial anisotropies are virtually wiped out by Silk damping and are very small. Even though the deflection angle is larger than the scale in consideration we may approximate $\mathcal{M} \sim \nabla\psi\cdot\nabla\mathcal{M}$. Setting the intrinsic $C_\ell = 0$ we obtain

$$\begin{aligned}\tilde{C}_\ell &\simeq \int \frac{d^2\ell'}{(2\pi)^2} C_{\ell'}\left[\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')\right]^2 C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|} \simeq C^\psi_\ell \int \frac{d^2\ell'}{(2\pi)^2} C_{\ell'}\left[\boldsymbol{\ell}'\cdot\boldsymbol{\ell}\right]^2 \\ &= \ell^2 C^\psi_\ell \int_0^\infty \frac{d\ell'}{4\pi}\ell'^3 C_{\ell'}\,. \end{aligned} \tag{7.31}$$

On very small scales, where intrinsic anisotropies are negligible, the lensed anisotropy power spectrum is given by the power of the deflection angle on this scale multiplied with the integrated anisotropy power on all scales.

To determine the general formula for the lensed CMB anisotropy spectrum, we consider the correlation function. As before we set the lensed temperature anisotropy equal to

$$\widetilde{\mathcal{M}}(\mathbf{x}) = \mathcal{M}(\mathbf{x}+\alpha(\mathbf{x}))\,,$$

where $\alpha = \nabla\psi$ is the deflection angle. For $\mathbf{r} = \mathbf{x} - \mathbf{x}', r = |\mathbf{r}|$ the lensed correlation function $\tilde{\xi}(r)$ is given by

$$\begin{aligned}\tilde{\xi}(r) &= \langle\widetilde{\mathcal{M}}(\mathbf{x})\widetilde{\mathcal{M}}(\mathbf{x}')\rangle = \langle\mathcal{M}(\mathbf{x}+\boldsymbol{\alpha})\mathcal{M}(\mathbf{x}'+\boldsymbol{\alpha}')\rangle \\ &= \int\frac{d^2\boldsymbol{\ell}}{2\pi}\int\frac{d^2\boldsymbol{\ell}'}{2\pi}\left\langle e^{-i\boldsymbol{\ell}\cdot(\mathbf{x}+\boldsymbol{\alpha})}e^{i\boldsymbol{\ell}'\cdot(\mathbf{x}'+\boldsymbol{\alpha}')}\right\rangle\left\langle\mathcal{M}(\boldsymbol{\ell})\widetilde{\mathcal{M}}(\boldsymbol{\ell}')\right\rangle \\ &= \int\frac{d^2\boldsymbol{\ell}}{(2\pi)^2}C_\ell e^{-i\boldsymbol{\ell}\mathbf{r}}\left\langle e^{i\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}'-\boldsymbol{\alpha})}\right\rangle . \end{aligned} \tag{7.32}$$

Here we have used the fact that the CMB anisotropies and the deflection angle are virtually uncorrelated and we can therefore write the expectation value of the product $\left\langle e^{-i\boldsymbol{\ell}\cdot(\mathbf{x}+\boldsymbol{\alpha})}e^{i\boldsymbol{\ell}'\cdot(\mathbf{x}'+\boldsymbol{\alpha}')}\mathcal{M}(\boldsymbol{\ell})\widetilde{\mathcal{M}}(\boldsymbol{\ell}')\right\rangle$ as the product of the expectation values.

We assume that linear perturbations are Gaussian so that $\boldsymbol{\alpha}$ is a Gaussian field. Hence $\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}')$ is a Gaussian random variable with mean $\langle\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}')\rangle = 0$ and variance $\langle[\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}')]^2\rangle$. The expectation value of its exponential is given by, see Ex. 7.3

$$\langle e^{i\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}'-\boldsymbol{\alpha})}\rangle = \exp\left(-\frac{1}{2}\left\langle\left[\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}'-\boldsymbol{\alpha})\right]^2\right\rangle\right) .$$

To calculate the variance of $\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}') = \boldsymbol{\ell}\cdot(\boldsymbol{\alpha}(\mathbf{x})-\boldsymbol{\alpha}(\mathbf{x}+\mathbf{r}))$ we define

$$A_{ij}(\mathbf{r}) = \langle\alpha_i(\mathbf{x})\alpha_j(\mathbf{x}+\mathbf{r})\rangle = \langle\nabla_i\psi(\mathbf{x})\nabla_j\psi(\mathbf{x}+\mathbf{r})\rangle = \int\frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\ell_i\ell_j C_\ell^\psi e^{i\mathbf{r}\cdot\boldsymbol{\ell}} . \tag{7.33}$$

By statistical homogeneity and isotropy, for fixed $r = |\mathbf{r}|$, this symmetric, matrix depends on directions only via $\mathbf{r}$. Therefore it is of the form

$$A_{ij}(\mathbf{r}) = \frac{1}{2}A_0(r)\delta_{ij} - A_2(r)\left[\hat{\mathbf{r}}_i\hat{\mathbf{r}}_j - \frac{1}{2}\delta_{ij}\right] . \tag{7.34}$$

To determine the functions A_0 and A_2 we first take the trace of A_{ij}. This yields

$$A_0(r) = \int_0^\infty\frac{d\ell\,\ell^3}{(2\pi)^2}C_\ell^\psi\int_0^{2\pi}e^{i\ell r\cos\phi} = \int_0^\infty\frac{d\ell\,\ell^3}{2\pi}C_\ell^\psi J_0(r\ell) . \tag{7.35}$$

For the last equals sign we made use of Eq. (5.102). We then contract A_{ij} with $\hat{\mathbf{r}} = \mathbf{r}/r$,

$$\begin{aligned}A_{ij}(\mathbf{r})\hat{\mathbf{r}}_i\hat{\mathbf{r}}_j &= \frac{1}{2}\left(A_0(r)-A_2(r)\right) = \int_0^\infty\frac{d\ell\,\ell^3}{(2\pi)^2}C_\ell^\psi\int_0^{2\pi}\cos^2\phi e^{i\ell r\cos\phi} \\ &= \int_0^\infty\frac{d\ell\,\ell^3}{(2\pi)^2}C_\ell^\psi\int_0^{2\pi}\frac{1}{2}[1+\cos(2\phi)]\phi e^{i\ell r\cos\phi} \\ &= \frac{1}{2}\int_0^\infty\frac{d\ell\,\ell^3}{2\pi}C_\ell^\psi\left(J_0(r\ell)-J_2(r\ell)\right) . \end{aligned} \tag{7.36}$$

We have again used Eq. (5.102) for the last equality. Together with Eq. (7.35) this determines $A_2(r)$,

$$A_2(r) = \int_0^\infty \frac{d\ell\,\ell^3}{2\pi} C_\ell^\psi J_2(r\ell)\,. \tag{7.37}$$

Inserting these results in the variance, we find $A_2(0) = 0$)

$$\begin{aligned}\langle\left[\boldsymbol{\ell}\cdot(\boldsymbol{\alpha}'-\boldsymbol{\alpha})\right]^2\rangle &= 2\ell_i\ell_j\left(\langle\alpha_i\alpha_j\rangle - \langle\alpha_i'\alpha_j\rangle\right)\\ &= \ell^2\left[A_0(0) - A_0(r) + A_2(r)\cos(2\phi)\right]\,.\end{aligned}$$

Inserting this in the correlation function for the lensed anisotropies yields

$$\tilde{\xi}(r) = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} C_\ell\,\exp[-i\ell r\cos\phi]\exp\left[-\frac{\ell^2}{2}\left[A_0(0) - A_0(r) + A_2(r)\cos(2\phi)\right]\right]. \tag{7.38}$$

This expression is exact. With the relation

$$\tilde{C}_{\ell'} = \frac{1}{4\pi}\int_0^\infty r\,dr\,J_0(r\ell')\tilde{\xi}(r)\,,$$

we can obtain the lensed power spectrum from it.

Bessel functions of imaginary arguments are related to the modified Bessel function (see Appendix A4.3), so that Eq. (5.101) leads to

$$\begin{aligned}\exp(-y\cos\phi) &= J_0(iy) + 2\sum_{n=1}^\infty i^n J_n(iy)\cos(n\phi)\\ &= I_0(y) + 2\sum_{n=1}^\infty (-1)^n I_n(y)\cos(n\phi),\end{aligned} \tag{7.39}$$

so that

$$\frac{1}{2\pi}\int_0^{2\pi} d\phi\,\exp(-y\cos\phi)\cos(n\phi) = (-1)^n I_n(y)\,. \tag{7.40}$$

With this and Eq. (5.101), we can perform the angular integration,

$$\begin{aligned}\tilde{\xi}(r) = \int \frac{\ell\,d\ell}{2\pi} C_\ell\,\exp\left[-\frac{\ell^2}{2}\left[A_0(0) - A_0(r)\right]\right]\\ \times\left(I_0(r\ell) + 2\sum_{n=1}^\infty I_n(\ell^2 A_2(r)/2)J_{2n}(r\ell)\right).\end{aligned} \tag{7.41}$$

Note that even though the modified Bessel functions grow exponentially $I_n(r) \to e^r/\sqrt{2\pi r}$ for large arguments, the combination $\exp[-\frac{\ell^2}{2}[A_0(0) - A_0(r)]]I_n(\ell^2 A_2(r)/2) \to 0$ for large ℓ, since $A_0(0) - A_0(r) > A_2(r)$ for all values of r (see Fig. 7.2).

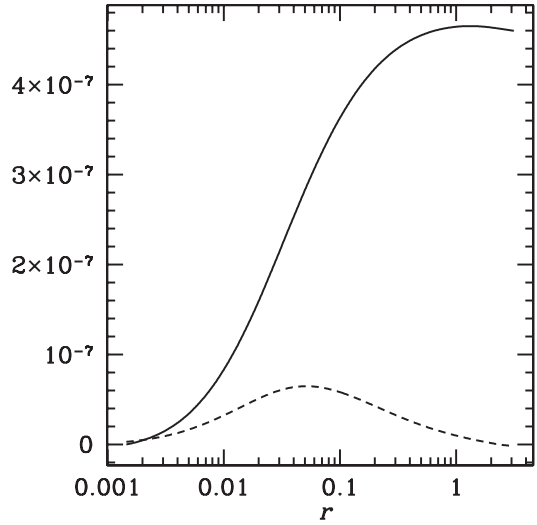


Fig. 7.2. The functions $A_0(0) - A_0(r)$ (solid) and $A_2(r)$ (dashed) are shown as functions of the separation angle r (in radians). The underlying cosmological model is a typical concordance model.

Since $A_2(0) = 0$ and $I_n(0) = \delta_{n0}$, the variance of the lensed CMB anisotropies remains unchanged,

$$\tilde{\xi}(0) = \int \frac{\ell\, d\ell}{2\pi} C_\ell = \xi(0)\,. \tag{7.42}$$

Weak lensing only alters photon directions and hence the spatial structure of the correlation function. The power is redistributed by weak lensing but no power is lost.

A simpler but also accurate expression for the correlation function can be obtained if we approximate the exponential

$$\exp\left[-\frac{\ell^2}{2}\left[A_0(0) - A_0(r) + A_2(r)\cos(2\phi)\right]\right]$$
$$\simeq \exp\left[-\frac{\ell^2}{2}\left[A_0(0) + A_0(r)\right]\right]\left(1 - \frac{\ell^2}{2}A_2(r)\cos(2\phi)\right)\,.$$

Note that this is not an expansion in the deflection angle $\boldsymbol{\alpha}$. The longitudinal part of the correlation function, $\langle \boldsymbol{\alpha}\cdot\boldsymbol{\alpha}'\rangle$ is fully taken into account and we have expanded only the traceless part A_2. A change in the direction of $\boldsymbol{\alpha}'$ with respect to the direction of $\boldsymbol{\alpha}$ contributes to this part. But since the deflection angle has most power on large scales $\ell \sim 50$, this change is small for scales corresponding to $\ell \gtrsim 1000$. As one sees in Fig. 7.2, the function $A_2(r)$ is much smaller than $A_0(0) - A_0(r)$ on all scales and it peaks at $r \sim 0.05$ which corresponds to $\ell \sim 20$, after which it decays like a power law.

Inserting this expansion in Eq. (7.38) we find

$$\tilde{\xi}(r) \simeq \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} C_\ell \exp[-i\ell r\cos\phi] \exp\left[-\frac{\ell^2}{2}\left[A_0(0) - A_0(r)\right]\right] \times \left(1 - \frac{\ell^2}{2} A_2(r)\cos(2\phi)\right) = \int_0^\infty \frac{\ell\, d\ell}{2\pi} C_\ell \exp\left[-\frac{\ell^2}{2}\left[A_0(0) - A_0(r)\right]\right] \times \left(J_0(r\ell) + \frac{\ell^2}{2} A_2(r) J_2(r\ell)\right) . \tag{7.43}$$

Equation (7.43) is a very good approximation which can be used for all ℓs for which CMB lensing is relevant. The $\tilde{C}_\ell$s can be obtained from Eq. (7.43) with the help of Eq. (5.103),

$$\tilde{C}_{\ell'} = \int_0^\infty \ell\, d\ell\, C_\ell \int_0^\infty r dr\, \exp\left[-\frac{\ell^2}{2}\left[A_0(0) - A_0(r)\right]\right] \times \left(J_0(r\ell')J_0(r\ell) + \frac{\ell^2}{2} A_2(r) J_0(r\ell')J_2(r\ell)\right) . \tag{7.44}$$

Observing that $A(0) - A(r) \lesssim 10^{-6}$, we may neglect the exponential for small ℓ. The integral over r of the first term in the parentheses then simply gives $\ell^{-1}\,\delta(\ell - \ell')$ and reproduces the unlensed spectrum. For larger values of ℓ the exponential reduces power and induces a broadening of the δ-function. For very large ℓs the second term also becomes relevant, but $A_2(r) < 10^{-7}$ for all values of r. Therefore, if $\ell^2 A_2(r)$ becomes relevant, the damping exponent $\ell^2[A_0(0) - A_0(r)]$ is several times bigger (see Fig. 7.2), so that this term never dominates.

In Fig. 7.3 the lensed CMB anisotropy power spectrum is shown. The large ℓ approximation given in Eq. (7.31) is also indicated as a dashed line.

7.4 Lensing of the CMB polarization

In this section we study how polarization is affected by lensing. We work again in the flat sky approximation which is sufficient for $\ell \gtrsim 20$. There are, in principle, two contributions: First, like for temperature anisotropies, the direction $\mathbf{n}$ in which a given photon is received has been deflected by the deflection angle $\boldsymbol{\alpha}$ from the direction in which it has been emitted, $\mathbf{n}' = \mathbf{n} + \boldsymbol{\alpha}$. Second, the polarization tensor is parallel-transported along the perturbed photon geodesics. To lowest order this means that the orientation of the polarization in the observed direction $\mathbf{n}$ and in the lensed direction $\mathbf{n}'$ is the same if it is determined which respect to a basis which is parallel-transported from $\mathbf{n}$ to $\mathbf{n}'$. Since the distance between $\mathbf{n}$ and $\mathbf{n}'$ is already first order, we may neglect the perturbation of the gravitational field along the geodesic from $\mathbf{n}$ to $\mathbf{n}'$. In the flat sky approximation, this simply means that we have to measure polarization with respect to the same basis $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\epsilon}^{(2)}$ in both points. With this, the second effect is automatically taken care of, to first order.

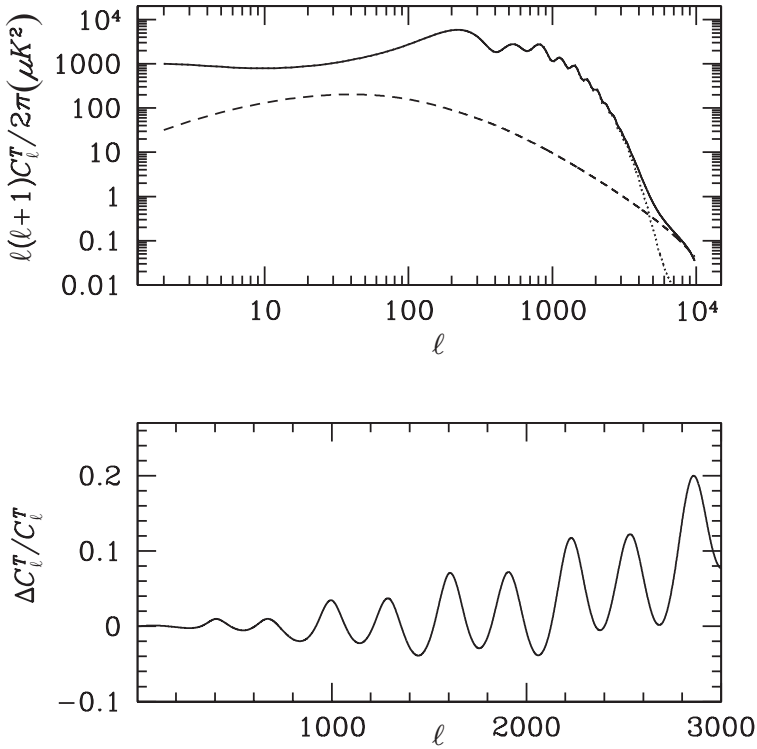


Fig. 7.3. Top panel: the lensed CMB temperature anisotropy spectrum is shown (solid). Underlaid is the unlensed spectrum (dotted). The large ℓ approximation for the lensed CMB spectrum is also indicated (dashed).
Bottom panel: the fractional difference between the lensed and non-lensed CMB spectrum.

We introduce, as in Chapter 5, see Eq. (5.5),

$$\mathbf{e}_\pm = \frac{1}{\sqrt{2}}(\boldsymbol{\epsilon}^{(1)} \pm i\boldsymbol{\epsilon}^{(2)}) \text{ and}$$

$$\mathcal{P} \equiv 2\mathbf{e}_+^i \mathbf{e}_+^j \mathcal{P}_{ij} = \mathcal{Q} + i\mathcal{U} \text{ , so that } \bar{\mathcal{P}} \equiv 2\bar{\mathbf{e}}_+^i \bar{\mathbf{e}}_+^j \mathcal{P}_{ij} = 2\mathbf{e}_-^i \mathbf{e}_-^j \mathcal{P}_{ij} = \mathcal{Q} - i\mathcal{U} \text{ .}$$

Expanding $\mathcal{Q} \pm i\mathcal{U}$ in Fourier space and direction, we have, see Eq. (5.35)

$$\mathcal{Q} \pm i\mathcal{U} = \int \frac{d^3k}{(2\pi)^3} \sum_{\ell=2}^{\infty} \sum_{m=-2}^{2} \times \left(\mathcal{E}_\ell^{(m)}(t,\mathbf{k}) \pm i\mathcal{B}_\ell^{(m)}(t,\mathbf{k})\right) {}_{\pm 2}G_{\ell m}(\mathbf{x},\mathbf{n}) \text{ ,} \tag{7.45}$$

with

$${}_sG_{\ell m}(\mathbf{x},\mathbf{n}) = (-i)^\ell \sqrt{\frac{4\pi}{2\ell+1}} e^{i\mathbf{k}\cdot\mathbf{x}} {}_sY_{\ell m}(\mathbf{n}) \text{ .} \tag{7.46}$$

The spin-2 spherical harmonics are given by (see Appendix A4.2.4) ${}_{\pm2}Y_{\ell m}(\mathbf{x}) = 2\sqrt{(\ell-2)!/(\ell+2)!}\nabla_{\mathbf{e}_\pm}\nabla_{\mathbf{e}_\pm}Y_{\ell m}$. As we have already seen in Chapter 5, in the flat sky approximation they become

$$ {}_{\pm2}Y_{\ell m} = {}_{\pm2}Y_{\boldsymbol{\ell}}(\mathbf{x}) = \frac{2}{\ell^2}\mathbf{e}^i_\pm\mathbf{e}^j_\pm\nabla_i\nabla_j e^{i\boldsymbol{\ell}\cdot\mathbf{x}} = -e^{\pm2i\phi}e^{i\boldsymbol{\ell}\cdot\mathbf{x}}\,. \tag{7.47}$$

Here ϕ denotes the angle which $\boldsymbol{\ell}$ encloses with the x axis. The relation of the polarization field with its Fourier transforms $\mathcal{E}(\boldsymbol{\ell})$ and $\mathcal{B}(\boldsymbol{\ell})$ is given in Chapter 5 in Eqs. (5.104)–(5.107).

7.4.1 The lensed polarization power spectrum

We again start by expanding the polarization tensor in the deflection angle up to second order. This is a good approximation only when considering angular scales which are much larger than the deflection angle, i.e. up to about $\ell \sim 1000$. We will have to do better in a second approach, but this approximation helps us to develop an intuition for the modifications of CMB polarization by lensing.

We shall see that even if, initially, perturbations are purely scalar and therefore do not have B-modes, the lensed polarization will develop B-modes. This is the most important effect from lensing: it generates B-modes from scalar perturbations so that B-modes are no longer an unambiguous sign of gravity waves.

7.4.2 Approximation for small deflection angles

As for the temperature anisotropies, we Taylor expand the polarization tensor to second order,

$$\begin{aligned}\tilde{P}_{ij}(\mathbf{x}) &= P_{ij}(\mathbf{x}+\nabla\psi)\\ &\simeq P_{ij}(\mathbf{x}) + \nabla^m\psi\nabla_m P_{ij}(\mathbf{x}) + \frac{1}{2}\nabla^m\psi\nabla^n\psi\nabla_n\nabla_m P_{ij}(\mathbf{x})\,.\end{aligned}$$

Since parallel-transporting in the flat sky just means keeping the polarization basis $\mathbf{e}_\pm$ constant, the same expansion is also valid for $\mathcal{P} = \mathcal{Q} + i\mathcal{U}$ and $\bar{\mathcal{P}} = \mathcal{Q} - i\mathcal{U}$. Fourier transforming the above expression leads to the same convolution integrals as we obtained for the lensed temperature anisotropies in Eq. (7.26). With the help of Eqs. (5.104)–(5.107) we find

$$\begin{aligned}\left(\tilde{\mathcal{E}}(\boldsymbol{\ell}) \pm i\tilde{\mathcal{B}}(\boldsymbol{\ell})\right)e^{2i\phi_\ell} \simeq{}& \left(\mathcal{E}(\boldsymbol{\ell}) \pm i\mathcal{B}(\boldsymbol{\ell})\right)e^{2i\phi_\ell}\\ &- \int\frac{d^2\boldsymbol{\ell}'}{2\pi}\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')\psi(\boldsymbol{\ell}-\boldsymbol{\ell}')[\mathcal{E}(\boldsymbol{\ell}') \pm i\mathcal{B}(\boldsymbol{\ell}')]e^{2i\phi'_\ell}\\ &-\frac{1}{2}\int\frac{d^2\boldsymbol{\ell}_1}{2\pi}\int\frac{d^2\boldsymbol{\ell}_2}{2\pi}\boldsymbol{\ell}_1\cdot(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2-\boldsymbol{\ell})\boldsymbol{\ell}_1\cdot\boldsymbol{\ell}_2\\ &\times[\mathcal{E}(\boldsymbol{\ell}_1) \pm i\mathcal{B}(\boldsymbol{\ell}_1)]e^{2i\phi_{\ell_1}} \times \psi(\boldsymbol{\ell}_2)\bar{\psi}(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)\,.\end{aligned} \tag{7.48}$$

In the flat sky approximation, the E- and B-polarization spectra and the T–E cross polarization spectrum are of the form

$$\langle \mathcal{E}(\boldsymbol{\ell})\bar{\mathcal{E}}(\boldsymbol{\ell}')\rangle = \delta^2(\boldsymbol{\ell}-\boldsymbol{\ell}')C_\ell^{(\mathcal{E})}\,, \qquad \langle \mathcal{B}(\boldsymbol{\ell})\bar{\mathcal{B}}(\boldsymbol{\ell}')\rangle = \delta^2(\boldsymbol{\ell}-\boldsymbol{\ell}')C_\ell^{(\mathcal{B})}\,,$$
$$\langle \mathcal{E}(\boldsymbol{\ell})\bar{\mathcal{M}}(\boldsymbol{\ell}')\rangle = \delta^2(\boldsymbol{\ell}-\boldsymbol{\ell}')C_\ell^{(\mathcal{E}\mathcal{M})}\,.$$

Multiplying Eq. (7.48) with its complex conjugate, with itself or with the expression for lensed temperature anisotropies in $\boldsymbol{\ell}$-space, Eq. (7.26) and keeping only lowest-order expressions in C_ℓ^ψ, we obtain,

$$\begin{aligned}\tilde{C}_\ell^{(\mathcal{E})} + \tilde{C}_\ell^{(\mathcal{B})} &= C_\ell^{(\mathcal{E})} + C_\ell^{(\mathcal{B})} + \int \frac{d^2\ell'}{(2\pi)^2}[\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')]^2 C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|}\left[C_{\ell'}^{(\mathcal{E})} + C_{\ell'}^{(\mathcal{B})}\right] \\ &\quad - \left[C_\ell^{(\mathcal{E})} + C_\ell^{(\mathcal{B})}\right]\int \frac{d^2\ell'}{(2\pi)^2}(\boldsymbol{\ell}'\cdot\boldsymbol{\ell})^2 C^\psi_{\ell'}\,, \qquad (7.49)\\ \tilde{C}_\ell^{(\mathcal{E})} - \tilde{C}_\ell^{(\mathcal{B})} &= C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})} \\ &\quad + \int \frac{d^2\ell'}{(2\pi)^2}[\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')]^2 e^{4i(\phi_{\ell'}-\phi_\ell)} C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|}\left[C_{\ell'}^{(\mathcal{E})} - C_{\ell'}^{(\mathcal{B})}\right] \\ &\quad - \left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right]\int \frac{d^2\ell'}{(2\pi)^2}(\boldsymbol{\ell}'\cdot\boldsymbol{\ell})^2 C^\psi_{\ell'}\,, \qquad (7.50)\\ \tilde{C}_\ell^{(\mathcal{E}\mathcal{M})} &= C_\ell^{(\mathcal{E}\mathcal{M})} + \int \frac{d^2\ell'}{(2\pi)^2}[\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')]^2 e^{2i(\phi_{\ell'}-\phi_\ell)} C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|} C_{\ell'}^{(\mathcal{E}\mathcal{M})} \\ &\quad - C_\ell^{(\mathcal{E}\mathcal{M})}\int \frac{d^2\ell'}{(2\pi)^2}(\boldsymbol{\ell}'\cdot\boldsymbol{\ell})^2 C^\psi_{\ell'}\,. \qquad (7.51)\end{aligned}$$

For these results we have made use of the fact that $\mathcal{B}$ is uncorrelated with both, $\mathcal{E}$ and $\mathcal{M}$. In the angular integration of the above expressions, only the real part contributes. Also noting that the scalar product $\boldsymbol{\ell}'\cdot\boldsymbol{\ell} = \ell'\ell\cos(\phi_{\ell'}-\phi_\ell)$ only depends on the angle difference $\phi \equiv \phi_{\ell'} - \phi_\ell$ the angular integral in (7.50) is of the form

$$\begin{aligned}&\int_0^{2\pi} d\phi\; f(\cos\phi)e^{4i\phi} \\ &= \int_0^{2\pi} d\phi\; f(\cos\phi)\big[\cos(4\phi) + 4i\sin\phi\cos\phi(\cos^2\phi - \sin^2\phi)\big] \\ &= \int_0^{2\pi} d\phi\; f(\cos\phi)\cos(4\phi)\,.\end{aligned}$$

The imaginary part of such an integral vanishes since $\int_0^{2\pi} f(\cos\phi)\sin\phi\, d\phi = \int_{-\pi}^{\pi} f(\cos\phi)\sin\phi\, d\phi = 0$ for arbitrary functions of $\cos\phi$. Correspondingly we have

$$\begin{aligned}\int_0^{2\pi} d\phi\; f(\cos\phi)e^{2i\phi} &= \int_0^{2\pi} d\phi\; f(\cos\phi)[\cos(2\phi) + 2i\sin\phi\cos\phi] \\ &= \int_0^{2\pi} d\phi\; f(\cos\phi)\cos(2\phi)\,.\end{aligned}$$

We may therefore replace $e^{4i\phi} \to \cos(4\phi) = \cos^2(2\phi) - \sin^2(2\phi)$ and $e^{2i\phi} \to \cos(2\phi)$. With the definition (7.29) of the mean square deflection angle, $R^\psi = (4\pi)^{-1}\int_0^\infty d\ell\, \ell^3 C_\ell^\psi$, we then find

$$\tilde{C}_\ell^{(\mathcal{E})} = (1-\ell^2 R^\psi) C_\ell^{(\mathcal{E})} + \int \frac{d^2\ell'}{(2\pi)^2} [\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')]^2 C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|} \times \left[C_{\ell'}^{(\mathcal{E})}\cos^2 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}}) + C_{\ell'}^{(\mathcal{B})}\sin^2 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}})\right], \tag{7.52}$$

$$\tilde{C}_\ell^{(\mathcal{B})} = (1-\ell^2 R^\psi) C_\ell^{(\mathcal{B})} + \int \frac{d^2\ell'}{(2\pi)^2} [\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')]^2 C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|} \times \left[C_{\ell'}^{(\mathcal{B})}\cos^2 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}}) + C_{\ell'}^{(\mathcal{E})}\sin^2 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}})\right], \tag{7.53}$$

$$\tilde{C}_\ell^{(\mathcal{EM})} = (1-\ell^2 R^\psi) C_\ell^{(\mathcal{EM})} + \int \frac{d^2\ell'}{(2\pi)^2} [\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}-\boldsymbol{\ell}')]^2 C^\psi_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|} C_{\ell'}^{(\mathcal{EM})}\cos 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}})\,. \tag{7.54}$$

As we see from Eq. (7.53), even if there is no unlensed B-mode, $C^{(\mathcal{B})} = 0$, like for purely scalar perturbations, the lensing deflection induces a non-zero B-spectrum, $\tilde{C}^{(\mathcal{B})} \neq 0$. On relatively large scales, $\ell \ll \ell'$, the lensed B-mode induced by a pure primordial E-mode is roughly[1]

$$\tilde{C}_\ell^{(\mathcal{B})} \sim \int \frac{d^2\ell'}{(2\pi)^2}\ell'^4 C_{\ell'}^\psi C_{\ell'}^{(\mathcal{E})}\sin^2 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}}) = \int \frac{d\ell'}{4\pi}\ell'^5 C_{\ell'}^\psi C_{\ell'}^{(\mathcal{E})}\,. \tag{7.55}$$

This is an ℓ-independent constant. On large scales, the B-mode power spectrum induced by lensing is white noise. The contribution to the power per logarithmic interval, $d\log(\ell) = d\ell/\ell$ is

$$d\tilde{C}_\ell^{(\mathcal{B})} = \frac{1}{2}\ell^4 C_\ell^\psi \frac{\ell^2 C_\ell^{(\mathcal{E})}}{2\pi}\,,$$

half the product of the power of the deflection angle and the E-polarization, see Fig. 7.4.

At very small scales, $\ell \gg \ell'$ we can approximate the lensed E- and B-spectra by

$$\tilde{C}_\ell^{(\mathcal{E})} \simeq C_\ell^\psi \int \frac{d^2\ell'}{(2\pi)^2}[\boldsymbol{\ell}'\cdot\boldsymbol{\ell}]^2 C_{\ell'}^{(\mathcal{E})}\cos^2 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}}) = \frac{1}{2}\ell^2 C_\ell^\psi R_E, \tag{7.56}$$

$$\tilde{C}_\ell^{(\mathcal{B})} \simeq C_\ell^\psi \int \frac{d^2\ell'}{(2\pi)^2}[\boldsymbol{\ell}'\cdot\boldsymbol{\ell}]^2 C_{\ell'}^{(\mathcal{E})}\sin^2 2(\phi_{\boldsymbol{\ell}'}-\phi_{\boldsymbol{\ell}}) = \frac{1}{2}\ell^2 C_\ell^\psi R_E = \tilde{C}_\ell^{(\mathcal{E})}\,, \tag{7.57}$$

[1] We integrate over ℓ' and so, in principle, the inequality $\ell \ll \ell'$ does not strictly make sense. What we mean, of course, is that ℓ is much smaller than those values of ℓ', which mainly contribute to the above integral. In the same sense we shall use $\ell \gg \ell'$ below.

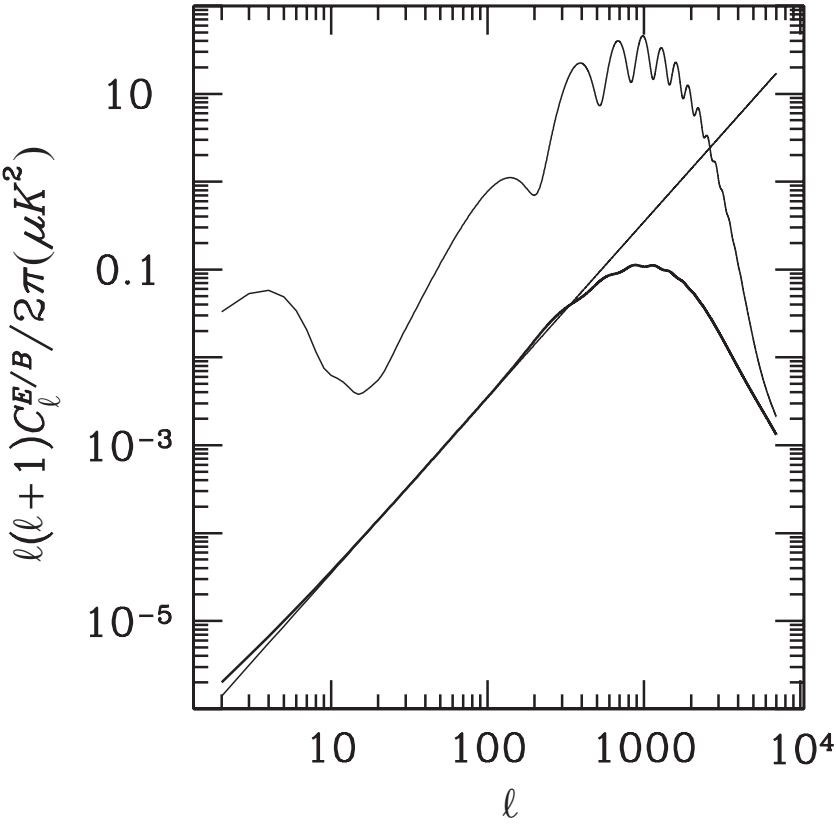


Fig. 7.4. The B-mode power spectrum induced from a pure E-mode by lensing is shown (thick solid curve). The lensed E-power spectrum (thin solid curve) is also indicated. The thin straight line traces the white noise approximation (7.55) which is excellent for $\ell < 200$.

where we have introduced the variance of the gradient of polarization,

$$R_E = \frac{1}{4\pi}\int d\ell\, \ell^3 C_\ell^{(\mathcal{E})} = \langle|\nabla\mathcal{Q}|^2\rangle = \langle|\nabla\mathcal{U}|^2\rangle \sim \frac{2\times 10^7(\mu K)^2}{T_0^2}\,. \quad (7.58)$$

The order of magnitude of this numerical value can be estimated from Fig. 7.4 by noting that at its maximum, $\ell \sim 1000$, the E-polarization spectrum is about $\ell^2 C_\ell^{(\mathcal{E})}/(2\pi) \sim 40(\mu K)^2/T_0^2$. The results (7.56) and (7.57) are valid for a pure E-primordial spectrum, but are not significantly modified if primordial B-modes are also present since the latter usually contribute very little on small scales.

The lensed and unlensed E-power spectrum and E–T-correlation spectrum are compared in Fig. 7.5. The lensed E- and B-power spectra from purely scalar primordial perturbations are shown in Fig. 7.6.

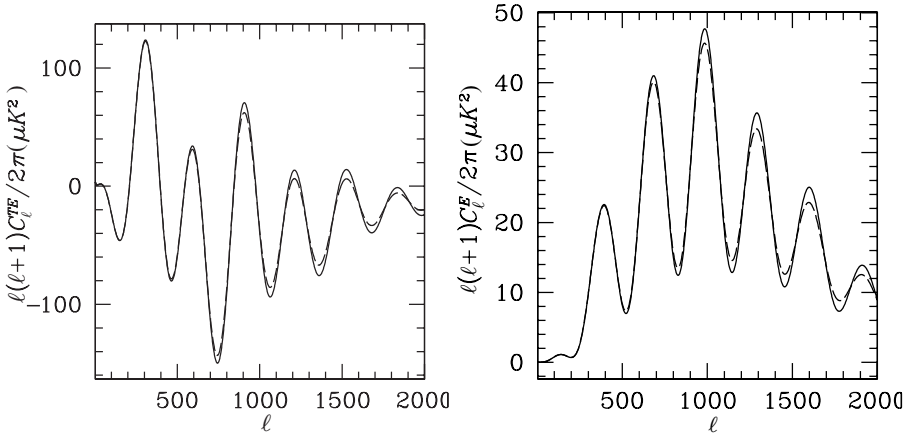


Fig. 7.5. The unlensed E–T-correlation and E-power spectra are compared with the smoother lensed spectra (dashed).

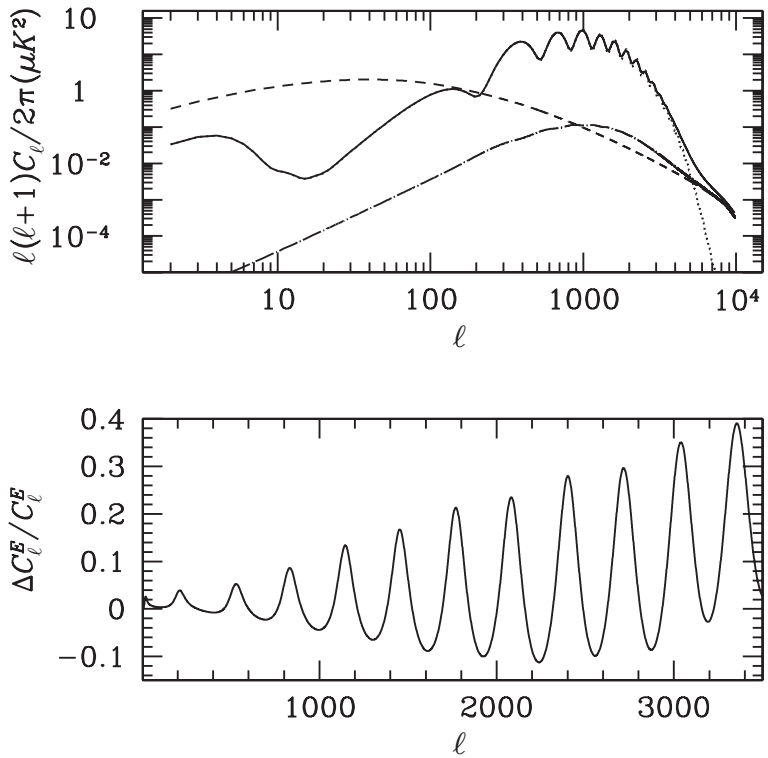


Fig. 7.6. The lensed E- (solid) and the induced B- (long dashed) power spectra are shown. The deflection angle spectrum (short dashed) and the unlensed E-power spectrum (dotted) are also indicated. The bottom panel shows the relative difference of the lensed and unlensed E-spectra.

7.4.3 Arbitrary deflection angles

We now derive an expression which is also valid for large ℓs where the deflection angle is no longer smaller than the scale of interest. As for the temperature anisotropies, we study the modification of the correlation function by lensing.

As in Chapter 5, we consider two points $\mathbf{x}$ and $\mathbf{x}'$ and define the polarization basis along $\mathbf{r} = \mathbf{x} - \mathbf{x}'$. The rotated polarization is given by

$$\mathcal{P}_r(\mathbf{x}) = e^{-2i\phi_r}\mathcal{P}(\mathbf{x}) \,.$$

The correlation functions of this rotated polarization are defined in Eqs. (5.108)–(5.110). The calculation of the lensed correlation function $\tilde{\xi}_+$ is exactly analogous to that for the temperature anisotropy, one just has to replace C_ℓ by $C_\ell^{(\mathcal{E})} + C_\ell^{(\mathcal{B})}$.

$$\begin{aligned}\tilde{\xi}_+(r) &= \left\langle \bar{\mathcal{P}}(\mathbf{x}+\boldsymbol{\alpha})\mathcal{P}(\mathbf{x}'+\boldsymbol{\alpha}')\right\rangle \\ &= \frac{1}{2\pi}\int \ell\, d\ell \left[C_\ell^{(\mathcal{E})} + C_\ell^{(\mathcal{B})}\right] e^{-(\ell^2/2)(A_0(0)-A_0(r))} \\ &\quad \times \left(I_0(\ell^2 A_2(r)/2)J_0(r\ell) + 2\sum_{n=1}^{\infty} I_n(\ell^2 A_2(r)/2)J_{2n}(r\ell)\right) . \end{aligned} \tag{7.59}$$

For ξ_- and $\xi_\times$, the situation is somewhat different because of the exponentials $e^{im\phi_r}$. We insert the Fourier transform of $\mathcal{P}(\mathbf{x})$ given in Eq. (5.104) in the expression for ξ_-, we find

$$\begin{aligned}\tilde{\xi}_-(r) &= \left\langle \mathcal{P}(\mathbf{x}+\boldsymbol{\alpha})\mathcal{P}(\mathbf{x}'+\boldsymbol{\alpha}')e^{-i4\phi_r}\right\rangle \\ &= \int \frac{d^2\ell}{2\pi}\frac{d^2\ell'}{2\pi}\langle[\mathcal{E}(\boldsymbol{\ell}) + i\mathcal{B}(\boldsymbol{\ell})][\mathcal{E}^*(\boldsymbol{\ell}') + i\mathcal{B}^*(\boldsymbol{\ell}')]\rangle \\ &\quad \times e^{i(\boldsymbol{\ell}\cdot\mathbf{x}-\boldsymbol{\ell}'\cdot\mathbf{x}')}e^{i(2\phi_\ell+2\phi_{\ell'}-4\phi_r)}\langle e^{i(\boldsymbol{\ell}\cdot\boldsymbol{\alpha}-\boldsymbol{\ell}'\cdot\boldsymbol{\alpha}')}\rangle \\ &= \int \frac{d^2\ell}{(2\pi)^2}\left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right] e^{ir\ell\cos\phi}e^{4i\phi} \\ &\quad \times \exp\left(-\frac{\ell^2}{2}[A_0(0) - A_0(r) + \cos(2\phi)A_2(r)]\right) . \end{aligned}$$

For the second equals sign we have used the expression for the E- and B-polarization spectra and we have set $\phi = \phi_\ell - \phi_r$.

Of the factor $e^{4i\phi}$ only the real part survives integration over ϕ, since the imaginary part can be written in the form $f(\cos\phi)\sin\phi$. Furthermore, $\cos 4\phi = \cos^2 2\phi - \sin^2 2\phi = 2\cos^2 2\phi - 1$

$$\begin{aligned}\tilde{\xi}_-(r) = \int \frac{d^2\ell}{(2\pi)^2}\left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right] e^{-\ell^2(A_0(0)-A_0(r))/2}e^{ir\ell\cos\phi} \\ \times \left[2\cos^2 2\phi - 1\right]\exp\left(-\frac{\ell^2}{2}\cos(2\phi)A_2(r)\right) .\end{aligned}$$

We now observe that

$$\cos^2 2\phi \exp\left(-\beta\cos(2\phi)\right) = \frac{d^2}{d\beta^2}\exp\left(-\beta\cos(2\phi)\right)\ .$$

Also using Eq. (7.39) we obtain

$$\begin{aligned}\tilde{\xi}_-(r) = \int \frac{d^2\ell}{(2\pi)^2}\left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right] e^{-\ell^2(A_0(0)-A_0(r))/2} e^{ir\ell\cos\phi} \\ \times\Big(2I_0''(\ell^2A_2(r)/2) - I_0(\ell^2A_2(r)/2) + 2\sum_{n=1}^{\infty}(-1)^n \\ \times[2I_n''(\ell^2A_2(r)/2) - I_n(\ell^2A_2(r)/2)]\cos(2n\phi)\Big)\ .\end{aligned}$$

Here primes indicate the derivative with respect to the argument. Integration over ϕ finally yields

$$\begin{aligned}\tilde{\xi}_-(r) &= \int \frac{\ell\, d\ell}{2\pi}\left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right] e^{-\ell^2(A_0(0)-A_0(r))/2} \\ &\quad\times\Big([2I_0''(\ell^2A_2(r)/2) - I_0(\ell^2A_2(r)/2)]J_0(\ell r) \\ &\quad+2\sum_{n=1}^{\infty}[2I_n''(\ell^2A_2(r)/2) - I_n(\ell^2A_2(r)/2)]J_{2n}(\ell r)\Big), \qquad (7.60) \\ &\simeq \int \frac{\ell\, d\ell}{2\pi}\left[C_\ell^{(\mathcal{E})} - C_\ell^{(\mathcal{B})}\right] e^{-\ell^2(A_0(0)-A_0(r))/2} \\ &\quad\times\left(J_4(\ell r) + \frac{\ell^2}{4}A_2(r)[J_2(\ell r) + J_6(\ell r)]\right)\ . \qquad (7.61)\end{aligned}$$

For the last line we have expanded the general expression to first order in $A_2(r)$. To obtain an accuracy of better than cosmic variance one has to also include the term $\propto A_2(r)^2$, see Challinor & Lewis (2005). A similar calculation gives the cross correlation function,

$$\begin{aligned}\tilde{\xi}_\times(r) &= \int \frac{\ell\, d\ell}{2\pi} C_\ell^{(\mathcal{EM})}\, e^{-\ell^2(A_0(0)-A_0(r))/2} \\ &\quad\times\left(I_0'(\ell^2A_2(r)/2)J_0(\ell r) + 2\sum_{n=1}^{\infty}(-1)^n I_n'(\ell^2A_2(r)/2)J_{2n}(\ell r)\right), \qquad (7.62) \\ &\simeq \int \frac{\ell\, d\ell}{2\pi} C_\ell^{(\mathcal{EM})}\, e^{-\ell^2(A_0(0)-A_0(r))/2}\left(J_2(\ell r) + \frac{\ell^2}{4}A_2(r)[J_0(\ell r) + J_4(\ell r)]\right). \qquad (7.63)\end{aligned}$$

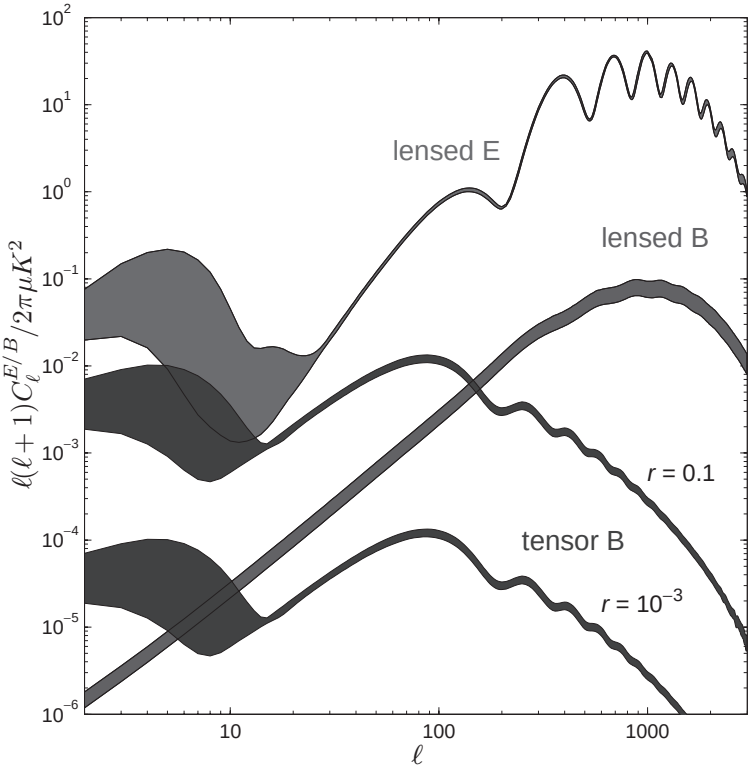


Fig. 7.7. The 95% confidence regions for the polarization spectra from a compilation of the CMB and large-scale structure data available in 2005. The spectral index is fixed, the other parameters are marginalized over a flat ΛCDM model. The tensor B-mode spectra from tensors for $r = 0.1$ and 10^{-3} are also indicated. (Figure from Lewis & Challinor (2006).)

Like for the temperature anisotropy, the polarization power spectra and correlation functions are related via two-dimensional Fourier transforms. The relations between polarization spectra and correlation functions are the same as those for the unlensed quantities given in Chapter 5.

In Fig. 7.7 we show the lensed E- and B-mode spectra for a fixed spectral index and a ΛCDM model. The B-mode spectra from tensors for $r = 0.1$ and $r = 10^{-3}$ are also indicated.

Considering the B-polarization induced by lensing of E-polarization as the only (Gaussian) noise in an all-sky polarization experiment, one finds that the primordial tensor B-mode is detectable for $r \geq 10^{-3}$. If $r < 10^{-3}$, a method must be found to subtract the lensing contribution to the B-polarization spectrum. This is not impossible. At small scales, $\ell > 1000$, the B-mode is nearly entirely due to lensing and can therefore help to determine the spectrum of the lensing potential.

Once we know the latter, we can, in principle, invert our expressions for the lensed spectra to obtain the 'delensed' primordial spectra. The procedure can be applied iteratively. In a first step, one may assume that the B-spectrum is purely due to lensing, but neglect the effect of lensing in the E-spectrum. Determining the spectrum of the lensing potential in this approximation, one can now calculate the first-order delensed E-spectrum and from it the new lensed B-spectrum. The difference of this and the measured B-spectrum is a first estimate for the primordial B-spectrum.

I suppose that this procedure converges rapidly, but this has not been shown as far as I know. Present methods hope to be able to identify the lensing signal with about 10% accuracy (Lewis & Challinor, 2006).

The lensing potential comes dominantly from low redshift and may also be determined or at least constrained by other observations, like e.g. weak lensing especially on small scales. On large scales, where perturbations are linear, we can obtain a first approximation to the lensing potential from parameter estimation, neglecting lensing, which determines the Bardeen potentials Φ, Ψ and then via Eq. (7.19) the lensing potential.

All calculations done in this chapter approximate the sky as flat. This approximation is very good within patches of the size of the deflection angle and up to a few degrees. But if we want to determine the modifications of the low C_ℓs by lensing, sky curvature has to be taken into account. The result from such a curved sky treatment for the correlation between two directions $\mathbf{n}_1$ and $\mathbf{n}_2$ is very similar to our formulae (7.41) and (7.59), (7.61) and (7.63). Only the sums over the modified Bessel functions become double sums over m and m' and the ordinary Bessel functions of r are replaced by elements $D^\ell_{mm'}$ of the representation matrix of the rotation which turns $\mathbf{n}_1$ into $\mathbf{n}_2$. The detailed expressions can be found in Lewis & Challinor (2006).

7.5 Non-Gaussianity

There is also another mean which may help to single out the lensing contribution to the CMB power spectra. This is statistical in nature: we usually assume that primordial fluctuations are Gaussian. However, lensing, being a second-order effect, is not Gaussian. Generically it has a non-vanishing 3-point and connected 4-point correlation function. This, in principle, also offers a possibility of identifying the lensing part of the CMB anisotropies and polarization. However, since all forms of noise, foregrounds, instrumental noise etc. are, in general, non-Gaussian, it may be difficult to use it to remove lensing. Nevertheless, the 3-point and connected 4-point correlation functions from lensing can be calculated and they have a well

defined scale dependence which may help to identify them. In Lewis & Challinor (2006) the subject of non-Gaussianity is treated in more detail.

7.6 Other second-order effects

Before this chapter, we have calculated CMB anisotropies and polarization entirely within linear perturbation theory. In this chapter we have discovered, that for $\ell > 1000$, a 5% accuracy requires to include lensing. At $\ell \gtrsim 3000$, lensing induces changes of order unity and more. Only for $\ell \lesssim 400$, neglecting lensing is accurate to better than $\frac{1}{2}\%$. This naturally brings up the question: are there other second-order effects which are similarly important? The answer to this question is 'probably not, at least not upto $\ell \sim 2000$'.

Several second-order effects have been considered in the literature, but a systematic study of second-order CMB anisotropies and polarization is still lacking. Here I briefly explain the physical effects that have been studied so far, but we do not enter into their calculation. We shall, however discuss the Sunyaev–Zel'dovich (SZ) effect in the next chapter in some detail.

- **Lensing by clusters.** So far we have discussed lensing mainly using the linear lensing potential. However, CMB photons are also lensed by non-linear structures like individual clusters. Statistically this effect can be taken into account by using the non-linear lensing potential on small scales.
- **Ostriker–Vishniac effect.** This is a second-order Doppler term which comes from the fact that the optical depth is proportional to the electron density and the Doppler term therefore has a second-order contribution of the form $\mathbf{n} \cdot \mathbf{V}_b D_b$, where D_b is the baryon density fluctuation, $\mathbf{V}_b$ the baryon velocity and $\mathbf{n}$ the photon direction. It has been argued in Hu & White (1996) that this term is less affected by Silk damping than the first-order Doppler term and may thus become important on small scales, $\ell \gg 1000$. Calculations show, however, that the effect is smaller than the lensing contribution for all $\ell \lesssim 3000$, see Lewis & Challinor (2006).
- **SZ effect.** Clusters contain a hot plasma with a temperature of several keVs. As we shall see in the next chapter, whenever CMB photons pass through a plasma with hot electrons, their spectrum is modified in a well defined way. This is the *thermal SZ effect*. It can be distinguished from primordial CMB anisotropies by its spectral signature. However, clusters usually have a coherent peculiar velocity and CMB photons which scatter off hot electrons of a cluster also acquire a Doppler shift, the so-called *kinetic SZ effect*, which has exactly the same spectrum as ordinary CMB anisotropies. The kinetic SZ effect is typically of the same order as the Vishniac effect.
- **Rees–Sciama effect.** The gravitational potential from linear perturbation theory is constant in a CDM background and decaying in a late ΛCDM background. However, once density perturbations become non-linear, the gravitational potential also starts growing.

This leads to a late integrated Sachs–Wolfe effect on very small scales. Estimations show that this effect is probably subdominant on all scales (Seljak, 1996a).

- **Patchy reionization.** As we discussed in Chapter 6, reionization is supposed to be caused by the radiation of the first, probably very massive stars. It is reasonable to expect that these stars have not formed everywhere at the same time and with the same number density. Therefore, reionization was probably earlier in some patches than in others, leading to more rescattering and therefore damping of CMB anisotropies and regeneration of polarization in some places than in others. It is not clear what and how strong the signature of this 'patchy reionization' is, but it probably is relevant only on very small scales where it may be comparable with the kinetic SZ and Ostriker–Vishniac effects. A study with somewhat unrealistic parameters can be found in Knox *et al.* (1998).

Exercises

(The exercises marked with an asterisk are solved in Appendix A10.6.)

Ex. 7.1 The deflection angle from a point mass∗

Consider a point mass M with gravitational potential $\Psi = GM/r$. Approximate the Schwarzschild metric for this mass by

$$ds^2 = -(1+2\Psi)\,dt^2 + (1-2\Psi)\,d\mathbf{x}^2 \,.$$

Show that the light deflection in this metric to first order in Ψ is given by $\boldsymbol{\alpha} = \varphi\mathbf{e}$ where $\mathbf{e}$ is the normal to the original photon direction $\mathbf{n}$ in the plane defined by $\mathbf{n}$ and the position of the mass, and

$$\varphi = \frac{4GM}{d} \,, \tag{7.64}$$

where d is the impact parameter of the photon trajectory (i.e. its closest distance to the mass M).

Ex. 7.2 Lensing of a scale-invariant power spectrum

In Section 7.3.1 we show that lensing of a scale-invariant spectrum by a small deflection angle, $|\boldsymbol{\alpha}| \ll \pi/\ell$, can be approximated by

$$\tilde{C}_\ell \simeq (1-R^\psi)C_\ell + \ell^2 C_\ell \int \frac{d^2\ell_1}{(2\pi)^2} \frac{[(\boldsymbol{\ell}_1-\boldsymbol{\ell})\cdot\boldsymbol{\ell}_1]^2}{(\boldsymbol{\ell}-\boldsymbol{\ell}_1)^2} C^\psi_{\ell_1} \,.$$

Bring the above integral into the form

$$\int \frac{d^2\ell_1}{(2\pi)^2} \frac{[(\boldsymbol{\ell}_1-\boldsymbol{\ell})\cdot\boldsymbol{\ell}_1]^2}{(\boldsymbol{\ell}-\boldsymbol{\ell}_1)^2} C^\psi_{\ell_1} = \int_0^\infty \frac{\ell_1\,d\ell_1}{(2\pi)^2} C^\psi_{\ell_1} \int_0^{2\pi} d\phi\, \frac{(\ell_1^2-\ell_1\ell\cos\phi)^2}{\ell_1^2+\ell^2-2\ell_1\ell\cos\phi} \,.$$

Using complex integration, show that the angular integral gives

$$\int_0^{2\pi} d\phi \frac{(\ell_1^2-\ell_1\ell\cos\phi)^2}{\ell_1^2+\ell^2-2\ell_1\ell\cos\phi} = \pi\left(\ell_1^2 + \theta(\ell_1-\ell)(\ell_1^2-\ell^2)\right) \,.$$

Here θ is the Heaviside function,

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \,. \end{cases}$$

This implies the result (7.30).

Ex. 7.3 Expectation values of Gaussian variables

Show that for a Gaussian variable X with mean zero and variance σ we have

$$\langle e^{iX} \rangle = e^{-\sigma^2/2} \,.$$

8

The CMB spectrum

The observed frequency power spectrum of the CMB is perfectly approximated by a Planck spectrum, see Fig. 1.7. In our units, $\hbar = c = 1$, it is given by

$$I(\omega) = 4\pi f(\omega) = \frac{1}{\pi^2} \frac{\omega^3}{e^{\omega/T} - 1} \,. \tag{8.1}$$

No deviation from this spectrum has been observed so far.

In this chapter we discuss physical processes which might lead to spectral distortions. We first introduce the collisional processes relevant at temperatures $T < m_e$. These are elastic Compton scattering, double Compton scattering and Bremsstrahlung. We derive the Boltzmann equation for this processes and calculate the relevant timescales. In Section 8.2 we analyse how the injection of high-energy photons e.g. by the decay of a long-lived unstable particle modifies the CMB spectrum. In the final section, we discuss what happens when the CMB photons pass through a hot electron gas affected only by elastic Compton scattering. We shall see that this leads, in general, to a so-called Compton-y distortion of the spectrum. We estimate the effect from the passage of CMB photons through a cluster of galaxies and discuss observations.

8.1 Collisional processes in the CMB

8.1.1 Generalities

At very high temperature many collisional processes keep the cosmic background radiation in thermal equilibrium with itself and all other particles. As we saw in Section 1.4, at $T \simeq 1.4$ MeV weak interactions drop out of equilibrium and neutrinos cease to interact. They are not heated by the decay of electron–positron pairs, which takes place at $T \sim m_e \simeq 500$ keV. Below that temperature, but before recombination, elastic Compton scattering, $e + \gamma \to e + \gamma$, double Compton

scattering, $e + \gamma \rightarrow e + 2\gamma$ and Bremsstrahlung, $e + X \rightarrow e + X + \gamma$ keep the CMB thermalized. Here X denotes an atomic nucleus, usually a proton or a helium-4 nucleus.

As we shall see below, at a redshift of about $z_\mu \simeq 10^7$ also Bremsstrahlung and double Compton drop out of equilibrium and only elastic Compton scattering is still active. This can still distribute the CMB photons in energy, but it does not change their number density. Therefore, energy injection after z_μ leads to a Bose–Einstein distribution with a non-vanishing chemical potential.

In this section we derive the equations which govern the evolution of the photon distribution function in the temperature range $m_e > T > T_{\rm rec}$. We also calculate the timescales for the above mentioned processes and the redshifts above which they are faster than the expansion timescale, i.e., above which they efficiently keep the photon distribution in thermal and 'chemical' equilibrium (by the latter we mean that no chemical potential is developed).

8.1.2 Elastic Compton scattering and the Kompaneets equation

We want to derive a differential equation which describes the thermalization of photons when they interact with electrons which also have a thermal distribution but may be at a different temperature $T_e \neq T$. We consider a photon with initial energy ω and final energy ω' and a non-relativistic electron with initial velocity $v \ll 1$ and final velocity $v' \ll 1$. Hence we must also require $\omega, \omega' \ll m_e$. In the centre of mass system, energy and momentum of the two particles remain unchanged, of course, in a two-body interaction. This is a simple consequence of energy and momentum conservation. However, in the laboratory frame, if initially the electron momentum is much larger than the photon momentum, after the collision the photon will have gained energy while the electron has lost and vice versa. Let us first study the process in the frame in which the initial electron momentum vanishes, so that by energy and momentum conservation

$$\omega\mathbf{n} = m_e\mathbf{v}' + \omega'\mathbf{n}' \quad \text{and} \quad \omega = \frac{1}{2}m_e v'^2 + \omega' = \frac{1}{2m_e}(\omega\mathbf{n} - \omega'\mathbf{n}')^2 + \omega' \,, \tag{8.2}$$

where $\mathbf{n}$ and $\mathbf{n}'$ denote the photon direction before and after the collision. If we neglect all terms of order ω/m_e, we find $\omega = \omega'$. This would imply no change in the photon frequency. In this approximation we obtain non-relativistic Thomson scattering where only the direction of the photon but not its energy is affected. The energy difference is of the order ω^2/m_e, where ω is either ω or ω'. Taking this difference into account to lowest order we may set $\omega^2 = \omega'^2 = \omega\omega'$ in the term

proportional to $1/(2m_e)$. With this Eq. (8.2) yields

$$\frac{\omega}{\omega'} = 1 + \left(\frac{\omega}{m_e}\right)(1 - \cos\vartheta)\,, \tag{8.3}$$

where ϑ is the scattering angle, $\cos\vartheta = \mathbf{n}\cdot\mathbf{n}'$. To first order in ω/m_e this can be written as

$$\frac{\omega' - \omega}{\omega} \equiv \frac{\Delta\omega}{\omega} = -\left(\frac{\omega}{m_e}\right)(1 - \cos\vartheta)\,. \tag{8.4}$$

In a generic frame, denoting the photon and electron 4-momentum before and after the scattering process by $p_\gamma = \omega(1, \mathbf{n})$ and $p_e = (E, \mathbf{p})$ respectively $p'_\gamma = \omega'(1, \mathbf{n}')$ and $p'_e = (E', \mathbf{p}')$ energy–momentum conservation implies

$$(p'_e)^2 = m_e^2 = (p_e + p_\gamma - p'_\gamma)^2 = m_e^2 + 2p_e(p_\gamma - p'_\gamma) - 2p_\gamma p'_\gamma\,,$$

which yields

$$\begin{aligned} 0 &= E(\omega - \omega') - \mathbf{p}\cdot(\omega\mathbf{n} - \omega'\mathbf{n}') - \omega\omega' + \omega\omega'\mathbf{n}\cdot\mathbf{n}' \\ &= E(\omega - \omega') - \mathbf{p}\cdot\mathbf{n}'(\omega - \omega') + \omega(\omega - \omega')(1 - \mathbf{n}\cdot\mathbf{n}') \\ &\quad + \omega\mathbf{p}\mathbf{n}' - \omega^2(1 - \mathbf{n}\cdot\mathbf{n}') - \omega\mathbf{p}\cdot\mathbf{n}\,. \end{aligned}$$

Defining $x_e \equiv \omega/T_e$, we obtain for the energy difference ($\frac{T_e}{E} \cong \frac{P^2}{2m_e^2} - \frac{V^2}{2}$)

$$\Delta \equiv \frac{\omega' - \omega}{T_e} = \frac{x_e\mathbf{p}(\mathbf{n}' - \mathbf{n}) - x_e^2 T_e(1 - \mathbf{n}\cdot\mathbf{n}')}{E - \mathbf{p}\cdot\mathbf{n}' + x_e T_e(1 - \mathbf{n}\cdot\mathbf{n}')} \simeq \frac{x_e\mathbf{p}(\mathbf{n}' - \mathbf{n})}{m_e}\,. \tag{8.5}$$

The energy transfer, $\omega' - \omega$ is of order ωv, i.e. it is suppressed by a factor v.

To calculate how the photon distribution $f(\omega)$ changes by elastic Compton scattering we write down the Boltzmann equation. Neglecting the expansion of the Universe and perturbations we have (see Section 4.5)

$$\frac{\partial f}{\partial t}(\omega) = \frac{df_+}{dt}(\omega) - \frac{df_-}{dt}(\omega) \equiv C[f](\omega)\,, \tag{8.6}$$

where df_+/dt denotes the phase space density of photons which are scattered into the energy range $[\omega, \omega + d\omega]$ per unit time and df_-/dt denotes the density of photons scattered out of this energy range. We assume that f is independent of direction and position. Contrary to the situation in Section 4.5 we now consider a distribution of electrons in momentum space which we denote by $f_e(E)$. The collision integral now becomes

$$\begin{aligned} &C[f](\omega) \\ &\quad = \int d^3p \int d\Omega_{\mathbf{n}'} \frac{d\sigma}{d\Omega}\left\{f(\omega')[1 + f(\omega)]f_e(E) - f(\omega)[1 + f(\omega')]f_e(E')\right\}\,. \end{aligned} \tag{8.7}$$

Here d^3p denotes integration over electron momenta with $E = \sqrt{m_e^2 + p^2} \simeq m_e + p^2/2m_e$ and $d\Omega_{\mathbf{n}'}$ denotes integration over photon directions. The variables $\omega' = \omega + T_e\Delta$ and $E' = E - T_e\Delta$ are eliminated via Eq. (8.5). The factors $1 + f$ take into account the quantum effect of stimulated emission for photons which are bosons. For the electrons we should, in principle, include a factor $[1 - f_e]$ due to their fermionic nature, but we assume that the electron gas is sufficiently diluted, $f_e \ll 1$, so that we may neglect this quantum correction. The Compton scattering cross section is given by Eq. (4.109)

$$\frac{d\sigma}{d\Omega} = \frac{3}{16\pi}\sigma_T \left(1 + (\mathbf{n}\cdot\mathbf{n}')^2\right) .$$

Strictly speaking, this is the cross section in the electron rest frame and when transforming it to the laboratory frame the photon directions and $(\mathbf{n}\cdot\mathbf{n}')^2$ change due to aberration. But as we shall argue, this effect can be neglected for non-relativistic electrons (up to order v^2).

We now expand the integrand of Eq. (8.7) to second order in the small energy transfer $\omega' - \omega = E - E' \sim \mathcal{O}(\omega v)$.

$$\begin{aligned} f(\omega') &= f(\omega) + \Delta\frac{\partial f}{\partial x_e} + \frac{\Delta^2}{2}\frac{\partial^2 f}{\partial x_e^2} + \cdots \\ f_e(E') &= f_e(E) - T_e\Delta\frac{\partial f_e}{\partial E} + \frac{T_e^2\Delta^2}{2}\frac{\partial^2 f_e}{\partial E^2} + \cdots \\ &= f_e(E)\left[1 + \Delta + \frac{\Delta^2}{2} + \cdots\right] . \end{aligned}$$

For the last equality sign we have assumed that the electrons obey a Maxwell distribution, $f_e(E) \propto \exp(-E/T)$. Inserting this expansion in Eq. (8.7), we find

$$\frac{\partial f}{\partial t}(\omega) = \left[\frac{\partial f}{\partial x_e} + f(1+f)\right] I_1 + \frac{1}{2}\left[\frac{\partial^2 f}{\partial x_e^2} + 2(1+f)\frac{\partial f}{\partial x_e} + f(1+f)\right] I_2 \tag{8.8}$$

with

$$I_1 = \int d^3p \int d\Omega_{\mathbf{n}'} \frac{d\sigma}{d\Omega} f_e(E)\Delta , \tag{8.9}$$

$$I_2 = \int d^3p \int d\Omega_{\mathbf{n}'} \frac{d\sigma}{d\Omega} f_e(E)\Delta^2 . \tag{8.10}$$

We want to calculate these integrals up to order v^2. The second integral is readily performed. Since the lowest-order approximation to Δ is already of order v we can simply set $\Delta^2 = (x_e/m_e)^2(\mathbf{p}\cdot(\mathbf{n}' - \mathbf{n}))^2$. Inserting this in Eq. (8.10) and choosing

the p_3-direction along $\mathbf{n}' - \mathbf{n}$ we obtain

$$\begin{aligned} I_2 &= \frac{x_e^2}{m_e^2} \int \frac{d\sigma}{d\Omega}\, d\Omega \int d^3p f_e(E) p^2 (\mathbf{n}' - \mathbf{n})^2 \cos^2\theta \\ &= \frac{4\pi x_e^2}{3m_e^2} \int \frac{d\sigma}{d\Omega} (\mathbf{n}' - \mathbf{n})^2\, d\Omega \int_0^\infty dp\; p^4 f_e(E) \\ &= \frac{T_e n_e x_e^2}{m_e} \int \frac{d\sigma}{d\Omega} (\mathbf{n}' - \mathbf{n})^2 \,. \end{aligned}$$

We have used the fact that to lowest order in v, $f_e' = -(p/m_e T_e) f_e$ so that $p^4 f_e = -m T_e p^3 f_e' = m_e T_e [-(p^3 f_e)' + 3p^2 f_e]$. The p-integral over the first term in the square bracket does not contribute while the integral over the second term gives $3n_e/(4\pi)$, where n_e denotes the electron density. We still have to integrate over scattering angles,

$$\begin{aligned} \int \frac{d\sigma}{d\Omega} (\mathbf{n}' - \mathbf{n})^2 d\Omega &= \frac{3\sigma_T}{16\pi} \int d\Omega (1 + \cos^2\vartheta)(2 - 2\cos\vartheta) \\ &= \frac{3\sigma_T}{4} \int_{-1}^{1} d(\cos\vartheta)[1 + \cos^2\vartheta] = 2\sigma_T \,. \end{aligned}$$

We finally obtain

$$I_2 = 2 n_e \sigma_T \frac{T_e}{m_e} x_e^2 \,. \tag{8.11}$$

Note that $T_e/m_e \sim p^2/(2m_e^2) \sim v^2$ hence the term is of the required order of magnitude.

The calculation of I_1 is more tricky. To lowest order $\Delta \propto \mathbf{p}(\mathbf{n}' - \mathbf{n})$ so that the integral over d^3p vanishes. (Were this not the case, this term which is of order v would by far dominate all other contributions and also I_2.) We therefore have to include the next order. Expanding Δ to the next order gives

$$\Delta \simeq \frac{x_e}{m_e} \mathbf{p} \cdot (\mathbf{n}' - \mathbf{n}) \left[1 + \frac{\mathbf{p} \cdot \mathbf{n}'}{m} \right] - \frac{x_e^2}{m_e} T_e (1 - \mathbf{n} \cdot \mathbf{n}') \,.$$

In addition, we have to take into account the fact that the photon density seen by the electron (in its rest frame) is not $4\pi f \omega^2\, d\omega$ but $4\pi f \omega^2\, d\omega\, (1 - \mathbf{p} \cdot \mathbf{n}/m_e)^3 \simeq 4\pi f \omega^2\, d\omega\, (1 - 3\mathbf{p} \cdot \mathbf{n}/m_e)$, to lowest order in $v \simeq p/m_e$. Therefore, the integral which we really have to compute is not the one given in Eq. (8.9) but

$$I_1 = \int d^3p \int d\Omega_{\mathbf{n}'} \frac{d\sigma}{d\Omega} f_e(E) (1 - 3\mathbf{p} \cdot \mathbf{n}/m_e) \Delta \,, \tag{8.12}$$

with

$$\begin{aligned} (1 - 3\mathbf{p} \cdot \mathbf{n}/m_e)\Delta = \frac{x_e}{m_e} \mathbf{p} \cdot (\mathbf{n}' - \mathbf{n}) \left[1 + \frac{\mathbf{p} \cdot \mathbf{n}'}{m} - \frac{3\mathbf{p} \cdot \mathbf{n}}{m} \right] \\ - \frac{x_e^2}{m_e} T_e (1 - \mathbf{n} \cdot \mathbf{n}') + \mathcal{O}(v^3) \,. \end{aligned}$$

Of course this correction also applies to I_2, but there it is subdominant and does not enter up to order v^2. To order v^2, in principle, aberration also has to be taken into account. We should replace $\mathbf{n}\cdot\mathbf{n}'$ by the corresponding the scalar product in the electron rest frame, $\mathbf{n}_R(\mathbf{n},\mathbf{v})\cdot\mathbf{n}'_R(\mathbf{n}',\mathbf{v})$ in $d\sigma/d\Omega$. Here a subscript R denotes the electron rest frame (see Ex. 8.2). However the resulting expression will always be symmetrical in $\mathbf{n}$ and $\mathbf{n}'$. Since to lowest order it is multiplied with the anti-symmetrical factor $\mathbf{p}\cdot(\mathbf{n}'-\mathbf{n})$, its angular integral vanishes.

Using as before that

$$\int d^3p\,(\mathbf{p}\cdot\mathbf{n})^2 f_e(E) = m_e T_e n_e\,, \qquad \int d^3p\, f_e = n_e\,,$$

and

$$\int d^2p\,(\mathbf{p}\cdot\mathbf{n})(\mathbf{p}\cdot\mathbf{n}')f_e(E) = (\mathbf{n}\cdot\mathbf{n}')\frac{4\pi}{3}\int_0^\infty dp\, p^4 f_e = (\mathbf{n}\cdot\mathbf{n}')m_e T_e n_e\,,$$

we find

$$\int d^3p\, f_e(1-3\mathbf{p}\cdot\mathbf{n}/m_e)\Delta = \left[4\frac{T_e n_e x_e}{m_e} - \frac{x_e^2 T_e n_e}{m_e}\right](1-\mathbf{n}\cdot\mathbf{n}')\,.$$

The integral over photon directions gives

$$\int d\Omega\,\frac{d\sigma}{d\Omega}(1-\mathbf{n}\cdot\mathbf{n}') = \int d\Omega\,\frac{d\sigma}{d\Omega} = \sigma_T\,.$$

Putting this together we obtain

$$I_1 = \frac{\sigma_T n_e T_e}{m_e} x_e(4-x_e)\,.$$

Inserting the results for I_1 and I_2 in Eq. (8.8), we obtain the Kompaneets equation

$$\frac{m_e}{T_e}\frac{1}{n_e\sigma_T}\frac{\partial f}{\partial t} = \frac{1}{x_e^2}\frac{\partial}{\partial x_e}\left[x_e^4\left(\frac{\partial f}{\partial x_e} + f + f^2\right)\right]\,. \tag{8.13}$$

Solving the time-dependent Kompaneets equation in order to study 'Comptonization' of a photon distribution on thermal electrons in full generality can only be achieved numerically. However, there are important situations where meaningful analytical results can be obtained.

First of all, as it should, the photon number density remains unchanged by evolution under the Kompaneets equation,

$$\frac{dn_\gamma}{dt} \propto \int dx_e\, x_e^2\frac{\partial f}{\partial t} \propto \int dx_e \frac{\partial}{\partial x_e}\left[x_e^4\left(\frac{\partial f}{\partial x_e} + f + f^2\right)\right] = 0\,. \tag{8.14}$$

Furthermore, a Bose–Einstein distribution with temperature T_e, hence $f_{BE} = \left(e^{(\omega/T_e+\mu)}-1\right)^{-1}$ is the (unique) equilibrium solution of this equation, $\frac{d}{dt}f_{BE} = 0$.

Let us also write the equation in terms of the photon energy ω instead of $x_e = \omega/T_e$,

$$\frac{m_e}{n_e\sigma_T}\frac{\partial f}{\partial t} = \frac{1}{\omega^2}\frac{\partial}{\partial \omega}\left[\omega^4\left(T_e\frac{\partial f}{\partial \omega} + f + f^2\right)\right] . \tag{8.15}$$

We finally write the Kompaneets equation in the form

$$\frac{\partial f}{\partial t} = \frac{1}{\tau_K}\frac{1}{x_e^2}\frac{\partial}{\partial x_e}\left[x_e^4\left(\frac{\partial f}{\partial x_e} + f + f^2\right)\right] , \tag{8.16}$$

with

$$\tau_K = \frac{m_e}{T_e}\frac{1}{n_e\sigma_T} \simeq 10^{28}\,\mathrm{s}\,(1 - Y_{He}/2)^{-1}(\Omega_b h^2)^{-1}\frac{T}{T_e}z^{-4} , \tag{8.17}$$

where we have used

$$n_e = n_p = n_B(1 - Y_{He}/2) = \frac{3\Omega_b H_0^2(1 - Y_{He}/2)}{8\pi G m_p},$$

and we have set the present photon temperature to $T(z = 0) = T_0 = 2.7\,\mathrm{K}$ and m_p is the proton mass.

In equilibrium, $f = f_{BE} = 1/(e^{x_e+\mu} - 1)$, it is easy to see that the electron temperature is

$$T_e = \frac{1}{4}\frac{\int d\omega\,\omega^4 f(f + 1)}{\int d\omega\,\omega^3 f} . \tag{8.18}$$

This is the electron temperature whenever the electrons are in thermal equilibrium with the photons, even if the photons are not in thermal equilibrium with the electrons. The timescale for elastic Compton scattering of the electrons is of the order

$$t_e \simeq \frac{n_e}{n_\gamma}t_K \simeq 10^{-10}t_K ,$$

which is much shorter than all other timescales involved. Therefore, instead of including a kinetic equation for the electrons, we shall always assume that they are in thermal equilibrium with the photons and therefore follow a Boltzmann distribution at temperature T_e given by Eq. (8.18). Note that it needs only a fraction of about 10^{-10} of all photons for one scattering event on each electron. This huge mismatch, $n_\gamma \gg n_e$, leads to the somewhat unusual behaviour, that electrons are much more rapidly thermalized than photons.

8.1.3 Thermal Bremsstrahlung

According to the Larmor formula (see Jackson, 1975) an accelerated electron emits electromagnetic radiation. For non-relativistic electrons, the energy emitted per unit

time is

$$\frac{d\mathcal{E}}{dt} = \frac{2\alpha}{3}|\mathbf{a}(t)|^2 \,, \tag{8.19}$$

where $\mathbf{a}(t)$ denotes the acceleration and α is the fine structure constant, see Appendix A1.2. The radiation spectrum is obtained by Fourier transforming $\mathbf{a}$. If the period of acceleration is a short interval Δt, over which the velocity changes by an amount Δv, one obtains, see Padmanabhan (2000)

$$\frac{d\mathcal{E}}{d\omega} = \begin{cases} \frac{2}{3\pi}\alpha(\Delta v)^2 \,, & \omega \ll (\Delta t)^{-1} \\ 0 \,, & \omega \gg (\Delta t)^{-1} \,. \end{cases}$$

We now consider an electron which passes by an ion X of charge Ze, with impact parameter b and initial velocity v. We assume the change in the velocity to be small so that we can only take into account the component normal to the initial velocity and integrate the equation of motion, $\mathbf{a} \simeq \mathbf{a}_\perp = \mathbf{e}_\perp \alpha/(m_e r^2)$, along the unperturbed path. With $|\mathbf{e}_\perp| = b/\sqrt{b^2 + (vt)^2}$, this gives the velocity change

$$\Delta v = \frac{Z\alpha}{m_e}\int_{-\infty}^{\infty} \frac{b}{[b^2 + (vt)^2]^{3/2}}\, dt = \frac{2Z\alpha}{m_e b v} \,.$$

The acceleration is important during a time interval of about $\Delta t = b/v$ around the closest encounter at $t = 0$. Inserting this in the above expression for the radiated energy spectrum we obtain

$$\frac{d\mathcal{E}}{d\omega} = \begin{cases} \frac{8Z^2\alpha^3}{3\pi m_e^2 v^2 b^2} \,, & \text{if} \quad \omega \ll v/b \\ 0 \,, & \text{if} \quad \omega \gg v/b \,. \end{cases} \tag{8.20}$$

We now want to determine the energy emitted by an ion density n_i and an electron density n_e. The electron flux incident on one ion is simply $n_e v$ and the surface area with a given impact parameter b is $2\pi b\, db$. Multiplying by the ion density and integrating over the impact parameter, we obtain the energy emitted per volume per unit time and per frequency,

$$\frac{d\mathcal{E}(\omega, v)}{dV\, dt\, d\omega} = \frac{16\alpha^3 Z^2}{3m_e^2 v} n_i n_e \int_{b_{\min}}^{b_{\max}} \frac{db}{b} = \frac{16\alpha^3 Z^2}{3m_e^2 v} n_i n_e \log\left(\frac{b_{\max}}{b_{\min}}\right) \,. \tag{8.21}$$

Here $b_{\max}$ is determined by the maximal impact parameter which can still produce photons with frequency ω, $b_{\max} = v/\omega$. The minimum impact parameter can be estimated in two ways. First we can take it as the smallest value for which the straight line approximation which we have used to determine Δv is still reasonable, this is roughly when $\Delta v \sim v$. Inserting this in the above expression for Δv yields

$$b_{\min,1} = \frac{2Z\alpha}{m_e v^2} \,.$$

On the other hand, the uncertainty principle requires $p = m_e v > \hbar/b$ so that

$$b_{\min,2} = \frac{\hbar}{m_e v} \, .$$

The correct value for $b_{\min}$ is whichever of the two values is larger.

In general, one casts the uncertainty of this logarithmic term in a so-called Gaunt factor defined by

$$g_{\rm ff} = \frac{\sqrt{3}}{\pi} \log\left(\frac{b_{\max}}{b_{\min}}\right) \, . \tag{8.22}$$

The correct expression for the Gaunt factor has to be obtained by a quantum mechanical treatment (see e.g. Padmanabhan (2000)).

We now want to average Eq. (8.21) over electron velocities which follow a Maxwell distribution, $f_e \propto \exp(-m_e v^2/(2T_e))$

$$\frac{d\mathcal{E}(\omega,T)}{dV\,dt\,d\omega} = \frac{\int_{v_{\min}}^{\infty} dv \frac{d\mathcal{E}(\omega,v)}{dV\,dt\,d\omega} v^2 \exp(-m_e v^2/(2T_e))}{\int_0^{\infty} dv\, v^2 \exp(-m_e v^2/(2T_e))} \, . \tag{8.23}$$

Here $v_{\min}$ is the minimal velocity which can generate a photon of energy ω, $\omega = v_{\min}^2/(2m)$. Neglecting the weak velocity dependence of the Gaunt factor, the integral in the numerator is elementary and the one in the denominator simply gives the mean square velocity, $2T_e/m$, so that we arrive at

$$\frac{d\mathcal{E}(\omega,T)}{dV\,dt\,d\omega} = \frac{16\alpha^3 Z^2}{3m_e}\left(\frac{2\pi}{3m_e}\right)^{1/2} T_e^{-1/2} n_i n_e e^{-\omega/T_e} \bar{g}_{\rm ff}(T,\omega) \, . \tag{8.24}$$

The modification of this formula obtained from a correct quantum treatment of Bremsstrahlung can be absorbed in the dimensionless Gaunt factor.

This emitted Bremsstrahlung changes the photon energy spectrum. Since, according to Eq. (1.55),

$$\frac{d\mathcal{E}_\gamma}{dV d\omega} = \frac{d\rho_\gamma}{d\omega} = \frac{1}{\pi^2}\omega^3 f \, ,$$

we can translate Eq. (8.24) into an equation for the change of the photon distribution function. With $\sigma_T = 8\pi\alpha^2/(3m_e^2)$ we find

$$\left[\frac{df}{dt}\right]_{\rm ff\,em} = \frac{\sigma_T\, n_e}{x_e^3\, e^{x_e}}\, Q\, \bar{g}_{\rm ff}, \quad \text{with} \tag{8.25}$$

$$Q = 2\pi\sqrt{\frac{2\pi}{3}}\,\frac{\alpha}{T_e^3}\sqrt{\frac{m}{T_e}}\sum_i Z_i^2 n_i \, , \tag{8.26}$$

where we now sum over the contributions from different ion species. The 'em' in the above subscript stands for 'emission'. We discuss free–free absorption below. In the early Universe, we can approximate the ions by hydrogen and helium-4 only so that $\sum_i Z_i^2 n_i = n_B$. In the energy range of interest, we can approximate the Gaunt factor by, Rybicki & Lightman (1979)

$$\bar{g}_{\rm ff} \simeq \begin{cases} \frac{\sqrt{3}}{\pi} \ln\left(0.37\, e^{\pi/\sqrt{3}}/x_e\right) , & x_e \le 0.37 \\ 1 , & x_e \ge 0.37 . \end{cases} \tag{8.27}$$

Of course, there is not only Bremsstrahlung emission but also absorption. We can either calculate the latter directly or obtain it by the argument of detailed balance: emission and absorption have to cancel exactly in equilibrium, i.e. if the photon distribution is a Planck distribution at temperature $T = T_e$, $f = 1/(e^{x_e} - 1)$. With this and using the fact that absorption is proportional to f, we obtain

$$\left[\frac{df}{dt}\right]_{\rm ff\,ab} = -\left[\frac{df}{dt}\right]_{\rm ff\,em} f(e^{x_e} - 1) .$$

The kinetic equation for Bremsstrahlung then becomes

$$\left[\frac{df}{dt}\right]_{\rm ff} = \frac{\sigma_T\, n_e}{x_e^3\, e^{x_e}}\, Q\, \bar{g}_{\rm ff} \left[1 - f(e^{x_e} - 1)\right] . \tag{8.28}$$

It is convenient to write this equation as

$$\left[\frac{df}{dt}\right]_{\rm ff} = \frac{1}{\tau_{\rm ff}} \frac{\bar{g}_{\rm ff}}{x_e^3\, e^{x_e}} \left[1 - f(e^{x_e} - 1)\right] , \tag{8.29}$$

with

$$\tau_{\rm ff}^{-1} = 2\pi \sqrt{\frac{2\pi}{3}} n_e\, \sigma_T\, n_B \frac{\alpha}{T_e^3} \sqrt{\frac{m_e}{T_e}} \tag{8.30}$$

$$\tau_{\rm ff} \simeq 2.3 \times 10^{23} {\rm s}\, (1 - Y_{He}/2)^{-1} (\Omega_b h^2)^{-2} \left[\frac{T_e}{T}\right]^{7/2} (z+1)^{-5/2} .$$

Here $T = (1+z)T_0$ is the CMB temperature.

8.1.4 Double Compton scattering

It actually turns out that in most cosmological circumstances double Compton scattering is more efficient than Bremsstrahlung, even though the double Compton process $e + \gamma' \leftrightarrow e + \gamma_1 + \gamma$ is second order. We shall see that it is most efficient for very small energies of the second photon γ and therefore we neglect the small energy transfer, assuming that the photons γ' and γ_1 have the same frequency. The

angle integrated double Compton cross section then gives (Jauch & Rorlich, 1976)

$$\frac{d\sigma_{2\gamma}}{d\omega_1\, d\omega} = \frac{4\alpha}{3\pi}\sigma_T \left(\frac{\omega'}{m_e}\right)^2 \frac{1}{\omega}\,\delta(\omega_1 - \omega')\,. \tag{8.31}$$

Here ω' is the energy of the incoming photon, ω_1 is the energy of the incoming photon after the collision and ω is the energy of the photon generated by the collision.

To derive an equation for the distribution function f, we use the fact that the number density of photons in the energy interval ω and $\omega + d\omega$ is given by

$$dn(\omega) = \frac{1}{\pi^2} f(\omega)\omega^2\, d\omega\,.$$

With Eq. (8.31) we then obtain for double Compton emission

$$\begin{aligned}\frac{\omega^2}{\pi^2}\frac{\partial f(\omega)}{\partial t} &= n_e \int d\omega_1\, d\omega' \,\frac{d\sigma_{2\gamma}}{d\omega_1 d\omega}\frac{\omega'^2}{\pi^2} f(\omega') \\ \frac{\partial f(\omega)}{\partial t} &= \frac{4\alpha}{3\pi}\sigma_T\, n_e \int d\omega' \left(\frac{\omega'}{m_e}\right)^2 \frac{\omega'^2}{\omega^3} f(\omega')[f(\omega')+1][f(\omega)+1]\,.\end{aligned} \tag{8.32}$$

In the second equation we have multiplied by the factor $[f(\omega_1)+1][f(\omega)+1] = [f(\omega')+1][f(\omega)+1]$ to take into account stimulated emission.

To obtain the contribution from double Compton absorption, $\gamma + \gamma_1 + e \rightarrow \gamma' + e$, we can again use detailed balance and the fact that absorption is proportional to $f(\omega)f(\omega_1)[f(\omega')+1]$. The full equation for double Compton scattering then takes the form

$$\begin{aligned}\left[\frac{\partial f(\omega)}{\partial t}\right]_{2\gamma} = \frac{4\alpha}{3\pi}\sigma_T\, n_e \int d\omega' \left(\frac{\omega'}{m_e}\right)^2 \frac{\omega'^2}{\omega^3}\Bigg\{ & f(\omega')[f(\omega')+1][f(\omega)+1] \\ & - \exp\left(\frac{\omega}{T_e}\right) f(\omega')f(\omega)[f(\omega')+1]\Bigg\}\,.\end{aligned}$$

With $x_e = \omega/T_e$ and $x'_e = \omega'/T_e$ this yields

$$\begin{aligned}\left[\frac{\partial f}{\partial t}\right]_{2\gamma} &= \frac{4\alpha}{3\pi}\sigma_T\, n_e \left(\frac{T_e}{m_e}\right)^2 \frac{1}{x_e^3}[1 - f(e^{x_e} - 1)]I \\ &= \frac{1}{\tau_{2\gamma}}\frac{1}{x_e^3}[1 - f(e^{x_e} - 1)]I\,,\end{aligned} \tag{8.33}$$

with

$$I = \int dx'_e\, (x'_e)^4 f(x'_e)[1 + f(x'_e)]\,,$$

and

$$\tau_{2\gamma} = \frac{3\pi}{4\alpha\,\sigma_T n_e}\left(\frac{m_e}{T_e}\right)^2 \simeq 7\times 10^{39}\,\mathrm{s}\,\frac{1}{(1-Y_{He}/2)\Omega_b h^2}\left(\frac{T}{T_e}\right)^2 (1+z)^{-5}\,. \tag{8.34}$$

Strictly speaking, this equation is only valid for $x_e < 1$ since the double Compton cross section assumes that the energy of the photon generated by the collision is much smaller than the kinetic energy of the electron. But double Compton scattering becomes very inefficient at high energy, so that we can simply set $[\partial f(x_e)/\partial t]_{2\gamma} = 0$ for $x_e \geq 1$.

8.1.5 Timescales and redshifts

In this section we follow Hu & Silk (1993).

Taking into account Compton scattering, double Compton scattering and Bremsstrahlung, the kinetic equation for the photon distribution becomes

$$\frac{\partial f}{\partial t} = \left[\frac{\partial f}{\partial t}\right]_K + \left[\frac{\partial f}{\partial t}\right]_{\mathrm{ff}} + \left[\frac{\partial f}{\partial t}\right]_{2\gamma}\,.$$

In this equation, cosmic expansion is not taken into account. Accounting also for the dilution of photons due to expansion, we have to add a term $-3Hf$ on the right-hand side and we must take care to distinguish between cosmic time τ and conformal time t. Of course, the above are derivatives with respect to cosmic time so that the kinetic equation in the expanding Universe becomes

$$\frac{\partial f}{\partial \tau} = -\frac{3}{\tau_{\mathrm{exp}}} f + \left[\frac{\partial f}{\partial t}\right]_K + \left[\frac{\partial f}{\partial t}\right]_{\mathrm{ff}} + \left[\frac{\partial f}{\partial t}\right]_{2\gamma}\,. \tag{8.35}$$

Here $\tau_{\mathrm{exp}} = 1/H$ is the expansion timescale which can be approximated by

$$\tau_{\mathrm{exp}}(z) = 1/H(z) \simeq 5\times 10^{19}\,\mathrm{s}\,(z+z_{\mathrm{eq}})^{-1/2}(1+z)^{-3/2}\,, \tag{8.36}$$

where $z_{\mathrm{eq}} = 2.4\times 10^4 \Omega_m h^2$ is the redshift where the matter and radiation densities are equal, see Ex. 8.1. Equation (8.36) holds in a matter/radiation universe, i.e. as long as the cosmological constant (or any other dark energy component) and curvature are negligible.

Let us first compare the expansion time and the timescale for Compton scattering, τ_K. Since $[\partial f/\partial t]_K$ is not simply proportional to $\tau_K^{-1} f$ one has to compare the term on the right-hand side of the Kompaneets equation with $3f/\tau_{\mathrm{exp}}$. For Compton scattering to be efficient, one typically requires (Hu & Silk, 1993) $\tau_K/4 \leq \tau_{\mathrm{exp}}$,

which implies

$$z \geq z_K \sim 10^4 (\Omega_b h^2)^{-1/2} \,. \tag{8.37}$$

Here we have assumed $z_K \gg z_{\rm eq}$. At redshifts below z_K, elastic Compton scattering drops out of equilibrium and only Thomson scattering, which does not change the photon energies is relevant.

Also of interest is the redshift at which double Compton scattering becomes more important than Bremsstrahlung (i.e. free–free). Considering the redshift dependence of $\tau_{\rm ff}$ and $\tau_{2\gamma}$ it is clear that at high redshift double Compton is more efficient. For our estimates, we may replace the integral I in Eq. (8.33) by its value for a Planck spectrum, $I \simeq I_P \simeq 26$. Furthermore, both processes are most efficient at small $x_e \ll 1$ where Eqs. (8.29) and (8.33) with $I \simeq 26$ yield $[df/dt]_{2\gamma} \simeq 26/(\tau_{2\gamma} x_e^3)$ and $[df/dt]_{\rm ff} \simeq \bar{g}_{\rm ff}(x_e)/(\tau_{\rm ff} x_e^3)$. Equating these two terms, neglecting the small temperature difference, $T_e \simeq T$, yields

$$z_{{\rm ff},2\gamma} \sim 10^6 \left(\Omega_b h^2 \bar{g}_{\rm ff}(x_e)\right)^{2/5} \quad , \quad x_e \ll 1 \,. \tag{8.38}$$

Above this redshift, double Compton scattering is more efficient than free–free.

We also determine the energy at which double Compton scattering or Bremsstrahlung become as efficient as elastic Compton scattering. For this we use the fact that

$$\left[I_P^{-1} \tau_{2\gamma} \frac{x_e^3}{e^{x_e}-1}\right] \left[\frac{\partial f}{\partial t}\right]_{2\gamma} \sim \frac{1}{e^{x_e}-1} - f \quad \text{and}$$

$$\left[\frac{e^{x_e}}{\bar{g}_{\rm ff}} \tau_{\rm ff} \frac{x_e^3}{e^{x_e}-1}\right] \left[\frac{\partial f}{\partial t}\right]_{\rm ff} \sim \frac{1}{e^{x_e}-1} - f \,.$$

Hence, within the timescale in square brackets, double Compton and Bremsstrahlung respectively, are able to establish a Planck spectrum. Equating this timescale to the Compton timescale, $\tau_K/4$, we obtain a redshift dependent energy below which double Compton and Bremsstrahlung respectively, are efficient,

$$x_{c,2\gamma}(z) = 3 \times 10^{-6} \sqrt{z} \,, \tag{8.39}$$

$$x_{c,{\rm ff}}(z) = 77\, z^{-3/4} (\Omega_b h^2)^{1/2} \,. \tag{8.40}$$

At energies below $x_c \equiv \sqrt{x_{c,2\gamma}^2 + x_{c,{\rm ff}}^2}$, photon number changing processes are efficient and a Planck spectrum can be established, if $z > z_K$ so that elastic Compton scattering is still efficient. At energies above x_c, photons settle into a Bose–Einstein distribution, if $z > z_K$.

Correspondingly, comparing double Compton scattering and Bremsstrahlung with expansion, we find that at a given redshift z double Compton scattering or Bremsstrahlung are still efficient only for photon energies with $x_e < x_{{\rm exp},2\gamma}(z)$ or

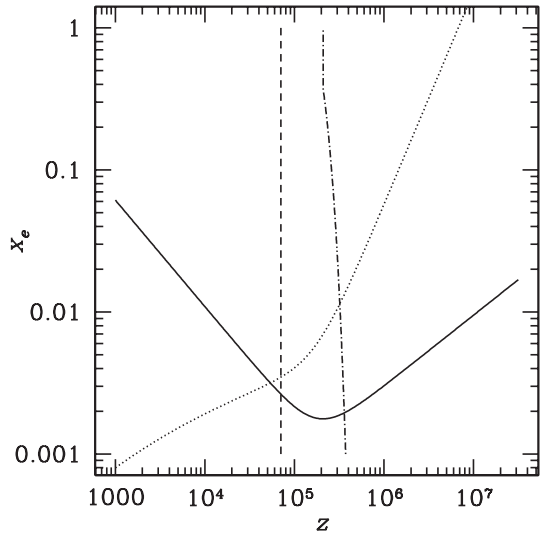


Fig. 8.1. The redshifts z_K (dashed) and $z_{\mathrm{ff},2\gamma}(x_e)$ (dash-dotted) are plotted together with the energies $\omega_c = T_e x_c$ and $\omega_{\mathrm{exp}} = T_e x_{\mathrm{exp}}$, $x_c(z)$ (solid) and $x_{\mathrm{exp}}(z)$ (dotted).

$x_e < x_{\mathrm{exp,ff}}(z)$ respectively with

$$x_{\mathrm{exp},2\gamma}(z) = 4.3 \times 10^{-10} \frac{z^{7/4}}{(z+z_{\mathrm{eq}})^{1/4}} (\Omega_b h^2)^{1/2} (1 - Y_{He}/2)^{1/2}\,, \qquad (8.41)$$

$$x_{\mathrm{exp,ff}}(z) = 1.1 \times 10^{-2} \frac{z^{1/2}}{(z+z_{\mathrm{eq}})^{1/4}} \Omega_b h^2 (1 - Y_{He}/2)^{1/2}\,. \qquad (8.42)$$

The energies below which photon number changing processes are still faster than expansion and can lead to establishing a thermal equilibrium are given by $x_e < x_{\mathrm{exp}} \equiv \sqrt{x_{\mathrm{exp},2\gamma}^2 + x_{\mathrm{exp,ff}}^2}$.

In Fig. 8.1 we plot z_K, $z_{\mathrm{ff},2\gamma}(x_e)$ as well as $x_c(z)$ and $x_{\mathrm{exp}}(z)$. When $x_c < 0.1$, say if $z < 10^8$, a Planck spectrum is established rapidly only for very small energies, $\omega < 0.1 \times T_e$, while at larger energies we first obtain a Bose–Einstein distribution. Nevertheless, if there is a short period of injection of photons at a redshift where $x_{\mathrm{exp}}(z) \gtrsim 1$, a Planck spectrum will be established eventually for the relevant regime of energies with $x_e \lesssim 1$. We may therefore say, that such processes still are fully thermalized if they happen sufficiently earlier than $z \sim 10^7$. However, if energy injection happens at $z \leq 10^7$, the Planck spectrum cannot be established anymore for energies with $x_e > x_{\mathrm{exp}}(z) \leq 1$.

It is also interesting to note that in the regime where Bremsstrahlung is more efficient than double Compton scattering, $z < z_{\mathrm{ff},2\gamma} \simeq 2 \times 10^6$, these processes are very inefficient anyway and can thermalize the spectrum only for

$x_e < x_{\rm exp}(2\times 10^6) \simeq 0.1$, the bulk part of the spectrum remains Bose–Einstein. Therefore, it is a good approximation to disregard Bremsstrahlung entirely in these considerations.

8.2 A chemical potential

Observational studies of the CMB spectrum have shown that it is very close to a blackbody, i.e. Planck spectrum, see Fig. 1.7. Up to this day, no deviations from a blackbody have been detected. The experimental bound for the reduced chemical potential comes mainly from the FIRAS experiment aboard the COBE satellite. It limits μ to (Fixsen *et al.*, 1996)

$$|\mu| \leq 9\times 10^{-5} \text{ at } 95\% \text{ confidence} . \tag{8.43}$$

This seems very small, but as we shall see, a quite violent event is needed to generate such a chemical potential. This comes from the fact that there are many more photons in the Universe than baryons or dark matter particles. Therefore, producing about one photon per dark matter particle will not induce a big chemical potential.

As we have seen in the previous section, if energy is injected into the CMB at a redshift $z_1 > z_K$ with $x_{\rm exp}(z_1) < 1$, we expect a spectral distortion which leads to a chemical potential on energies with $\omega/T_e \equiv x_e > x_{\rm exp}(z_1)$. In this section we want to estimate the chemical potential produced by a given energy input $\delta\rho$. We assume that the energy input happens rapidly. Furthermore, we neglect the part of the integrals over photon energies in which the spectrum has been able to relax to a Planck spectrum, i.e. $x < x_{\rm exp}(z_1)$. We must therefore assume $z_1 < z_\mu \simeq 6\times 10^6$ (see Fig. 8.1).

This can happen, for example via a 'long' lived, unstable particle which decays at redshift z_1. Some models of super-symmetry predict the existence of a 'next-to-lightest' super-symmetric particle which is rather long lived and decays quite late into the lightest super-symmetric particle which then plays the role of dark matter. But also the annihilation of a very light particle with mass $m \simeq T(1+z_1) \simeq 2.3\times 10^{-4}(1+z_1)$ eV, when the temperature drops below its mass threshold can provoke a chemical potential.

According to what we have learned in the previous section, since $z_1 > z_K$, a Bose–Einstein distribution is established rapidly at some temperature T_e and with chemical potential μ. The temperature T_e and the chemical potential μ are determined by

$$\rho = \frac{1}{\pi^2}T_e^4\int\frac{x_e^3\,dx_e}{e^{x_e+\mu}-1} = \rho_\gamma+\delta\rho = \frac{\pi^2}{15}T^4(1+\epsilon) , \quad \text{and} \tag{8.44}$$

$$n = \frac{1}{\pi^2}T_e^3\int\frac{x_e^2\,dx_e}{e^{x_e+\mu}-1} = n_\gamma+\delta n = \frac{2\zeta(3)}{\pi^2}T^3(1+\alpha) . \tag{8.45}$$

For the last equals sign we have introduced $\epsilon \equiv \delta\rho/\rho_\gamma$ and $\alpha \equiv \delta n/n_\gamma$. For ρ_γ and n_γ we use the expressions for a Planck spectrum given in Eqs. (1.55) and (1.62). For $z_1 \ll 10^7$, photon number changing processes are no longer active and $\alpha \equiv 0$. But here we keep the expressions general. Experimentally we know that $\mu \ll 1$. Expanding the integrals in Eqs. (8.44) and (8.45) to first order in μ yields

$$\rho = \frac{T_e^4}{\pi^2}\,[I_b(3) - 3\mu I_b(2)] = \frac{\pi^2}{15}T^4(1+\epsilon)$$
$$n = \frac{T_e^4}{\pi^2}\,[I_b(2) - 2\mu I_b(1)] = \frac{2\zeta(3)}{\pi^2}T^3(1+\alpha)\,.$$

Here $I_b(n) = \int_0^\infty dx\,\frac{x^n}{e^x-1} = \Gamma(n+1)\zeta(n+1)$ as defined in Eq. (1.56). Inserting $I_b(3) = \pi^4/15$, $I_b(2) = 2\zeta(3)$ and $I_b(1) = \pi^2/6$, we can write these equations as

$$\mu = \frac{6\zeta(3)}{\pi^2}\left[1 - \left(\frac{T}{T_e}\right)^3(1+\alpha)\right] \quad \text{and} \quad 1 - \left(\frac{T}{T_e}\right)^4(1+\epsilon) = \frac{90\zeta(3)}{\pi^4}\mu\,.$$

Since $|\mu| \ll 1$ and $\zeta(3) \simeq 1.2$, we must have $|1 - T/T_e| \ll 1$. We therefore expand $T_e/T = 1+\delta$ with $|\delta| \ll 1$ so that $(T_e/T)^n \simeq 1 + n\delta$. Inserting this approximation above, we can determine δ and μ in terms of ϵ and α. The above relations give (we neglect the second-order terms $\propto \delta\epsilon$ and $\delta\alpha$)

$$\mu = \frac{18\zeta(3)}{\pi^2}(\delta - \alpha/3)$$
$$4\delta - \epsilon = \frac{90\zeta(3)}{\pi^4}\mu\,,$$

so that

$$\delta = \frac{\epsilon - (540\zeta(3)^2/\pi^6)\alpha}{4[1 - 405\zeta(3)^2/\pi^6]} \simeq 0.64\frac{\delta\rho}{\rho} - 0.52\frac{\delta n}{n}\,, \tag{8.46}$$

$$\mu = \frac{3\zeta(3)}{2\pi^2[1 - 405\zeta(3)^2/\pi^6]}(3\epsilon - 4\alpha) \simeq 0.46\left(3\frac{\delta\rho}{\rho_\gamma} - 4\frac{\delta n}{n_\gamma}\right)\,. \tag{8.47}$$

First of all, we note that when photon number changing processes are still very rapid, the photon number will change so that $\delta n/n_\gamma = (3/4)(\delta\rho/\rho_\gamma)$ and no chemical potential is generated. The temperature is then modified by the injection of energy to $T \to T(1+\delta) = T_e = T(1+\epsilon/4)$, which is evident since in this case $\rho/\rho_\gamma = (T_e/T)^4$.

The situation is very different if $z_1 < z_\mu$ and photon number changing processes are no longer active. Then $\delta n = 0$ and

$$\mu \simeq 1.4\frac{\delta\rho}{\rho_\gamma}\,. \tag{8.48}$$

First of all, the chemical potential generated by such an energy injection is always positive. This is good, since a distribution with a negative chemical potential is not well defined for frequencies $\omega \leq \omega_c = -T_e\mu$. But since we know that double Compton scattering is still active at very low frequencies, this is not a real problem, as at these low frequencies a Planck spectrum would be established anyway.

Let us now estimate the chemical potential generated by the decay of a species of non-relativistic particles which contributes an energy density $\Omega_X H_0^2(1+z)^3$ before they decay. Since photons contribute the energy density $\Omega_\gamma H_0^2(1+z)^4$, assuming that a fraction f of the energy of these particles is heating up the CMB we have

$$\mu = 1.4\frac{\delta\rho}{\rho_\gamma} = 1.4\frac{f\Omega_X}{\Omega_\gamma(1+z_1)}\,,$$

where z_1 denotes the redshift of the decay. This formula is of course only valid if $z_1 < z_\mu \sim 10^7$ since an energy injection at higher redshift is still fully thermalized. Using $\Omega_\gamma = 5\times10^{-5}$ (see Appendix A1.3), the limit on the chemical potential can be translated into a limit for $f\Omega_X$,

$$f\Omega_X \leq 3\times10^{-3}\frac{1+z_1}{10^6}\,, \quad z_1 < 10^7\,. \tag{8.49}$$

This might appear as a small number, nevertheless it is more than the entire mass density in stars. If the decay product of the particle species X is supposed to be the dark matter, we need $\Omega_X h^2 > 0.14$. The above bound then implies $f < 0.01 z_1/10^6$, hence only a small fraction of the energy may be injected into standard model particles (other than neutrinos). If a particle decays at a redshift $z_1 \simeq 10^7$ partial thermalization, especially at small frequencies, $x_e \lesssim 1$ can still be achieved. In this case, the Boltzmann equation (8.35) with a source term describing the injection has to be solved numerically and the resulting 'chemical potential' depends on the frequency.

For a particle species which decays into photons when the temperature goes below its mass threshold we would expect $\delta\rho/\rho \sim$0.1–1. So that it would produce a chemical potential of order unity. This shows that no particle with mass $m < T_0(1+z_\mu) \simeq 230\,\mathrm{eV}$ which interacts relatively strongly with photons can exist. Of course such a particle would also be produced in accelerators, so that this does not come as a surprise.

8.3 The Sunyaev–Zel'dovich effect

Clusters of galaxies are permeated by a hot plasma of electrons and nuclei at a temperature of several keV. This is much hotter than CMB photons at redshifts $z \lesssim 1$. Therefore, in the Kompaneets equation (8.15) the first term on the right-hand side dominates.

Furthermore, the plasma is optically thin to Compton scattering so that we can neglect multiple scattering. The change of the photon distribution when passing through a cluster is then simply

$$\delta f = y\omega^{-2}\frac{\partial}{\partial\omega}\left(\omega^4\frac{\partial f}{\partial\omega}\right) , \tag{8.50}$$

where we have introduced the Compton-y parameter

$$y \equiv \sigma_T \int n_e \frac{T_e}{m}\, dr . \tag{8.51}$$

The integral extends through the cluster, and the approximation holds, if the optical depth $\tau = \sigma_T \int n_e\, dr \ll 1$. Inserting a blackbody spectrum, $f \propto (\exp(\omega/T) - 1)^{-1}$, we obtain with $x \equiv \omega/T \neq \omega/T_e$

$$\frac{\delta f}{f} = -y\frac{xe^x}{e^x-1}\left[4 - x\coth\left(\frac{x}{2}\right)\right] \simeq \begin{cases} -2y , & \text{if } x \ll 1 \\ yx^2, & \text{if } x \gg 1 . \end{cases} \tag{8.52}$$

This is the frequency dependent Sunyaev–Zel'dovich (SZ) effect . With the help of

$$\frac{\delta T}{T} = \frac{f}{T}\frac{\delta f}{f}\left(\frac{df}{dT}\right)^{-1} = \frac{e^x-1}{xe^x}\frac{\delta f}{f} ,$$

we can translate it in a frequency dependent temperature shift

$$\frac{\delta T}{T} = -y\left[4 - x\coth\left(\frac{x}{2}\right)\right] \simeq \begin{cases} -2y , & \text{if } x \ll 1 \\ yx, & \text{if } x \gg 1 . \end{cases} \tag{8.53}$$

When passing through a hot thermal plasma, the low-energy Rayleigh–Jeans regime of the photons' spectrum is depleted and the spectrum is enhanced at high energies, in the Wien tail. Photons are on average up-scattered in energy. The spectral change vanishes at $x_0 \simeq 3.8$ given by $0 = 4 - x_0\coth(x_0/2)$, see Fig. 8.2. For a CMB temperature of $T = 2.726$ this corresponds to a frequency of $\nu = 217$ GHz.

Numbers for a typical cluster are $T_e \simeq 7$ keV, $n_e \sim 10^{-2}\,\mathrm{cm}^{-3}$ and the diameter is of the order of $R \simeq 300$ kpc. Estimating the Compton-y parameter by $y \sim y_c = (\sigma_T T_e/m_e)n_e R$ we obtain $y_c \simeq 10^{-4}$. This is the order of magnitude of the SZ effect in clusters. Of course the true value can deviate substantially and will depend on the details of the cluster (LaRoque *et al.*, 2006; De Petris *et al.*, 2002). At present, SZ observers start to use the effect to generate cluster maps of the quantity $T_e n_e$ integrated along the line of sight. Not surprisingly, the best studied cluster so far is the Coma cluster which is closest to us (De Petris *et al.*, 2002). The SZ results for the Coma cluster are shown in Fig. 8.3. In general, the situation is complicated since clusters are complicated objects which may have different 'electron populations' one of them thermal and others not. Also the electron density might be a complicated function of position etc. Here we are not entering into all of

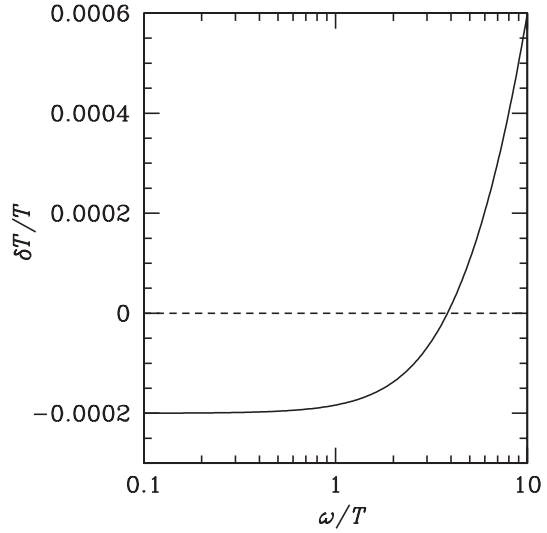


Fig. 8.2. The function $\delta T/T$ given in Eq. (8.53) is shown for $y = 10^{-4}$. Note that it passes through zero roughly at $x = \omega/T \simeq 3.8$.

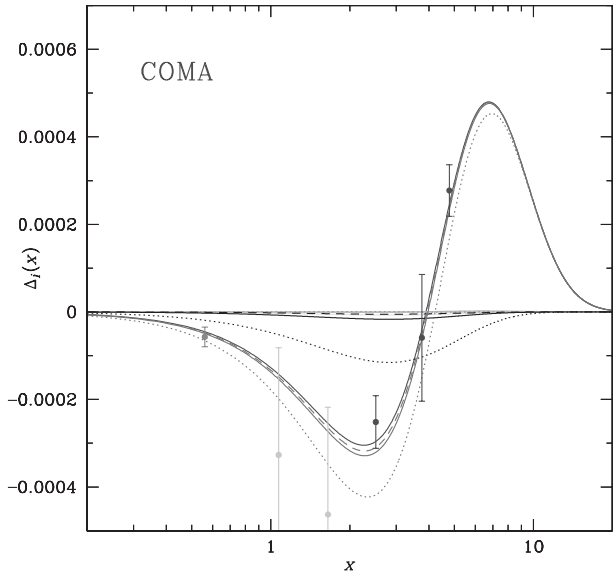


Fig. 8.3. The SZ distortion $\Delta_j(x) = x^3\,\delta f$ measured in the Coma cluster by WMAP (left-most point) OVRA (two subsequent points) and MITO (the three rightmost points) is compared with its theoretical best fit with $T_e = 8.2$ keV and optical depth $\tau_e = 4.9 \times 10^{-3}$ (middle solid curve). Other curves which add possible other effects are also shown. From Colafrancesco (2007).

these interesting difficulties of cluster physics, but we just mention some important points.

- The SZ effect in clusters can add to the CMB anisotropies on very small scales and may, depending on the cluster number density which is not well known, even dominate it above $\ell \sim 2000$. Fortunately it can be distinguished from primordial anisotropies due to its spectral signature, see Fig. 8.2.
- The motion of clusters induces in addition to the thermal SZ effect discussed above a temperature shift due to the bulk motion of the cluster,

$$\delta T/T = v_c \sigma_T \int n_e \, dr = v_c \tau_e \,,$$

  where v_c is the bulk motion in the direction of the line of sight and τ_e is the optical depth of the cluster. This effect which is spectrally identical to primordial CMB anisotropies is typically several times smaller than the thermal SZ effect. It goes under the name 'kinetic SZ effect'.
- The average effect from all clusters should contribute a mean Compton-y parameter in the Universe. Its amplitude strongly depends on the cluster distribution, but is estimated to be of the order of $\bar{y} \sim 10^{-7}$. This number is within reach of planned CMB spectrum experiments (Singal *et al.*, 2005; Kogut *et al.*, 2006).
- Clearly the SZ effect in clusters is mainly of interest for cluster physics. Together with X-ray observations which probe the square of the electron density, it allows us, in principle, to gain detailed information about the electron density and temperature distribution inside the clusters. Furthermore, the SZ effect, which represents the 'shadow' of the cluster in the CMB, is independent of redshift and might lead to the detection of clusters with high redshift which are too faint to be seen in optical or X-ray telescopes.

The fact that the observed average Compton-y parameter is so small $y \leq 10^{-5}$, leads to a limit on early reionization. Let us derive this limit for a simple toy model. We assume that the Universe is reionized at some redshift $z_{\rm ri}$. Then during the ionization process, the electrons also gain some kinetic energy which we estimate to be typically in the 10 eV range. This seems reasonable, if we do not want to assume that the reionizing photons have exactly the reionization energy, 13.6 eV, but some energy in this ball park. The remaining energy is then simply absorbed by the electron as kinetic energy. As the Universe evolves, since the electron momenta are redshifted $p \propto 1/a$, the temperature which is a measure of the kinetic energy of the electrons is also redshifted, $T_e \simeq p^2/2m \propto a^{-2}$. Already in Chapter 1 in our discussion below Eq. (1.93) we have seen that the temperature of non-relativistic particles decays like $1/a^2$. Denoting the electron temperature at reionization by $T_{\rm ri}$ we find that reionization should induce a Compton-y parameter given by

$$y = \frac{\sigma_T T_{\rm ri} n_e(t_0)}{m_e (z_{\rm ri}+1)^2} \int_{t_{\rm ri}}^{t_0} dt \, (z+1)^5 . \tag{8.54}$$

We expect $z_{\rm ri}$ to lie in the matter dominated phase of expansion and before the cosmological constant becomes relevant. During matter domination the Friedmann equation yields

$$dt = \frac{-dz}{(1+z)^{5/2}} \frac{1}{H_0\sqrt{\Omega_m}} ,$$

so that we obtain

$$y = \frac{\sigma_T T_{\rm ri} n_e(t_0)}{m_e H_0 \sqrt{\Omega_m}(z_{\rm ri}+1)^2} \int_0^{z_{\rm ri}} dz\,(z+1)^{5/2} = \frac{2(1+z_{\rm ri})^{3/2}\sigma_T T_{\rm ri} n_e(t_0)}{7 m_e H_0 \sqrt{\Omega_m}}$$
$$\simeq 5\times 10^{-7} \frac{\Omega_b h^2}{\sqrt{\Omega_m h^2}} (1+z_{\rm ri})^{3/2} \left(\frac{T_{\rm ri}}{10\,{\rm eV}}\right) . \tag{8.55}$$

For $\Omega_m h^2 = 0.13$ and $\Omega_b h^2 \cong 0.022$ the limit $y < 10^{-5}$ translates into the reionization redshift

$$z_{\rm ri} < 50 \left(\frac{T_{\rm ri}}{10\,{\rm eV}}\right)^{2/3} . \tag{8.56}$$

This is a truly interesting number and it reduces by a factor $(10^5 y)^{2/3}$ if we lower the limit on the y parameter. Already a y parameter $y < 10^{-6}$ would require a reionization redshift of $z_{\rm ri} < 10 \times (T_{\rm ri}/10\ {\rm eV})^{2/3}$. As discussed in Chapter 6 the CMB polarization spectrum favours $z_{\rm ri} \simeq 10$. Reducing the assumed electron temperature by a factor of 10 can help, but when $y < 10^{-7}$ we have no simple way out. Therefore, according to our understanding of the reionization process which took place probably at $z_{\rm ri} \sim 10$, this should have led to a Compton-y parameter of the order of $y \simeq 10^{-7}$–10^{-6}.

Exercises

Ex. 8.1 The Hubble parameter in a matter/radiation universe and the collision times
Using the Friedmann equation show that in a spatially flat universe containing only matter and radiation, the Hubble parameter is given by

$$H^2(z) = H_0^2 \Omega_r (1+z)^3 (z + z_{\rm eq} + 2) , \tag{8.57}$$
$$H \simeq 2\times 10^{-20}\,{\rm s}^{-1} (2 + z + z_{\rm eq})^{1/2} (1+z)^{3/2} , \tag{8.58}$$

where $1 + z_{\rm eq} = \Omega_m/\Omega_r \simeq 2.4\times 10^4 \Omega_m h^2$ is the redshift where the matter and radiation densities are equal. We use the relativistic density parameter for photons and three species of massive neutrinos, $\Omega_r = 4.19\times 10^{-5}\, h^{-2}$ and $H_0 = 3.24 \times 10^{-18}\, h\,{\rm s}^{-1}$ given in Appendices A1.2 and A1.3.
Using $n_e = (1 - Y_{He}/2) n_B$ also verify Eqs. (8.17), (8.30) and (8.34).

Ex. 8.2 Aberration
For a given non-relativistic electron velocity $\mathbf{v}$ and incoming and outgoing photon directions $\mathbf{n}$ and $\mathbf{n}'$ in the laboratory frame determine the scalar product $\mathbf{n}_R \cdot \mathbf{n}'_R$ of the photon directions in the electron rest frame.

Final remarks

The goal of this book has been to give you an overview of one of the most successful and fascinating topics of cosmology, the physics of the cosmic microwave background. Its success is best illustrated by the two Nobel prizes the subject has led to: Penzias & Wilson (1978) for the discovery of the CMB and Mather & Smoot (2006) for the detailed measurement of its spectrum and for the discovery of the fluctuations.

I have concentrated on the theoretical side of the topic not only because this is my expertise, but also because I believe that this subject is mature enough for a text book. On the experimental side, certainly there is much to tell and tremendous progress has been made in the last 15 years, but I think, hope, that this is not the end of it. There will be much more to come and therefore a book on CMB experiments could be only a snapshot of the present situation. The theory of the CMB, on the other hand, is in many of its aspects basically complete, so that I can hope that this book may have some lasting value for students who want to learn about the topic and also for researchers in the field who want to obtain a rather detailed overview.

I am afraid that despite a big effort there are still some misprints or errors in the book. If you, dear reader, have spotted one, please let me know (ruth.durrer@physics.unge.ch) so that I can correct it in forthcoming editions.

Appendix 1

Fundamental constants, units and relations

Here we summarize some useful relations between units and the values of physical constants that are used throughout this book.

A1.1 Conversion factors, units

In a system of units where $\hbar = c = k_{\text{Boltzmann}} = 1$, as is often used in this book, all units can be expressed in terms of a unit of energy like, e.g., the GeV or a length scale like, e.g., cm. We then have

$$\begin{aligned} 1\,\text{GeV} &= 1.6022 \times 10^{-3}\,\text{erg} \\ &= 1.1605 \times 10^{13}\,\text{K} \\ &= 1.7827 \times 10^{-24}\,\text{g} \\ &= 5.0684 \times 10^{13}\,\text{cm}^{-1} \\ &= 1.5192 \times 10^{24}\,\text{s}^{-1} \end{aligned}$$

The relation I always remember by heart for order of magnitude estimates is $1 =$ 200 MeV fm. Here 'fm' is 1 femtometre (or fermi), $1\,\text{fm} = 10^{-15}\,\text{m}$.

Other useful relations are

$$\begin{aligned} 1\ \text{parsec (pc)} &= 3.2612\ \text{light years} \ = 3.0856 \times 10^{18}\,\text{cm} \\ 1\ \text{Mpc} &= 10^{6}\,\text{pc} \simeq 3 \times 10^{24}\text{cm} \simeq 10^{14}\,\text{s} \\ 1\,\text{g cm}^{-3} &= 4.3102 \times 10^{-18}\,\text{GeV}^4 \\ \text{(Astronomical unit)}\ \ 1\,\text{AU} &= 1.4960 \times 10^{13}\,\text{cm} \\ 1\,(\text{Gauss})^2/8\pi &= 1.9084 \times 10^{-40}\,\text{GeV}^4 \\ \text{(Jansky)}\ \ 1\,\text{Jy} &= 10^{-23}\,\text{erg cm}^{-2}\,\text{s}^{-1}\,\text{Hz}^{-1} \\ &= 2.4730 \times 10^{-48}\,\text{GeV}^3 \\ 1\,\text{yr} &\simeq \pi \times 10^{7}\,\text{s} \\ \text{(Radian)}\ \ 1\,\text{rad} &= (180/\pi)\ \text{degrees} = 57.266\ \text{degrees} \\ \text{(Steradian)}\ \ 1\,\text{sr} &= 1\,\text{rad}^2 \ = \ 3.283 \times 10^{3}\ \text{degrees}^2 \end{aligned}$$

A1.2 Constants

A1.2.1 Fundamental constants

$$
\begin{aligned}
\text{Planck's constant} \quad & \hbar = 1 \ = h/(2\pi) \\
& = 1.0546 \times 10^{-27}\,\text{cm}^2\,\text{g}\,\text{s}^{-1} \\
\text{Speed of light} \quad & c = 1 \ = \ 2.9979 \times 10^{10}\,\text{cm}\,\text{s}^{-1} \\
\text{Fine structure constant} \quad & \alpha \equiv \ \frac{e^2}{4\pi} = 1/137.036 \\
\text{Gravitational constant} \quad & G = \ 6.673 \times 10^{-8}\,\text{cm}^3\text{g}^{-1}\text{s}^{-2} \\
\text{Planck mass} \quad & m_P = \ 1.2211 \times 10^{19}\,\text{GeV} \\
& = \ 2.1768 \times 10^{-5}\,\text{g} \\
\text{Planck length} \quad & \ell_P = \ 8.189 \times 10^{-20}\,\text{GeV}^{-1} \\
& = \ 1.616 \times 10^{-33}\,\text{cm} \\
\text{Planck time} \quad & \tau_P = \ 8.189 \times 10^{-20}\,\text{GeV}^{-1} \\
& = \ 5.3904 \times 10^{-44}\,\text{s} \\
\text{Electron mass} \quad & m_e = \ 0.5110\,\text{MeV} \\
\text{Proton mass} \quad & m_p = \ 938.27\,\text{MeV} \\
\text{Neutron mass} \quad & m_n = \ 939.57\,\text{MeV} \\
\text{Rydberg} \quad & 1\,\text{Ry} = \alpha^2 m_e/2 = 13.606\,\text{eV} \\
\text{Thomson cross section} \quad & \sigma_T \equiv \ 8\pi\alpha^2/3m_e^2 = 6.65246 \times 10^{-25}\,\text{cm}^2 \\
\text{Bohr radius} \quad & a_0 \equiv \ \frac{1}{\alpha m_e} = 5.2918 \times 10^{-9}\,\text{cm} \\
\text{Bohr magneton} \quad & \mu_0 \equiv \ \frac{e}{2m_e} = 5.7884 \times 10^{-18}\,\frac{\text{GeV}}{\text{Gauss}} \\
\text{Avogadro's number} \quad & N_A = \ 6.022 \times 10^{23} \\
\text{Stefan–Boltzmann constant} \quad & a_{SB} \equiv \pi^2/15 = 0.658 \\
& = 7.566 \times 10^{-15}\,\text{erg}\,\text{cm}^{-3}\,\text{K}^{-4}\,.
\end{aligned}
$$

A1.2.2 Important constants

$$
\begin{aligned}
\text{Solar mass} \quad & M_\odot = \ 1.989 \times 10^{33}\,\text{g} = 1.116 \times 10^{57}\,\text{GeV} \\
\text{Solar radius} \quad & R_\odot = \ 6.9598 \times 10^{10}\,\text{cm} = 3.527 \times 10^{24}\,\text{GeV}^{-1} \\
\text{Luminosity of the Sun} \quad & L_\odot = \ 3.90 \times 10^{33}\text{erg}\,\text{s}^{-1} = 1.6 \times 10^{12}\,\text{GeV}^{-2} \\
\text{Mass of the Earth} \quad & M_\oplus = \ 5.977 \times 10^{27}\,\text{g} \\
& = \ 3.357 \times 10^{51}\,\text{GeV} \\
\text{Solar magnitude} \quad & m_\odot = \ -26.85, \quad \text{(apparent)} \\
& \mathcal{M}_\odot = \ 4.72, \quad \text{(absolute)} \\
\text{Distance modulus} \quad & m - \mathcal{M} = \ 5\log(D/10\,\text{pc}) \\
\text{Hubble constant} \quad & H_0 = \ 100\,h\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1} \\
& = \ 2.1332\,h \times 10^{-42}\text{GeV} \\
\text{where} \quad & 0.5 < h < 0.8
\end{aligned}
$$

Hubble time, distance	$H_0^{-1} =$	$3.0856 \times 10^{17}\, h^{-1}$ s
	$=$	$9.7776 \times 10^{9}\, h^{-1}$ yr
	$=$	$2997.9\, h^{-1}$ Mpc
	$=$	$9.2503 \times 10^{27} h^{-1}$ cm
Critical density $\rho_c = 3H_0^2/8\pi G$	$=$	$1.8791\, h^2 \times 10^{-29}$ g cm^{-3}
	$=$	$8.0992\, h^2 \times 10^{-47}$ GeV^4
	$=$	$1.0540\, h^2 \times 10^{4}$ eV cm^{-3}
	$=$	$11.2\, h^2$ (proton masses)/m^3
CMB temperature	$T_0 =$	2.725 K
	$=$	2.35×10^{-13} GeV
Neutrino temperature	$T_\nu =$	1.945 K $= (4/11)^{1/3} T_0$

A1.3 Useful relations

Photons

Number density	$n_\gamma =$	411 cm^{-3}
Entropy density	$s_\gamma =$	2900 cm^{-3} $= 3.602 n_\gamma$
Energy density	$\rho_\gamma =$	2.01×10^{-51} GeV^4
Density parameter	$\Omega_\gamma h^2 =$	2.48×10^{-5}

Neutrino (per species)

Number density	$n_\nu =$	112 cm^{-3}
Entropy density	$s_\nu =$	470 cm^{-3} $= 4.202 n_\nu$
Energy density	$\rho_\nu =$	3.08×10^{-53} GeV^4
Density parameter	$\Omega_\nu h^2 =$	5.63×10^{-6}

Relativistic entropy	$s_0 =$	4310 cm$^{-3} = s_\gamma + 3 s_\nu$
Relativistic density parameter	$\Omega_{\gamma 3\nu} h^2 =$	4.17×10^{-5}
Baryon density	$\Omega_B h^2 =$	$3.639 \times 10^{7} \eta_B$
	$=$	0.022 ± 0.002
Baryons per photon $n_B/n_\gamma =$	$\eta_B =$	$(5.2 \pm 0.5) \times 10^{-10}$
Age of the Universe		
for $T > T_{\rm eq}$:	$\tau =$	$2.42\,\text{s} \times (1\,\text{MeV}/T)^2/\sqrt{g_*}$
	$=$	$0.30118 (m_P/T^2)/\sqrt{g_*}$
for $T < T_{\rm eq}$, $\Lambda = 0$:	$\tau =$	$\dfrac{2.057 \times 10^{17}}{\sqrt{\Omega_m h^2}} (1+z)^{-3/2}$ s
	$=$	$\dfrac{7.504 \times 10^{11}}{\sqrt{\Omega_m h^2}} (T/1\,\text{eV})^{-3/2}$ s

Matter density	$\Omega_m h^2 =$	0.13 ± 0.02
Equivalence redshift	$z_{\rm eq} =$	$2.4 \times 10^4 (\Omega_m h^2)$
Equivalence temperature	$T_{\rm eq} =$	$5.6\ {\rm eV} (\Omega_m h^2)$
Decoupling redshift	$z_{\rm dec} \simeq$	1089 ± 2
Decoupling temperature	$T_{\rm dec} \simeq$	$2970 \pm 10\ {\rm K} = 0.26\ {\rm eV}$
Decoupling time	$\tau_{\rm dec} \simeq$	$10^{13} (0.14/\Omega_m h^2)^{1/2}\ {\rm s}$
Recombination redshift	$z_{\rm rec} \simeq$	1360
Nucleosynthesis temperature	$T_{\rm nuc} \simeq$	$0.08\ {\rm MeV} = 9 \times 10^8\ {\rm K}$
Time of nucleosynthesis	$\tau_{\rm nuc} \simeq$	$206\ {\rm s}$
Age of the Universe	$\tau_0 =$	$(1.37 \pm 0.03) \times 10^{10}\ {\rm yr}$

Appendix 2

General relativity

Throughout this book it is assumed that the reader is familiar with the basics of general relativity as presented, e.g., in Wald (1984). This appendix does not present an introduction to general relativity but just fixes the notation used throughout this book. Furthermore, we calculate the curvature tensor for a FL universe.

A2.1 Notation

We consider a four-dimensional pseudo-Riemannian spacetime given by a manifold $\mathcal{M}$ and a metric g with signature $(-,+,+,+)$. For a given choice of coordinates $(x^\mu)_{\mu=0}^3$ the metric is given by the ten components of a 4×4 symmetric tensor,

$$g = ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu\,. \tag{A2.1}$$

Contra- and covariant tensor fields on a pseudo-Riemannian manifold are equivalent. Their indices can be lowered and raised with the metric, e.g.,

$$g_{\beta\nu}T^{\alpha\nu} = T^\alpha_\beta = g^{\alpha\mu}T_{\mu\beta}\,. \tag{A2.2}$$

Here $g^{\alpha\mu}$ is the inverse of the metric such that $g^{\alpha\mu}g_{\mu\beta} = \delta^\alpha_\beta$, and we adopt Einstein's summation convention: indices which appear as subscripts and superscripts are summed over.

The Christoffel symbols are defined by

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2}g^{\mu\nu}\left[\partial_\alpha g_{\nu\beta} + \partial_\beta g_{\nu\alpha} - \partial_\nu g_{\alpha\beta}\right]. \tag{A2.3}$$

Here ∂_μ indicates a partial derivative w.r.t. the coordinate x^μ, this is sometimes also simply denoted by a comma $\partial_\mu f \equiv f_{,\mu}$. Covariant derivatives are indicated by a semi-colon, or by the symbol ∇.

A geodesic $\gamma(t)$ with $X = \dot\gamma$ is a solution to the differential equation

$$\nabla_X X = 0\,, \quad X^\mu\partial_\mu X^\nu + \Gamma^\nu_{\alpha\beta}X^\alpha X^\beta = \frac{d^2X^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta}\frac{dX^\alpha}{ds}\frac{dX^\beta}{ds}0\,, \tag{A2.4}$$

where the second equation expresses the first equation in components. The vector field $X = \dot\gamma$ is given by $X = X^\mu\partial_\mu = \frac{dX^4}{ds}$. We often conveniently identify a vector field with the partial derivative in its direction. A tensor field T of rank (p,q) is parallel transported along the vector field X if

$$\nabla_X T = 0\,, \quad X^\mu T^{\alpha_{i_1}\cdots\alpha_{i_p}}_{\beta_{j_1}\cdots\beta_{j_q};\mu} = 0\,. \tag{A2.5}$$

Covariant derivatives of a tensor field are given by

$$T^{\alpha_{i_1}\cdots\alpha_{i_p}}_{\beta_{j_1}\cdots\beta_{j_q};\mu} = T^{\alpha_{i_1}\cdots\alpha_{i_p}}_{\beta_{j_1}\cdots\beta_{j_q},\mu} + \Gamma^{\alpha_{i_1}}_{\mu\sigma} T^{\sigma\cdots\alpha_{i_p}}_{\beta_{j_1}\cdots\beta_{j_q}} + \cdots - \Gamma^{\sigma}_{\mu\beta_{j_1}} T^{\alpha_{i_1}\cdots\alpha_{i_p}}_{\sigma\cdots\beta_{j_q}} - \cdots \ . \tag{A2.6}$$

The Riemann curvature tensor is defined by

$$R^{\alpha}_{\beta\mu\nu} = \Gamma^{\alpha}_{\nu\beta,\mu} - \Gamma^{\alpha}_{\mu\beta,\nu} + \Gamma^{\rho}_{\beta\nu}\Gamma^{\alpha}_{\mu\rho} - \Gamma^{\rho}_{\beta\mu}\Gamma^{\alpha}_{\nu\rho} \ . \tag{A2.7}$$

The tensor $R_{\alpha\beta\mu\nu} = g_{\alpha\sigma} R^{\sigma}_{\beta\mu\nu}$ is anti-symmetric in the first $(\alpha\beta)$ and second $(\mu\nu)$ pair of indices and symmetric in the exchange of the pairs, $(\alpha\beta) \leftrightarrow (\mu\nu)$. The Bianchi identities read

$$\Sigma_{(\beta\mu\nu)} R^{\alpha}_{\beta\mu\nu} = 0 \qquad \text{1st Bianchi identity.} \tag{A2.8}$$

$$\Sigma_{(\mu\nu\sigma)} R^{\alpha}_{\beta\mu\nu;\sigma} = 0 \qquad \text{2nd Bianchi identity .} \tag{A2.9}$$

Here $\Sigma_{(\beta\mu\nu)}$ denotes the sum over all cyclic permutations of these three indices.

The Ricci tensor and the Riemann scalar are given by

$$R_{\mu\nu} = R^{\alpha}_{\mu\alpha\nu} \ , \qquad R = R^{\mu}_{\mu} = R_{\mu\nu} g^{\mu\nu} \ . \tag{A2.10}$$

With these sign conventions, the curvature of the sphere is positive, and changing the order of covariant derivatives of a vector field X yields

$$\nabla_\mu \nabla_\nu X^\alpha - \nabla_\nu \nabla_\mu X^\alpha = R^{\alpha}_{\sigma\mu\nu} X^\sigma \ . \tag{A2.11}$$

The Einstein tensor is defined as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \ . \tag{A2.12}$$

The second Bianchi identity and the symmetries of the Riemann tensor imply $G^{\nu}_{\mu;\nu} = 0$.

The field equations of general relativity relate the curvature to the energy–momentum tensor $T_{\mu\nu}$ via Einstein's equation,

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \ , \tag{A2.13}$$

where G denotes Newton's constant, $G = m_P^{-2}$. The second Bianchi identity ensures that $T_{\mu\nu}$ is covariantly conserved, $T^{\nu}_{\mu;\nu} = 0$. Equation (A2.13) can also be derived from an action principle with

$$S = S_{\rm grav} + S_{\rm mat} \ .$$

Here $S_{\rm mat}$ is the usual matter action and

$$S_{\rm grav} = \frac{m_P^2}{16\pi} \int d^4x \, \sqrt{-g} R \tag{A2.14}$$

is the Hilbert action. A somewhat tedious but standard calculation gives (see e.g., Wald, 1984)

$$\delta S_{\rm grav} = -\frac{m_P^2}{16\pi} \int d^4x \, \sqrt{-g} G^{\mu\nu} \delta g_{\mu\nu} \ . \tag{A2.15}$$

The Einstein equation implies then that the energy–momentum tensor can be obtained by varying the matter action w.r.t. the metric,

$$T^{\mu\nu} = 2\sqrt{-g} \frac{\delta S_{\rm mat}}{\delta g_{\mu\nu}} \ .$$

By construction, this energy–momentum tensor is always symmetric, but it does, in general, not agree with the canonical energy–momentum tensor. Of course the conserved quantities (if any!) are the same for both definitions.

The Weyl tensor specifies the degrees of freedom of the Riemann tensor which are not determined by the Ricci tensor (or Einstein tensor). It is the traceless part of $R^{\alpha}_{\beta\mu\nu}$. In n dimensions, $n \geq 3$, it is given by

$$C_{\alpha\beta\mu\nu} = R_{\alpha\beta\mu\nu} - \frac{2}{n-2}\left(g_{\alpha[\mu}R_{\nu]\beta} + g_{\beta[\mu}R_{\nu]\alpha}\right) - \frac{2}{(n-1)(n-2)} R g_{\alpha[\mu}g_{\nu]\beta}\,. \tag{A2.16}$$

Here $[\mu\nu]$ denotes anti-symmetrization in the indices μ and ν. The Weyl tensor has the same symmetries like the Riemann tensor but all its traces vanish. It describes the degrees of freedom of the curvature (gravitational field) in source-free spacetime, hence it describes gravity waves.

An introduction to general relativity can be found e.g., in the books by Straumann (2004) or Wald (1984).

A2.2 The Lie derivative

For a vector field X with flux ϕ_t^X the Lie derivative of a tensor field T of arbitrary rank is defined by

$$L_X T = \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left(\left(\phi_\epsilon^X\right)^* T - T\right)\,. \tag{A2.17}$$

Here $\left(\phi_\epsilon^X\right)^*$ denotes the pullback of the map $\phi_t^X : \mathcal{M} \to \mathcal{M} : p \mapsto \gamma_p(t)$, where γ_p is the integral curve to X with starting point p. The existence and uniqueness of solutions to ordinary differential equations tells us that for sufficiently small t, ϕ_t^X is a local diffeomorphism. If $T(t)$ denotes the value of the tensor field T at the position $\gamma_p(t)$ we also have

$$L_X T(p) = \left.\frac{d}{dt}\right|_{t=0} T(t)\,. \tag{A2.18}$$

Hence the Lie derivative in direction X vanishes if the tensor field T is conserved along integral curves of X. Furthermore, for small t we have

$$\left(\phi_t^X\right)^* T = T + t L_X T + \mathcal{O}(t^2). \tag{A2.19}$$

In coordinates the Lie derivative becomes (see e.g., Wald, 1984)

$$L_X T^{\alpha_{i_1}\cdots\alpha_{i_p}}_{\beta_{j_1}\cdots\beta_{j_q}} = X^\mu T^{\alpha_{i_1}\cdots\alpha_{i_p}}_{\beta_{j_1}\cdots\beta_{j_q},\mu} - X^{\alpha_{i_1}}{}_{,\sigma}\, T^{\sigma\cdots\alpha_{i_p}}_{\beta_{j_1}\cdots\beta_{j_q}} - \cdots + X^\sigma{}_{,\beta_{j_1}}\, T^{\alpha_{i_1}\cdots\alpha_{i_p}}_{\sigma\cdots\beta_{j_q}} + \cdots\,. \tag{A2.20}$$

A2.3 Friedmann metric and curvature

The Friedmann metric is given by

$$ds^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu = -d\tau^2 + a^2(\tau)\gamma_{ij}\,dx^i\,dx^j = a^2(\tau)[-dt^2 + \gamma_{ij}\,dx^i\,dx^j]. \tag{A2.21}$$

The Christoffel symbols with respect to cosmic or conformal time are

		cosmic time τ	conformal time t	
Γ^0_{00}	$=$	0	$\frac{\dot{a}}{a}$,	(A2.22)
Γ^i_{00}	$=$	0	0 ,	(A2.23)
Γ^0_{i0}	$=$	0	0 ,	(A2.24)
Γ^i_{j0}	$=$	$\frac{a'}{a}\delta^i_j = H\delta^i_j$	$\frac{\dot{a}}{a}\delta^i_j = \mathcal{H}\delta^i_j$,	(A2.25)
Γ^0_{ij}	$=$	$a' a \gamma_{ij}$	$\frac{\dot{a}}{a}\gamma_{ij}$,	(A2.26)
Γ^k_{ij}	$= {}^{(3)}\Gamma^k_{ij} = \frac{1}{2}\gamma^{km}\left(\gamma_{im,j} + \gamma_{jm,i} - \gamma_{ij,m}\right)$		${}^{(3)}\Gamma^k_{ij}$,	(A2.27)

where ${}^{(3)}\Gamma^k_{ij}$ denotes the three-dimensional Christoffel symbols of the metric γ which depend on the coordinate system chosen on the spatial slices. The over-dot indicates a derivative w.r.t. conformal time t while the prime indicates a derivative w.r.t. cosmic time τ.

The non-vanishing components of the Riemann and Ricci curvature tensors in **cosmic time** τ are then given by

$$R^0_{i0j} = a'' a \gamma_{ij} \,, \tag{A2.28}$$

$$R^i_{00j} = \frac{a''}{a}\delta^i_j \,, \tag{A2.29}$$

$$R^i_{jkm} = {}^{(3)}R^i_{jkm} + (a')^2\left(\delta^i_k\gamma_{jm} - \delta^i_m\gamma_{jk}\right) \,, \tag{A2.30}$$

$$R_{00} = -3\frac{a''}{a} \,, \tag{A2.31}$$

$$R_{ij} = \left[a'' a + 2\left(a'^2 + K\right)\right]\gamma_{ij} \,, \tag{A2.32}$$

$$R = 6\left[\frac{a''}{a} + H^2 + \frac{K}{a^2}\right] \,, \tag{A2.33}$$

while in **conformal time** t we have

$$R^0_{i0j} = \left(\frac{\dot{a}}{a}\right)^{\cdot}\gamma_{ij} = \dot{\mathcal{H}}\gamma_{ij} \,, \tag{A2.34}$$

$$R^i_{00j} = \left(\frac{\dot{a}}{a}\right)^{\cdot}\delta^i_j = \dot{\mathcal{H}}\delta^i_j \,, \tag{A2.35}$$

$$R^i_{jkm} = {}^{(3)}R^i_{jkm} + \mathcal{H}^2\left(\delta^i_k\gamma_{jm} - \delta^i_m\gamma_{jk}\right) \,, \tag{A2.36}$$

$$R_{00} = -3\left(\frac{\dot{a}}{a}\right)^{\cdot} = -3\dot{\mathcal{H}} \,, \tag{A2.37}$$

$$R_{ij} = \left[\dot{\mathcal{H}} + 2\left(\mathcal{H}^2 + K\right)\right]\gamma_{ij} \,, \tag{A2.38}$$

$$R = \frac{6}{a^2}\left[\dot{\mathcal{H}} + \mathcal{H}^2 + K\right] \,. \tag{A2.39}$$

The curvature on the three-dimensional slices of constant time is given by

$$ {}^{(3)}R^{i}_{jkm} = K\left(\delta^{i}_{k}\gamma_{jm} - \delta^{i}_{m}\gamma_{jk}\right) , \tag{A2.40} $$

$$ {}^{(3)}R_{ij} = 2K\gamma_{ij} \qquad \text{and} \tag{A2.41} $$

$$ {}^{(3)}R = 6K . \tag{A2.42} $$

Appendix 3

Perturbations

In this appendix we present the intermediate results in the calculation of the perturbed Einstein equations for a given 'Fourier mode' k. We also determine the Weyl tensor. All the results are for conformal time t.

A3.1 Scalar perturbations

We work in the longitudinal gauge,

$$ds^2 = a^2\left(-(1+2\Psi Q^{(S)})\,dt^2 + (1-2\Phi Q^{(S)})\gamma_{ij}\,dx^i\,dx^i\right)\,. \tag{A3.1}$$

Here $Q^{(S)}$ is an eigenfunction of the spatial Laplacian with eigenvalue $-k^2$ (see Section 2.2.2).

A3.1.1 The Christoffel symbols

$$\delta\Gamma^0_{00} = \dot\Psi Q^{(S)}\,, \qquad \delta\Gamma^0_{0j} = -k\Psi Q^{(S)}_j\,, \tag{A3.2}$$

$$\delta\Gamma^j_{00} = -k\Psi Q^{(S)j}\,, \quad \delta\Gamma^j_{i0} = -\dot\Phi\delta^j_i Q^{(S)}\,, \tag{A3.3}$$

$$\delta\Gamma^0_{ij} = \left[-2\mathcal{H}(\Psi+\Phi) - \dot\Phi\right] Q^{(S)}\gamma_{ij}\,, \tag{A3.4}$$

$$\delta\Gamma^j_{im} = k\Phi\left[\delta^j_i Q^{(S)}_m + \delta^j_m Q^{(S)}_i - \gamma_{im}Q^{(S)j}\right]\,. \tag{A3.5}$$

A3.1.2 The Riemann tensor

$$\delta R^0_{00j} = \delta R^0_{0ij} = 0, \tag{A3.6}$$

$$\delta R^0_{i0j} = -\left[2\dot{\mathcal{H}}(\Psi+\Phi) + \mathcal{H}(\dot\Psi+\dot\Phi) + \ddot\Phi - \frac{k^2}{3}\Psi\right]\gamma_{ij}Q^{(S)} - k^2\Psi Q^{(S)}_{ij}\,, \tag{A3.7}$$

$$\delta R^0_{ijm} = -k\left[\mathcal{H}\Psi + \dot\Phi\right]\left(\gamma_{ij}Q^{(S)}_m - \gamma_{im}Q^{(S)}_j\right)\,, \tag{A3.8}$$

$$\delta R^i_{00j} = \left[\frac{k^2}{3}\Psi - \mathcal{H}(\dot\Psi + \dot\Phi) - \ddot\Phi\right]\delta^i_j Q^{(S)} - k^2\Psi Q^{(S)i}_j \,, \tag{A3.9}$$

$$\delta R^i_{0jm} = -k\left[\dot\Phi + \mathcal{H}\Psi\right]\left(\delta^i_j Q^{(S)}_m - \delta^i_m Q^{(S)}_j\right) \,, \tag{A3.10}$$

$$\delta R^i_{j0m} = k\left[\mathcal{H}\Psi + \dot\Phi\right]\left(\delta^i_m Q^{(S)}_j - \gamma_{jm} Q^{(S)i}\right) \,, \tag{A3.11}$$

$$\begin{aligned}\delta R^i_{jmn} = &-2\left[\mathcal{H}^2(\Psi+\Phi) + \mathcal{H}\dot\Phi + \frac{1}{3}k^2\Phi\right]\left(\delta^i_m\gamma_{jn} - \delta^i_n\gamma_{jm}\right)Q^{(S)} \\ &- k^2\Phi\left(\delta^i_n Q^{(S)}_{jm} - \delta^i_m Q^{(S)}_{jn} + Q^{(S)i}_n\gamma_{jm} - Q^{(S)i}_m\gamma_{jn}\right) \,.\end{aligned} \tag{A3.12}$$

A3.1.3 The Ricci and Einstein tensors

The perturbation if the Ricci tensor is

$$\delta R_{00} = \left[3\mathcal{H}(\dot\Psi + \dot\Phi) - k^2\Psi + 3\ddot\Phi\right]Q^{(S)} \,, \tag{A3.13}$$

$$\delta R_{0j} = -2k\left[\mathcal{H}\Psi + \dot\Phi\right]Q^{(S)}_j \,, \tag{A3.14}$$

$$\begin{aligned}\delta R_{ij} = &\left[-2(\dot{\mathcal{H}} + 2\mathcal{H}^2)(\Psi+\Phi) - \mathcal{H}\dot\Psi + \frac{k^2}{3}\Psi - \ddot\Phi - 5\mathcal{H}\dot\Phi\right. \\ &\left. - \frac{4}{3}k^2\Phi\right]\gamma_{ij}Q^{(S)} + k^2(\Phi-\Psi)Q^{(S)}_{ij} \,.\end{aligned} \tag{A3.15}$$

The perturbation of the Riemann scalar then becomes

$$\delta R = -\frac{2}{a^2}\left[6(\dot{\mathcal{H}} + \mathcal{H}^2)\Psi + 3\mathcal{H}\dot\Psi - k^2\Psi + 9\mathcal{H}\dot\Phi + 3\ddot\Phi + 2(k^2 - 3K)\Phi\right]Q^{(S)} \,. \tag{A3.16}$$

For the Einstein tensor we find

$$\delta G^0_0 = \frac{2}{a^2}\left[3\mathcal{H}^2\Psi + 3\mathcal{H}\dot\Phi + (k^2 - 3K)\Phi\right]Q^{(S)} \,, \tag{A3.17}$$

$$\delta G^0_j = \frac{2}{a^2}k\left[\mathcal{H}\Psi + \dot\Phi\right]Q^{(S)}_j \,, \tag{A3.18}$$

$$\delta G^j_0 = -\frac{2}{a^2}k\left[\mathcal{H}\Psi + \dot\Phi\right]Q^{(S)j} \,, \tag{A3.19}$$

$$\begin{aligned}\delta G^i_j = &\frac{2}{a^2}\left[(2\dot{\mathcal{H}} + \mathcal{H}^2)\Psi + \mathcal{H}\dot\Psi - \frac{k^2}{3}\Psi + \ddot\Phi + 2\mathcal{H}\dot\Phi + \left(\frac{k^2}{3} - K\right)\Phi\right]\delta^i_j Q^{(S)} \\ &+ \frac{k^2}{a^2}(\Phi - \Psi)Q^{(S)i}_j \,.\end{aligned} \tag{A3.20}$$

A3.1.4 The Weyl tensor

The Weyl tensor from scalar perturbations only has an 'electric' component, i.e., all the components are determined by

$$C^0_{i0j} \equiv -E_{ij} = \frac{k^2}{2}(\Phi + \Psi)Q^{(S)}_{ij} \,. \tag{A3.21}$$

More precisely we have

$$C_{0i0j} = a^2 E_{ij} \,, \tag{A3.22}$$

$$C_{0ijk} = 0 \,, \tag{A3.23}$$

$$C_{ijk\ell} = g_{ik} E_{j\ell} + g_{j\ell} E_{ik} - g_{jk} E_{i\ell} - g_{i\ell} E_{jk} \,. \tag{A3.24}$$

A3.2 Vector perturbations

We work in the vector gauge defined in Eq. (2.58),

$$ds^2 = a^2 \left(-dt^2 + 2\sigma\, Q_i^{(V)}\, dt dx^i + \gamma_{ij}\, dx^i\, dx^i \right) \,, \tag{A3.25}$$

where $\sigma^i = \sigma\, Q^{(V)i}$ is divergence-free and $Q_{ij}^{(V)} = -\frac{1}{2k}(Q_{i|j}^{(V)} + Q_{j|i}^{(V)})$.

A3.2.1 The Christoffel symbols

$$\delta\Gamma^0_{00} = 0 \,, \qquad \delta\Gamma^0_{0j} = \mathcal{H}\sigma\, Q_j^{(V)} \,, \tag{A3.26}$$

$$\delta\Gamma^j_{00} = [\dot\sigma + \mathcal{H}\sigma] Q^{(V)j} \,, \quad \delta\Gamma^j_{i0} = \frac{1}{2}\sigma \left(Q_{|i}^{(V)j} - Q_i^{(V)|j} \right) \,, \tag{A3.27}$$

$$\delta\Gamma^0_{ij} = k\sigma\, Q_{ij}^{(V)} \,, \qquad \delta\Gamma^j_{im} = -\mathcal{H}\sigma\, \gamma_{im} Q^{(V)j} \,. \tag{A3.28}$$

A3.2.2 The Riemann tensor

$$\delta R^0_{00j} = \dot{\mathcal{H}}\sigma\, Q^{(V)j} \,, \qquad \delta R^0_{0ij} = 0 \,, \tag{A3.29}$$

$$\delta R^0_{i0j} = k\, [\dot\sigma + \mathcal{H}\sigma]\, Q_{ij}^{(V)} \,, \tag{A3.30}$$

$$\delta R^0_{ijm} = -k\sigma \left(Q_{ij|m}^{(V)} - Q_{im|j}^{(V)} \right) \,, \tag{A3.31}$$

$$\delta R^i_{00j} = k\, [\dot\sigma + \mathcal{H}\sigma]\, Q_j^{(V)i} \,, \tag{A3.32}$$

$$\begin{aligned} \delta R^i_{0jm} = {} & \left[K + \mathcal{H}^2 \right] \sigma \left(\delta^i_j\, Q_m^{(V)} - \delta^i_m\, Q_j^{(V)} \right) \\ & + k\sigma \left[\left(Q_m^{(V)i} \right)_{|j} - \left(Q_j^{(V)i} \right)_{|m} \right] \,, \end{aligned} \tag{A3.33}$$

$$\begin{aligned} \delta R^i_{j0m} = {} & -\sigma \left[\mathcal{H}^2 \left(\delta^i_m\, Q_j^{(V)} - \gamma_{jm}\, Q^{(V)i} \right) + \dot{\mathcal{H}} \gamma_{jm}\, Q^{(V)i} \right. \\ & \left. - \frac{1}{2} \left(Q_j^{(V)|i} - Q_{|j}^{(V)i} \right)_{|m} \right] \,, \end{aligned} \tag{A3.34}$$

$$\delta R^i_{jmn} = k\mathcal{H}\sigma \left(\delta^i_m\, Q_{jn}^{(V)} - \delta^i_n\, Q_{jm}^{(V)} + Q_m^{(V)i} \gamma_{jn} - Q_n^{(V)i} \gamma_{jm} \right) \,. \tag{A3.35}$$

A3.2.3 The Ricci and Einstein tensors

The perturbation if the Ricci tensor is

$$\delta R_{00} = 0\,, \tag{A3.36}$$

$$\delta R_{0j} = \left[K + \frac{1}{2}k^2 + 2\mathcal{H}^2 + \dot{\mathcal{H}}\right]\sigma\, Q_j^{(V)}\,, \tag{A3.37}$$

$$\delta R_{ij} = k\left[\dot{\sigma} + 2\mathcal{H}\sigma\right] Q_{ij}^{(V)}\,. \tag{A3.38}$$

The vector perturbation of the Riemann scalar of course vanishes. For the Einstein tensor we find

$$\delta G_0^0 = 0\,, \tag{A3.39}$$

$$\delta G_j^0 = \frac{2K - k^2}{2a^2}\sigma\, Q_j^{(V)}\,, \tag{A3.40}$$

$$\delta G_0^j = \frac{1}{a^2}\left[2(\mathcal{H}^2 - \dot{\mathcal{H}}) + K + \frac{k^2}{2}\right]\sigma\, Q^{(V)j}\,, \tag{A3.41}$$

$$\delta G_j^i = \frac{k}{a^2}\left[\dot{\sigma} + 2\mathcal{H}\sigma\right] Q_j^{(V)i}\,. \tag{A3.42}$$

A3.2.4 The Weyl tensor

$$C^0_{\ i0j} = -\frac{k}{2}\dot{\sigma}^{(V)} Q_{ij}^{(V)} \equiv -E_{ij}^{(V)}\,, \tag{A3.43}$$

$$C_{ijk\ell} = g_{ik}E_{j\ell}^{(V)} + g_{j\ell}E_{ik}^{(V)} - g_{jk}E_{i\ell}^{(V)} - g_{i\ell}E_{jk}^{(V)}\,, \tag{A3.44}$$

$$\begin{aligned} C^0_{\ jlm} &\equiv \epsilon_{lmi}B^{(V)i}{}_j \\ &= \frac{1}{2}\sigma\left[Q^{(V)}_{l|jm} - Q^{(V)}_{m|jl} - \frac{k^2}{2}\gamma_{jl}Q_m^{(V)} + \frac{k^2}{2}\gamma_{jm}Q_l^{(V)}\right]. \end{aligned} \tag{A3.45}$$

All other components are determined by symmetry.

A3.3 Tensor perturbations

The metric is given by

$$ds^2 = a^2\left(-dt^2 + \left(\gamma_{ij} + 2H\,Q_{ij}^{(T)}\right)dx^i\,dx^i\right)\,, \tag{A3.46}$$

where $H_{ij} = H\,Q_{ij}^{(T)}$ is symmetric, traceless and divergence-free. For tensor perturbations all scalar- and vector-type quantities vanish and we shall not write them down here. The non-vanishing tensor perturbations are

A3.3.1 The Christoffel symbols

$$\delta\Gamma^i_{0j} = \dot{H}\,Q_j^{(T)i}, \qquad \delta\Gamma^0_{ij} = (2\mathcal{H}H + \dot{H})Q_{ij}^{(T)}\,, \tag{A3.47}$$

$$\delta\Gamma^i_{jm} = H\left(Q^{(T)i}_{j|m} + Q^{(T)i}_{m|j} - Q^{(T)|i}_{mj}\right)\,. \tag{A3.48}$$

A3.3.2 The Riemann tensor

$$\delta R^0_{i0j} = \left[\ddot{H} + \mathcal{H}\dot{H} + 2\dot{\mathcal{H}}H\right] Q^{(T)}_{ij} \,, \tag{A3.49}$$

$$\delta R^0_{ijm} = -\dot{H}\left(Q^{(T)}_{ij|m} - Q^{(T)}_{im|j}\right) \,, \tag{A3.50}$$

$$\delta R^i_{00j} = -\left[\ddot{H} + \mathcal{H}\dot{H}\right] Q^{(T)i}_{j} \,, \tag{A3.51}$$

$$\delta R^i_{0jm} = \dot{H}\left(Q^{(T)i}_{m|j} - Q^{(T)i}_{j|m}\right) \,, \tag{A3.52}$$

$$\delta R^i_{j0m} = -\dot{H}\left(Q^{(T)|i}_{jm} - Q^{(T)i}_{m|j}\right) \,, \tag{A3.53}$$

$$\begin{aligned}\delta R^i_{jmn} = {} & 2\mathcal{H}^2 H\left(\delta^i_m Q^{(T)}_{jn} - \delta^i_n Q^{(T)}_{jm}\right) \\ & + H\left(Q^{(T)i}_{j|nm} - Q^{(T)i}_{j|mn} + Q^{(T)i}_{n|jm} - Q^{(T)i}_{m|jn} + Q^{(T)|i}_{jm\ |n} - Q^{(T)|i}_{jn\ |m}\right) \\ & + \mathcal{H}\dot{H}\left(\delta^i_m Q^{(T)}_{jn} - \delta^i_n Q^{(T)}_{jm} - Q^{(T)i}_{n}\gamma_{jm} + Q^{(T)i}_{m}\gamma_{jn}\right) \,.\end{aligned} \tag{A3.54}$$

A3.3.3 The Ricci and Einstein tensors

$$\delta R_{ij} = \left[\ddot{H} + 2\mathcal{H}\dot{H} + (2\dot{\mathcal{H}} + 4\mathcal{H}^2 + k^2 + 6K)H\right] Q^{(T)}_{ij} \,, \tag{A3.55}$$

$$\delta G^i_{\ j} = \left[\ddot{H} + 2\mathcal{H}\dot{H} + (k^2 + 2K)H\right] Q^{(T)i}_{j} \,. \tag{A3.56}$$

A3.3.4 The Weyl tensor

$$C^0_{\ i0j} \equiv -E^{(T)}_{ij} = -\frac{1}{2}(\partial_t^2 - k^2)H\, Q^{(T)}_{ij} \,, \tag{A3.57}$$

$$C_{ijk\ell} = g_{ik}E^{(T)}_{j\ell} + g_{j\ell}E^{(T)}_{ik} - g_{jk}E^{(T)}_{i\ell} - g_{i\ell}E^{(T)}_{jk} \,, \tag{A3.58}$$

$$C^0_{\ jlm} \equiv \epsilon_{lmk}B^{(T)k}_{\ \ \ j} = -\dot{H}\left[Q^{(T)}_{jl|m} - Q^{(T)}_{jm|l}\right]. \tag{A3.59}$$

All other components are determined by symmetry.

Appendix 4

Special functions

A4.1 Legendre polynomials and Legendre functions

The **Legendre polynomials** form an orthonormal set of polynomials on the interval $[-1, 1]$. The lowest-order polynomials are $P_0 = 1$ and $P_1 = x$. The higher-order polynomials can then be obtained via the Gram–Schmidt orthogonalization procedure starting from the monomial x^n. They obey the normalization condition

$$\int_{-1}^{1} dx\; P_\ell(x) P_{\ell'}(x) = \frac{2}{2\ell+1}\,\delta_{\ell\ell'}\;. \tag{A4.1}$$

The Legendre polynomials can also be obtained via the recursion relation

$$(\ell+1)P_{\ell+1}(x) = (2\ell+1)x\,P_\ell(x) - \ell P_{\ell-1}(x)\;. \tag{A4.2}$$

They obey the differential equation

$$(1-x^2)P_\ell'' - 2x\,P_\ell' + \ell(\ell+1)P_\ell = 0\;. \tag{A4.3}$$

Rodrigues' formula

$$P_\ell(x) = \frac{1}{2^\ell \ell!}\frac{d^\ell}{dx^\ell}\left(x^2-1\right)^\ell\;. \tag{A4.4}$$

The lowest-order Legendre polynomials are given by

$$P_0 = 1\;, \tag{A4.5}$$

$$P_1 = x\;, \tag{A4.6}$$

$$P_2 = \frac{1}{2}(3x^2-1)\;, \tag{A4.7}$$

$$P_3 = \frac{1}{2}(5x^3-3x)\;, \tag{A4.8}$$

$$P_4 = \frac{1}{8}(35x^4-30x^2+3)\;. \tag{A4.9}$$

Clearly $P_\ell(-x) = (-1)^\ell P_\ell(x)$. Via induction, using Eq. (A4.2), one finds that

$$P_\ell(1) = 1\;. \tag{A4.10}$$

The Legendre polynomials obey the limiting relation

$$\lim_{\ell\to\infty} P_\ell\left(\cos(\theta/\ell)\right) = J_0(\theta)\;. \tag{A4.11}$$

Here J_0 is the Bessel function of order zero (see Appendix A4.3).

The associated **Legendre functions** are defined by

$$P_{\ell\, m}(x) = (1-x^2)^{m/2}\frac{d^m P_\ell(x)}{dx^m} = (1-x^2)^{m/2}\frac{1}{2^\ell \ell!}\frac{d^{\ell+m}}{dx^{\ell+m}}\left(x^2-1\right)^\ell \ , \tag{A4.12}$$

for $0 \le m \le \ell$. (We use the notation of Abramowitz & Stegun (1970) with $P_{\ell\, m} = (-1)^m P_\ell^m$.) The Legendre functions with $-\ell \le m < 0$ are given via the relation

$$P_{\ell -m} = \frac{(\ell-m)!}{(\ell+m)!}P_{\ell\, m} \ . \tag{A4.13}$$

From the above definition and Eq. (A4.10) one obtains

$$P_{\ell\, m}(1) = \delta_{m0} \ . \tag{A4.14}$$

The Legendre functions solve the differential equation

$$(1-x^2)P''_{\ell\, m} - 2xP'_{\ell\, m} + \left[\ell(\ell+1) - \frac{m^2}{1-x^2}\right]P_{\ell\, m} = 0 \ . \tag{A4.15}$$

They are in principle defined for arbitrary complex degree ℓ and order m as (meromorphic) functions of complex variables x. We shall only need them for integer m and non-negative integer ℓs with $|m| \le \ell$ and $x \in [-1, 1]$. In this interval and with these values of order and degree they are singularity-free and analytic.

The Legendre functions satisfy the following (and several more) recurrence relations:

$$P_{\ell\, m+1} = \frac{2mx}{\sqrt{1-x^2}}P_{\ell\, m} - [\ell(\ell+1) - m(m+1)]\, P_{\ell\, m-1} \ , \tag{A4.16}$$

$$xP_{\ell\, m} = \frac{\ell+m}{2\ell+1}P_{\ell-1\, m} + \frac{\ell-m+1}{2\ell+1}P_{\ell+1\, m} \ , \tag{A4.17}$$

$$\frac{dP_{\ell\, m}}{dx} = \frac{1}{2\sqrt{1-x^2}}\left[P_{\ell\, m+1} - (\ell+m)(\ell-m+1)P_{\ell\, m-1}\right] \ , \tag{A4.18}$$

$$(x^2-1)\frac{dP_{\ell\, m}}{dx} = \ell x P_{\ell\, m} + (m+\ell)P_{\ell-1\, m} \ , \tag{A4.19}$$

$$P_{\ell+1\, m} = P_{\ell-1\, m} + (2\ell+1)\sqrt{1-x^2}P_{\ell\, m-1} \ . \tag{A4.20}$$

The parity relation of the associated Legendre function is a simple consequence of their definition, $P_{\ell\, m}(-x) = (-1)^{\ell+m}P_{\ell\, m}(x)$. Also of importance for us is the orthogonality relation

$$\begin{aligned}&\int_{-1}^{1} P_{\ell\, m}(x)P_{\ell'\, m}(x)\, dx \\ &\quad = \int_0^\pi P_{\ell\, m}(\cos\vartheta)P_{\ell'\, m}(\cos\vartheta)\sin\vartheta\, d\vartheta = \frac{2}{2\ell+1}\frac{(\ell+m)!}{(\ell-m)!}\delta_{\ell\,\ell'} \ .\end{aligned} \tag{A4.21}$$

The derivation of most of these results and more can be found in Arfken & Weber (2001).

A4.2 Spherical harmonics

A4.2.1 The irreducible representations of the rotation group

Here we briefly repeat some basics about the rotation group and its irreducible representations. Much more can be found in most quantum mechanics books, e.g., Sakurai (1993). Here we are only interested in ordinary (i.e., not projective) representations and therefore integer values of the angular momentum. For a function Ψ on the sphere we define its transformation under rotations $R \in SO(3)$ by

$$[\mathcal{U}(R)\Psi](\mathbf{n}) \equiv \Psi\left(R^{-1}\mathbf{n}\right) \quad \forall\, \mathbf{n} \in \mathbb{S}^2 \,. \tag{A4.22}$$

This is clearly a unitary representation of the rotation group on $\mathcal{L}^2(\mathbb{S}^2)$, i.e. the Hilbert space of square integrable functions on the sphere.

The one-parameter subgroup of rotations around a given axis $\mathbf{e}$ is

$$R(\mathbf{e},\alpha)\mathbf{n} = \cos\alpha\mathbf{n} + [1-\cos\alpha]\,(\mathbf{e}\cdot\mathbf{n})\mathbf{e} + \sin\alpha\mathbf{e}\wedge\mathbf{n} \,. \tag{A4.23}$$

Its generator is defined by

$$\Omega(\mathbf{e})\mathbf{n} = \left.\frac{d}{d\alpha}R(\mathbf{e},\alpha)\mathbf{n}\right|_{\alpha=0} .$$

With Eq. (A4.23) we obtain

$$\Omega(\mathbf{e})_{ij} = \mathbf{e}_k I^k_{ij} \,, \quad \text{where} \quad I^k_{ij} = -\epsilon_{ijk} \,. \tag{A4.24}$$

The generator of $\mathcal{U}(R(\mathbf{e},\alpha))$ is the angular momentum in direction $\mathbf{e}$;

$$\left.\frac{d}{d\alpha}\mathcal{U}(R(\mathbf{e},\alpha))\right|_{\alpha=0} \equiv \mathcal{U}_*(I^j)\mathbf{e}_j = \frac{i}{\hbar}L^j\mathbf{e}_j \,,$$

with

$$\mathbf{L} = -i\hbar\mathbf{x}\wedge\nabla \,. \tag{A4.25}$$

In spherical coordinates (r,ϑ,φ) one finds

$$\mathbf{L} = i\hbar\begin{pmatrix} \sin\varphi\cot\vartheta\,\partial_\varphi + \cos\varphi\partial_\vartheta \\ \cos\varphi\cot\vartheta\,\partial_\varphi - \sin\varphi\partial_\vartheta \\ -\partial_\varphi \end{pmatrix} . \tag{A4.26}$$

One easily verifies that the matrices I_k and the operators L_k satisfy the commutation relations

$$\left[I_j, I_k\right] = \epsilon_{jkl}I_l \,, \tag{A4.27}$$

$$\left[L_j, L_k\right] = +i\hbar\epsilon_{jkl}L_l \,. \tag{A4.28}$$

Introducing also $L_\pm = L_1 \pm L_2$ and $\mathbf{L}^2 = L_1^2 + L_2^2 + L_3^2$ one finds the commutation relations

$$\left[\mathbf{L}^2, L_j\right] = 0 = \left[\mathbf{L}^2, L_\pm\right] \,, \tag{A4.29}$$

$$[L_3, L_\pm] = \pm\hbar L_\pm \,, \quad \text{and} \tag{A4.30}$$

$$L_\pm L_\mp = \mathbf{L}^2 - L_3^2 \pm \hbar L_3 \,. \tag{A4.31}$$

Let us now consider a representation of the rotation group on some finite-dimensional vector space $\mathcal{V}$. Since $\mathbf{L}^2$ and L_3 are commuting hermitian operators, we can find an orthonormal basis of simultaneous eigenvectors of $\mathbf{L}^2$ and L_3. We order them according to

their eigenvalue of L_3, so that the eigenvalue of ψ_1 is maximal. Let us call it $\hbar a$. Furthermore, $\hbar^2 b$ denotes the corresponding eigenvalue of $\mathbf{L}^2$. Hence $L_3\psi_1 = \hbar a\psi_1$ and $\mathbf{L}^2\psi_1 = \hbar^2 b\psi_1$. Equation (A4.30) gives $L_3L_+\psi_1 = \hbar(a+1)L_+\psi_1$. Hence $L_+\psi_1$ is an eigenvector of L_3 with eigenvalue $\hbar(a+1)$ or zero. Since $\hbar a$ is maximal, this implies $L_+\psi_1 = 0$. With Eq. (A4.31) therefore $0 = (\mathbf{L}^2 - L_3^2 - \hbar L_3)\psi_1 = \hbar^2(b - a(a+1))\psi_1$, so that $b = a(a+1)$. Applying L_- on ψ_1 and using again Eq. (A4.30), we find that $L_-\psi_1$ is also an eigenvector of L_3 with eigenvalue $\hbar(a-1)$. Repeated application of L_- shows that $(L_-)^m\psi_1$ is an eigenvector of L_3 with eigenvalue $\hbar(a-m)$. We finally arrive at the eigenvector with the lowest eigenvalue $\hbar(a-n)$ of L_3, let us call it

$$\psi_{n+1} = (L_-)^n\psi_1 / \parallel (L_-)^n\psi_1 \parallel .$$

Necessarily, $L_-\psi_{n+1} = 0$ since it would otherwise have an even lower eigenvalue of L_3. From this we conclude

$$0 = L_+(L_-)^{n+1}\psi_1 = (\mathbf{L}^2 - L_3^2 + \hbar L_3)(L_-)^n\psi_1 = \hbar^2(b - (a-n)^2 + a - n)(L_-)^n\psi_1 ,$$

so that $b = (a-n)^2 + n - a$. Together with the previous identity, $b = a(a+1)$, this implies $a = n/2$. Therefore, a must be an integer or half-integer number, the representation with $a = \ell$ is denoted by $D^{(\ell)}$ and has dimension $n+1 = 2\ell+1$. The induced representation of the generators (the Lie algebra) defines the angular momentum, $L_j = i\hbar D_*^{(\ell)}(I_j)$. The vector space which carries $D^{(\ell)}$ is denoted by $\mathcal{V}^{(\ell)}$. The vectors $\big((L_-^n)\psi_1 / \parallel (L_-^n)\psi_1 \parallel\big)_{n=0}^{2\ell}$ form an orthonormal basis of eigenvectors of L_3 and $\mathbf{L}^2$; the so-called canonical basis. The eigenvalues of L_3 are $\hbar\ell,\ \hbar(\ell-1),\ \ldots,\ -\hbar\ell$. The operator $\mathbf{L}^2$ is constant on $\mathcal{V}^{(\ell)}$ with eigenvalue $\hbar^2\ell(\ell+1)$. $D^{(0)}$ is the trivial representation, $D^{(0)}(R) = \mathbb{1}$, which is irreducible only on a one-dimensional space; and $D^{(1)}$ is the identical representation, $D^{(1)}(R) = R$.

In the next section, when we realize these representations on $\mathcal{L}^2(\mathbb{S}^2)$ we shall see that only the representations $D_*^{(\ell)}$ with integer ℓ can be lifted to representations of the rotation group. Half-integer ℓs give projective representation with $D^{(\ell)}(R_1)D^{(\ell)}(R_2) = \pm D^{(\ell)}(R_1R_2)$ which are relevant in quantum mechanics where a state is only defined up to a constant phase. The existence of particles with half-integer spin, fermions, is a purely quantum mechanical phenomenon.

Acting with $L_\pm$ we can pass from one basis vector to every other in $\mathcal{V}^{(\ell)}$. Hence the representation $D^{(\ell)}$ is irreducible, i.e., $\mathcal{V}^{(\ell)}$ contains no invariant subspaces. We have obtained all irreducible representations of the rotation group in this way.

A4.2.2 The Clebsch–Gordan decomposition

The tensor product, $\mathcal{V}^{(\ell)} \otimes \mathcal{V}^{(\ell')}$ carries the tensor representation $D^{(\ell)} \otimes D^{(\ell')}$. In general this representation is not irreducible but can be decomposed into a sum of irreducible representations. We show that

$$D^{(\ell)} \otimes D^{(\ell')} = \sum_{j=|\ell-\ell'|}^{\ell+\ell'} D^{(j)} . \tag{A4.32}$$

This sum is called the Clebsch–Gordan series.

Without loss of generality, we assume $\ell \geq \ell'$. The Leibnitz rule, implies that the induced representation on the generators is $(D^{(\ell)} \otimes D^{(\ell')})_* = (D_*^{(\ell)} \otimes \mathbb{1}) \oplus (\mathbb{1} \otimes D_*^{(\ell')})$. We denote the canonical basis on $\mathcal{V}^{(\ell)}$ by $(\psi_{\ell\, m})_{m=-\ell}^{\ell}$. L_3 takes once the maximal value on the state

$\psi_{\ell\,\ell}\otimes\psi_{\ell'\,\ell'}$, where we have $L_3(\psi_{\ell\,\ell}\otimes\psi_{\ell'\,\ell'}) = (L_3\psi_{\ell\,\ell})\otimes\psi_{\ell'\,\ell'} + \psi_{\ell\,\ell}\otimes(L_3\psi_{\ell'\,\ell'}) = \hbar(\ell+\ell')\psi_{\ell\,\ell}\otimes\psi_{\ell'\,\ell'}$. Hence $D^{(\ell)}\otimes D^{(\ell')}$ contains $D^{(j)}$ with $j=\ell+\ell'$ exactly once and it does not contain any higher angular momentum. However, there are two states with $L_3\phi = \hbar(\ell+\ell'-1)\phi$, namely the states $\psi_{\ell\,\ell-1}\otimes\psi_{\ell'\,\ell'}$ and $\psi_{\ell\,\ell}\otimes\psi_{\ell'\,\ell'-1}$, hence $D^{(\ell)}\otimes D^{(\ell')}$ must also contain $D^{(\ell+\ell'-1)}$ (except if $\ell'=0$). Furthermore, there are three states with $L_3\phi = \hbar(\ell+\ell'-2)\phi$, namely the states $\psi_{\ell\,\ell-2}\otimes\psi_{\ell'\,\ell'}$, $\psi_{\ell\,\ell-1}\otimes\psi_{\ell'\,\ell'-1}$ and $\psi_{\ell\,\ell}\otimes\psi_{\ell'\,\ell'-2}$, hence $D^{(\ell)}\otimes D^{(\ell')}$ must in addition contain $D^{(\ell+\ell'-2)}$. This goes on until the eigenvalue $\ell+\ell'-m$, with $m=2\ell'$ is reached, which has an eigenspace of dimension $m+1$ and which also implies that the representation $D^{(\ell-\ell')}$ is contained. For higher values of m the dimension of the eigenspace is reduced by one at each step and is therefore just sufficient to contain the eigenvectors of each of the representations $D^{(j)}$ already inferred. This can also be concluded from the fact that $(2\ell+1)(2\ell'+1) = \sum_{j=\ell-\ell'}^{\ell+\ell'}(2j+1)$ and therefore the dimension of the total space agrees with the sum of the dimensions of all the irreducible representations already defined. This proves the Clebsch–Gordan series.

The matrix which induces the change of basis from the tensor product basis to the canonical basis on each of the irreducible pieces of $\mathcal{V}^{(\ell)}\otimes\mathcal{V}^{(\ell')}$ defines the Clebsch–Gordan coefficients in the following way. We have seen that

$$\mathcal{V}^{(\ell)}\otimes\mathcal{V}^{(\ell')} = \mathcal{W}^{(\ell+\ell')}\oplus\mathcal{W}^{(\ell+\ell'-1)}\oplus\cdots\oplus\mathcal{W}^{(|\ell-\ell'|)}\,,$$

where $\mathcal{W}^{(j)}$ carries the representation $D^{(j)}$ of the rotation group. Let us denote the canonical basis in $\mathcal{W}^{(j)}$ by $\big(\phi_{j\,m}\big)_{m=-j}^{j}$. The transformation from the basis $(\psi_{\ell\,m}\otimes\psi_{\ell'\,m'})_{m,m'=-\ell,-\ell'}^{\ell,\ell'}$ to the basis $\Big(\big(\phi_{j\,m}\big)_{m=-j}^{j}\Big)_{j=|\ell-\ell'|}^{\ell+\ell'}$ is of the form

$$\phi_{j\,m_j} = \sum_{m,m'}\langle\ell,\ell';m,m'|j,m_j\rangle\psi_{\ell\,m}\otimes\psi_{\ell'\,m'}\,. \tag{A4.33}$$

The complex coefficients $\langle\ell,\ell';m,m'|j,m_j\rangle$ are called Clebsch–Gordan coefficients. In the literature they are often denoted by $\langle\ell,\ell';m,m'|j,m_j\rangle \equiv \langle\ell,\ell';m,m'|\ell,\ell';j,m_j\rangle$. We shall not repeat the redundant numbers ℓ,ℓ' in the second argument. From the above discussion it is clear that $\langle\ell,\ell';m,m'|j,m_j\rangle \neq 0$ only if $m_j = m+m'$ and $j\in\{\ell+\ell',\ell+\ell'-1,\ldots,|\ell-\ell'|\}$. The general formula for these coefficients is given below (Abramowitz & Stegun, 1970). They can also be computed with Mathematica.

$$\begin{aligned}
&\langle\ell_1,\ell_2;m_1,m_2|j,m_1+m_2\rangle\\
&\quad= \sqrt{\frac{(\ell_1+\ell_2-j)!(j+\ell_1-\ell_2)!(j+\ell_2-\ell_1)!(2j+1)}{(\ell_1+\ell_2+j+1)!}}\\
&\qquad\times\sum_k\left[\frac{(-1)^k\sqrt{(\ell_1+m_1)!(\ell_1-m_1)!(\ell_2+m_2)!(\ell_2-m_2)!}}{k!(\ell_1+\ell_2-j-k)!(\ell_1-m_1-k)!}\right.\\
&\qquad\left.\times\frac{\sqrt{(j+m_1+m_2)!(j-m_1-m_2)!}}{(\ell_2+m_2-k)!(j-\ell_2+m_1+k)!(j-\ell_1-m_2+k)!}\right].
\end{aligned} \tag{A4.34}$$

In the sum over k only the terms with a finite denominator contribute, hence $k\geq\max\{0,\ \ell_2-j-m_1,\ \ell_1-j+m_2\}$ and $k\leq\min\{\ell_1+\ell_2-j,\ \ell_1-m_1,\ \ell_2+m_2\}$.

Table A4.1. *The non-vanishing Clebsch–Gordan coefficients for* $\ell_2 = 1$. $\langle \ell, 1; m, m_2 | j, m + m_2 \rangle$.

j	$m_2 = 1$	$m_2 = 0$	$m_2 = -1$
$\ell + 1$	$\sqrt{\frac{(\ell+m+1)(\ell+m+2)}{(2\ell+1)(2\ell+2)}}$	$\sqrt{\frac{(\ell+m+1)(\ell-m+1)}{(2\ell+1)(\ell+1)}}$	$\sqrt{\frac{(\ell-m+1)(\ell-m+2)}{(2\ell+1)(2\ell+2)}}$
ℓ	$-\sqrt{\frac{(\ell+m+1)(\ell-m)}{2\ell(\ell+1)}}$	$\frac{m}{\sqrt{\ell(\ell+1)}}$	$\sqrt{\frac{(\ell-m+1)(\ell+m)}{2\ell(\ell+1)}}$
$\ell - 1$	$\sqrt{\frac{(\ell-m-1)(\ell-m)}{2\ell(2\ell+1)}}$	$-\sqrt{\frac{(\ell+m)(\ell-m)}{\ell(2\ell+1)}}$	$\sqrt{\frac{(\ell+m)(\ell+m-1)}{2\ell(2\ell+1)}}$

In Chapter 5 we need the Clebsch–Gordan coefficients $\langle \ell_1, \ell_2; m_1, m_2 | j, m_1 + m_2 \rangle$ for $\ell_2 \leq 2$. We therefore give the non-vanishing ones of these coefficients in Tables A4.1 and A4.2. Of course $\langle \ell_1, 0; m_1, 0 | \ell, m \rangle = \delta_{\ell_1 \ell}\, \delta_{m_1 m}$.

A4.2.3 Spherical harmonics of spin-0

The spherical harmonics are functions on the sphere. For a unit vector $\mathbf{n}$ defined by its polar angles (ϑ, φ) the spherical harmonics are given by

$$Y_{\ell m}(\mathbf{n}) = (-1)^m \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} e^{im\varphi} P_{\ell m}(\mu) \,, \qquad \mu = \cos\vartheta \,. \tag{A4.35}$$

From the parity transformation properties and the orthogonality of the associated Legendre functions, Eq. (A4.21), we conclude $Y_{\ell -m} = (-1)^m \bar{Y}_{\ell m}$ and

$$\int Y_{\ell m}(\mathbf{n}) \bar{Y}_{\ell' m'}(\mathbf{n}) d\Omega_{\mathbf{n}} = \delta_{\ell\ell'} \delta_{mm'} \,. \tag{A4.36}$$

We now show that the spherical harmonics $(Y_{\ell m})_{m=-\ell}^{\ell}$ carry the representation $D^{(\ell)}$.

From Eq. (A4.26) we know $L_3 = -i\hbar\partial_\varphi$. Therefore, the set of functions $f_{\ell m}$ which forms a canonical basis for the representation $D^{(\ell)}$ must be of the form $f_{\ell m} = \exp(im\varphi) g_{\ell m}(\mu)$. Furthermore, Eq. (A4.26) implies

$$\mathbf{L}^2 = L_1^2 + L_2^2 = L_3^2 = -\hbar^2 \left[\frac{1}{\sin\vartheta} \partial_\vartheta \, \sin\vartheta \, \partial_\vartheta + \frac{1}{\sin^2\vartheta} \partial_\varphi^2 \right] = -\hbar^2 \Delta \,,$$

where Δ denotes the Laplacian on the 2-sphere. For $f_{\ell m} = \exp(im\varphi) g_{\ell m}(\mu)$ we obtain

$$\Delta f_{\ell m} = \left[(1-\mu^2) \frac{d^2}{d\mu^2} - 2\mu \frac{d}{d\mu} - \frac{m^2}{1-\mu^2} \right] g_{\ell m}(\mu) \exp(im\varphi) \,.$$

With $\mathbf{L}^2 = \hbar^2 \ell(\ell+1)$ it follows that $g_{\ell m}(\mu)$ satisfies the differential equation of the associated Legendre function, Eq. (A4.15), hence $g_{\ell m} = c_{\ell m} P_{\ell m}(\mu)$. The constants $c_{\ell m}$ are chosen to normalize the functions $f_{\ell m}$. Furthermore, since $f_{\ell m}$ and $f_{\ell' m'}$ are eigenfunctions with different eigenvalues for some hermitian operator (L_3 if $m \neq m'$ and $\mathbf{L}^2$ if $\ell \neq \ell'$) they are certainly orthogonal. Hence the functions $f_{\ell m}$ are proportional to the

Table A4.2. *The non-vanishing Clebsch–Gordan coefficients for* $\ell_2 = 2$. $\langle \ell, 2; m, m_2 | j, m + m_2 \rangle$.

j	$m_2 = 2$	$m_2 = 1$
$\ell+2$	$\sqrt{\frac{(\ell+m+1)(\ell+m+2)(\ell+m+3)(\ell+m+4)}{(2\ell+1)(2\ell+2)(2\ell+3)(2\ell+4)}}$	$\sqrt{\frac{(\ell-m+1)(\ell+m+3)(\ell+m+2)(\ell+m+1)}{(2\ell+1)(\ell+1)(2\ell+3)(\ell+2)}}$
$\ell+1$	$-\sqrt{\frac{(\ell+m+1)(\ell+m+2)(\ell+m+3)(\ell-m)}{2\ell(\ell+1)(\ell+2)(2\ell+1)}}$	$-(\ell-2m)\sqrt{\frac{(\ell+m+2)(\ell+m+1)}{2\ell(2\ell+1)(\ell+1)(\ell+2)}}$
ℓ	$\sqrt{\frac{3(\ell+m+1)(\ell+m+2)(\ell-m-1)(\ell-m)}{(2\ell-1)2\ell(\ell+1)(2\ell+3)}}$	$-(1+2m)\sqrt{\frac{3(\ell-m)(\ell+m+1)}{\ell(2\ell-1)(2\ell+2)(2\ell+3)}}$
$\ell-1$	$-\sqrt{\frac{(\ell+m+1)(\ell-m-2)(\ell-m-1)(\ell-m)}{2(\ell-1)\ell(\ell+1)(2\ell+1)}}$	$(\ell+2m+1)\sqrt{\frac{(\ell-m)(\ell-m-1)}{\ell(\ell-1)(2\ell+1)(2\ell+2)}}$
$\ell-2$	$\sqrt{\frac{(\ell-m-3)(\ell-m-2)(\ell-m-1)(\ell-m)}{(2\ell-2)(2\ell-1)2\ell(2\ell+1)}}$	$-\sqrt{\frac{(\ell-m)(\ell-m-1)(\ell-m-2)(\ell+m)}{(\ell-1)(2\ell-1)\ell(2\ell+1)}}$

j	$m_2 = 0$	$m_2 = -1$
$\ell+2$	$\sqrt{\frac{3(\ell-m+2)(\ell-m+1)(\ell+m+2)(\ell+m+1)}{(2\ell+1)(2\ell+2)(2\ell+3)(\ell+2)}}$	$\sqrt{\frac{(\ell-m+3)(\ell-m+2)(\ell-m+1)(\ell+m+1)}{(2\ell+1)(\ell+1)(2\ell+3)(\ell+2)}}$
$\ell+1$	$m\sqrt{\frac{3(\ell-m+1)(\ell+m+1)}{\ell(2\ell+1)(\ell+1)(\ell+2)}}$	$(\ell+2m)\sqrt{\frac{(\ell-m+2)(\ell-m+1)}{2\ell(2\ell+1)(\ell+1)(\ell+2)}}$
ℓ	$\frac{3m^2-\ell(\ell+1)}{\sqrt{(2\ell-1)\ell(2\ell+3)(\ell+1)}}$	$(2m-1)\sqrt{\frac{3(\ell-m+1)(\ell+m)}{\ell(2\ell-1)(2\ell+2)(2\ell+3)}}$
$\ell-1$	$-m\sqrt{\frac{3(\ell-m)(\ell+m)}{(\ell-1)\ell(2\ell+1)(\ell+1)}}$	$-(\ell-2m+1)\sqrt{\frac{(\ell+m)(\ell+m-1)}{\ell(\ell-1)(2\ell+1)(2\ell+2)}}$
$\ell-2$	$\sqrt{\frac{3(\ell-m)(\ell-m-1)(\ell+m)(\ell+m-1)}{(2\ell-2)(2\ell-1)(2\ell+1)\ell}}$	$-\sqrt{\frac{(\ell-m)(\ell+m)(\ell+m-1)(\ell+m-2)}{(\ell-1)(2\ell-1)\ell(2\ell+1)}}$

j	$m_2 = -2$
$\ell+2$	$\sqrt{\frac{(\ell-m+1)(\ell-m+2)(\ell-m+3)(\ell-m+4)}{(2\ell+1)(2\ell+2)(2\ell+3)(2\ell+4)}}$
$\ell+1$	$\sqrt{\frac{(\ell-m+1)(\ell-m+2)(\ell-m+3)(\ell+m)}{\ell(2\ell+1)(\ell+1)(2\ell+4)}}$
ℓ	$\sqrt{\frac{3(\ell-m+1)(\ell-m+2)(\ell+m-1)(\ell+m)}{\ell(2\ell-1)(2\ell+2)(2\ell+3)}}$
$\ell-1$	$\sqrt{\frac{(\ell-m+1)(\ell+m-2)(\ell+m-1)(\ell+m)}{(\ell-1)\ell(2\ell+1)(2\ell+2)}}$
$\ell-2$	$\sqrt{\frac{(\ell+m-3)(\ell+m-2)(\ell+m-1)(\ell+m)}{(2\ell-2)(2\ell-1)2\ell(2\ell+1)}}$

spherical harmonics and obey the same normalization condition, i.e., they are the spherical harmonics $Y_{\ell m}$.

We can relate the spherical harmonic $Y_{\ell m}(\mathbf{n})$ to the matrix element $D^{(\ell)}_{m0}(R)$, where R is a rotation which turns $\mathbf{e}_z$ into $\mathbf{n}$. To do this we observe that the spherical harmonics of order ℓ form an orthonormal basis for the $(2\ell+1)$-dimensional space of functions on the sphere which transform with the representation $D^{(\ell)}$ under rotation. Let f be such a

function and (f_m) be its coefficients in the basis $Y_{\ell m}$. In other words,

$$f(\mathbf{n}) = \sum_m f_m Y_{\ell m}(\mathbf{n}) \,.$$

Under a rotation, the vector (f_m) transforms with $D^{(\ell)}_{m_1 m_2}$ so that

$$f(R^{-1}\mathbf{n}) = \sum_{m_1} \left(\sum_{m_2} D^{(\ell)}_{m_1 m_2}(R) f_{m_2} \right) Y_{\ell m_1}(\mathbf{n}) \,.$$

Considering the function with $f_{m_2} = \delta_{m m_2}$ this yields

$$Y_{\ell m}(R^{-1}\mathbf{n}) = \sum_{m_1} D^{(\ell)}_{m_1 m}(R) Y_{\ell m_1}(\mathbf{n}) \,. \tag{A4.37}$$

Let us now consider $\mathbf{n} = \mathbf{e}_z$. Using $Y_{\ell m}(\mathbf{e}_z) = \delta_{0m}\sqrt{(2\ell+1)/4\pi}$ we arrive at

$$Y_{\ell m}(R^{-1}\mathbf{e}_z) = \sqrt{\frac{2\ell+1}{4\pi}} D^{(\ell)}_{0m}(R) \,. \tag{A4.38}$$

If R is an (otherwise arbitrary) rotation which turns $\mathbf{n}$ into $\mathbf{e}_z$, so that $R^{-1}\mathbf{e}_z = \mathbf{n}$, we therefore have

$$Y_{\ell m}(\mathbf{n}) = \sqrt{\frac{2\ell+1}{4\pi}} D^{(\ell)}_{0m}(R) \,. \tag{A4.39}$$

Usually one chooses for R the rotation with Euler angles $(0, -\vartheta, -\varphi)$ for the unit vector $\mathbf{n}$ with polar angles (ϑ, φ). We denote the representation matrix of the rotation by Euler angles (α, β, γ) by $D^{(\ell)}_{mn}(\alpha, \beta, \gamma)$: first a rotation by angle γ around the z-axis, then a rotation by angle β around the y-axis and finally a rotation by angle α around the (new) z-axis. The inverse of the rotation (α, β, γ) is the rotation with Euler angles $(-\gamma, -\beta, -\alpha)$. Observing that the representation $D^{(\ell)}$ is unitary we find $D^{(\ell)}_{0m}(0, -\vartheta, -\varphi) = D^{(\ell)^{-1}}_{0m}(\varphi, \vartheta, 0) = \bar{D}^{(\ell)}_{m0}(\varphi, \vartheta, 0)$ so that we can also write

$$Y_{\ell m}(\mathbf{n}) = \sqrt{\frac{2\ell+1}{4\pi}} \bar{D}^{(\ell)}_{m0}(\varphi, \vartheta, 0) \,. \tag{A4.40}$$

With this it is now easy to show the addition theorem of spherical harmonics. Consider two directions $\mathbf{n}_1$ and $\mathbf{n}_2$ separated by an angle γ, $\cos\gamma = \mathbf{n}_1 \cdot \mathbf{n}_2$. We denote the rotation with Euler angles $(\varphi_1,\ \vartheta_1, 0)$ by R_1. It rotates $\mathbf{e}_z$ into $\mathbf{n}_1$. With Eqs. (A4.37) and (A4.40) we have

$$Y_{\ell 0}(R_1^{-1}\mathbf{n}_2) = \sum_m D^{(\ell)}_{m0}(R_1) Y_{\ell\, m}(\mathbf{n}_2) = \sqrt{\frac{4\pi}{2\ell+1}} \sum_m \bar{Y}_{\ell m}(\mathbf{n}_1) Y_{\ell m}(\mathbf{n}_2) \,. \tag{A4.41}$$

But since R_1^{-1} rotates $\mathbf{n}_1$ into $\mathbf{e}_z$, the polar angle ϑ of $R_1^{-1}\mathbf{n}_2$ is simply the angle between $\mathbf{n}_1$ and $\mathbf{n}_2$, so that $Y_{\ell 0}(R_1^{-1}\mathbf{n}_2) = \sqrt{(2\ell+1)/4\pi}\, P_\ell(\mathbf{n}_1 \cdot \mathbf{n}_2)$. Inserting this above yields the addition theorem for spherical harmonics,

$$\frac{2\ell+1}{4\pi} P_\ell(\mathbf{n}_1 \cdot \mathbf{n}_2) = \sum_{m=-\ell}^{\ell} \bar{Y}_{\ell m}(\mathbf{n}_1) Y_{\ell m}(\mathbf{n}_2) \,. \tag{A4.42}$$

The lowest ℓ spherical harmonics are given by

$$\ell = 0 \quad Y_{00} = \frac{1}{\sqrt{4\pi}} \,, \tag{A4.43}$$

$$\ell = 1 \quad \begin{cases} Y_{11} = -\sqrt{\frac{3}{8\pi}} \sin\vartheta \, e^{i\varphi} \,, \\ Y_{10} = \sqrt{\frac{3}{4\pi}} \cos\vartheta \,, \end{cases} \tag{A4.44}$$

$$\ell = 2 \quad \begin{cases} Y_{22} = \sqrt{\frac{15}{32\pi}} \sin^2\vartheta \, e^{2i\varphi} \,, \\ Y_{21} = -\sqrt{\frac{15}{8\pi}} \sin\vartheta \cos\vartheta \, e^{i\varphi} \,, \\ Y_{20} = \sqrt{\frac{5}{4\pi}} \left(\frac{3}{2}\cos^2\vartheta - \frac{1}{2}\right) \,, \end{cases} \tag{A4.45}$$

$$\ell = 3 \quad \begin{cases} Y_{33} = -\sqrt{\frac{35}{64\pi}} \sin^3\vartheta \, e^{3i\varphi} \,, \\ Y_{32} = \sqrt{\frac{105}{32\pi}} \sin^2\vartheta \cos\vartheta \, e^{2i\varphi} \,, \\ Y_{31} = -\sqrt{\frac{21}{16\pi}} \sin\vartheta \left(\frac{5}{2}\cos^2\vartheta - \frac{1}{2}\right) e^{i\varphi} \,, \\ Y_{30} = \sqrt{\frac{7}{4\pi}} \cos\vartheta \left(\frac{5}{2}\cos^2\vartheta - \frac{3}{2}\right) \,, \end{cases} \tag{A4.46}$$

$$Y_{\ell\,-m} = (-1)^m Y^*_{\ell m} \,. \tag{A4.47}$$

A4.2.4 Spherical harmonics of spin s[1]

We now consider tensor fields on the sphere. We can express their components in terms of the standard 'real' orthonormal basis, $\mathbf{e}_1 = \mathbf{e}_\vartheta = \partial_\vartheta$, $\mathbf{e}_2 = \mathbf{e}_\varphi = \frac{1}{\sin\vartheta}\partial_\varphi$ or in terms of the helicity basis

$$\mathbf{e}_+ = \frac{1}{\sqrt{2}} (\mathbf{e}_1 - i\mathbf{e}_2) \,, \quad \mathbf{e}_- = \frac{1}{\sqrt{2}} (\mathbf{e}_1 + i\mathbf{e}_2) \,. \tag{A4.48}$$

Here we identify, as is often done, a vector with the derivative in a given direction. A vector is then defined by its action on functions: $\mathbf{e}_1 f = \partial_\vartheta f$ and $\mathbf{e}_2 f = \frac{1}{\sin\vartheta}\partial_\varphi f$. Note also that the metric $ds^2 = d\vartheta^2 + \sin^2\vartheta \, d\varphi^2$ on the sphere has the components $g_{-+} = g_{+-} = 1$, and $g_{++} = g_{--} = 0$ in the helicity basis.

Under a rotation of the 'real' basis, $\mathbf{e}_1 \to \cos\gamma \mathbf{e}_1 - \sin\gamma \mathbf{e}_2$, $\mathbf{e}_2 \to \cos\gamma \mathbf{e}_2 + \sin\gamma \mathbf{e}_1$, the helicity basis transforms as $\mathbf{e}_+ \to e^{-i\gamma}\mathbf{e}_+$, $\mathbf{e}_- \to e^{i\gamma}\mathbf{e}_-$.

The components of a tensor field of rank r in the helicity basis transform under a rotation by

$$T^{\overbrace{+\cdots+}^{s}\overbrace{-\cdots-}^{r-s}} \to e^{i(2s-r)\gamma} T^{\overbrace{+\cdots+}^{s}\overbrace{-\cdots-}^{r-s}} \,. \tag{A4.49}$$

For example, the components of a vector transform as $V^+ \to e^{i\gamma} V^+$ and $V^- \to e^{-i\gamma} V^-$. (The vector itself $V = V^+\mathbf{e}_+ + V^-\mathbf{e}_-$ is invariant, hence the components transform 'contragradient' to the basis.) Components which transform with $e^{is\gamma}$ are called components of spin $|s|$ or of helicity s.

[1] A more detailed treatment of spin weighted spherical harmonics can be found in Goldberg (1967) and Newman & Penrose (1966).

We are mainly interested in symmetric (real) rank-2 tensors like the polarization. These have one spin-0 component $T^{+-} = T^{-+} = \frac{1}{2}\mathrm{tr}\, T \equiv I$, one spin-2 component with positive helicity, $T^{++} = \frac{1}{2}(T_{11} - T_{22}) - iT_{12}$ and one spin-2 component with negative helicity, $T^{--} = \frac{1}{2}(T_{11} - T_{22}) + iT_{12}$ (see Ex. A4.1).

We can write a symmetric rank-2 tensor in these components as

$$T = \frac{1}{2}\left[I\,\mathbb{1} + T^{++}\sigma_+ + T^{--}\sigma_-\right] , \tag{A4.50}$$

where $\sigma_\pm = \sigma^3 \pm i\sigma^1$ are given by the Pauli matrices,

$$\sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \qquad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

To expand a spin-s component of a tensor field on the sphere, one employs the spin weighted spherical harmonics. These are spin-s components of tensor fields on the 2-sphere. In the basis $(\mathbf{e}_\vartheta, \mathbf{e}_\varphi)$ they are given in terms of the irreducible representations of the rotation group by

$$ {}_sY_{\ell m}(\vartheta, \varphi) \equiv \sqrt{\frac{2\ell+1}{4\pi}}\,\bar{D}^{(\ell)}_{m\,-s}(\varphi, \vartheta, 0) , \tag{A4.51}$$

$$= (-1)^m \sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell+m)!(\ell-m)!}{(\ell+s)!(\ell-s)!}}(\sin\vartheta/2)^{2\ell} e^{im\varphi}$$

$$\times \sum_r \binom{\ell-s}{r}\binom{\ell+s}{r+s-m}(-1)^{\ell-r-s}(\cot\vartheta/2)^{2r+s-m} . \tag{A4.52}$$

Here the sum over r goes over those values for which the binomial coefficients are non-vanishing, this means $\max\{0, m-s\} \le r \le \min\{\ell - s, \ell + m\}$. (Remember $\binom{0}{0} = 1$.) The spin weighted spherical harmonics are defined for $|s| \le \ell$ and $|m| \le \ell$. For each fixed spin s they form a complete set of orthonormal functions on the sphere, so that

$$\int d\Omega_{\mathbf{n}}\, {}_sY^*_{\ell m}(\mathbf{n})\, {}_sY_{\ell' m'}(\mathbf{n}) = \delta_{\ell\ell'}\delta_{mm'} , \tag{A4.53}$$

and

$$\sum_{\ell m} {}_sY^*_{\ell m}(\mathbf{n})\, {}_sY_{\ell m}(\mathbf{n}) = \delta(\varphi - \varphi')\delta(\cos\vartheta - \cos\vartheta') . \tag{A4.54}$$

Since a rotation with angle ψ around the z axis simply rotates the matrix element $D^{(\ell)}_{m\,-s}$ by a factor $e^{-is\psi}$, we also have

$$\bar{D}^{(\ell)}_{m\,-s}(\varphi, \vartheta, \psi) = \sqrt{\frac{4\pi}{2\ell+1}}\, {}_sY_{\ell m}(\vartheta, \varphi)e^{is\psi} . \tag{A4.55}$$

Now let R_1 be the rotation with Euler angles $(\varphi_1, \vartheta_1, 0)$ which rotates $\mathbf{e}_z$ into $\mathbf{n}_1$ and R_2 the rotation with Euler angles $(\varphi_2, \vartheta_2, 0)$ which rotates $\mathbf{e}_z$ into $\mathbf{n}_2$. Let (α, β, γ) be the Euler angles of the rotation $R_1^{-1}R_2$, which first rotates $\mathbf{n}_2$ into $\mathbf{e}_z$ and then $\mathbf{e}_z$ into $\mathbf{n}_1$. We then have

$$D^{(\ell)}_{m\,-s}(\alpha, \beta, \gamma) = \sum_{m'} D^{(\ell)-1}_{mm'}(\varphi_1, \vartheta_1, 0)D^{(\ell)}_{m'\,-s}(\varphi_2, \vartheta_2, 0)$$

$$= \sum_{m'} \bar{D}^{(\ell)}_{m'm}(\varphi_1, \vartheta_1, 0)D^{(\ell)}_{m'\,-s}(\varphi_2, \vartheta_2, 0)$$

$$= \frac{4\pi}{2\ell+1}\sum_{m'} {}_s\bar{Y}_{\ell m'}(\mathbf{n}_2)\, {}_{-m}Y_{\ell m'}(\mathbf{n}_1) . \tag{A4.56}$$

Table A4.3. *The spin-2 spherical harmonics for $\ell = 2$.*

m	${}_2Y_{2m}$
2	$\frac{1}{8}\sqrt{\frac{5}{\pi}}(1-\cos\vartheta)^2 e^{2i\varphi}$
1	$\frac{1}{4}\sqrt{\frac{5}{\pi}}\sin\vartheta(1-\cos\vartheta)e^{i\varphi}$
0	$\frac{3}{4}\sqrt{\frac{5}{6\pi}}\sin^2\vartheta$
-1	$\frac{1}{4}\sqrt{\frac{5}{\pi}}\sin\vartheta(1+\cos\vartheta)e^{-i\varphi}$
-2	$\frac{1}{8}\sqrt{\frac{5}{\pi}}(1+\cos\vartheta)^2 e^{-2i\varphi}$

Using Eq. (A4.55) we find

$$\sqrt{\frac{4\pi}{2\ell+1}}\sum_{m'} {}_sY_{\ell m'}(\vartheta_2,\varphi_2)\,{}_{-m}\bar{Y}_{\ell m'}(\vartheta_1,\varphi_1) = {}_sY_{\ell m}(\beta,\alpha)e^{-is\gamma}\ . \tag{A4.57}$$

This is the generalized addition theorem for spin weighted spherical harmonics.

In analogy to $L_\pm$ as raising and lowering operators for the 'magnetic quantum number' m, we introduce the spin raising and lowering operators $\eth$ and $\eth^*$. They are defined by, $\mu = \cos\vartheta$,

$$\eth\ {}_sY_{\ell m} = \left(s\,\mathrm{ctg}\vartheta - \partial_\vartheta - \frac{i}{\sin\vartheta}\partial_\varphi\right){}_sY_{\ell m} \tag{A4.58}$$

$$= \left(\frac{s\mu}{\sqrt{1-\mu^2}} + \sqrt{1-\mu^2}\partial_\mu - \frac{i}{\sqrt{1-\mu^2}}\partial_\varphi\right){}_sY_{\ell m} \tag{A4.59}$$

$$= -(1-\mu^2)^{\frac{s}{2}}\left(\partial_\vartheta + \frac{i\partial_\varphi}{\sqrt{1-\mu^2}}\right)[(1-\mu^2)^{-s/2}\,{}_sY_{\ell m}],$$

$$\eth^*\ {}_sY_{\ell m} = \left(-s\,\mathrm{ctg}\vartheta - \partial_\vartheta + \frac{i}{\sin\vartheta}\partial_\varphi\right){}_sY_{\ell m} \tag{A4.60}$$

$$= \left(\frac{-s\mu}{\sqrt{1-\mu^2}} + \sqrt{1-\mu^2}\partial_\mu + \frac{i}{\sqrt{1-\mu^2}}\partial_\varphi\right){}_sY_{\ell m} \tag{A4.61}$$

$$= -(1-\mu^2)^{\frac{-s}{2}}\left(\partial_\vartheta - \frac{i\partial_\varphi}{\sqrt{1-\mu^2}}\right)[(1-\mu^2)^{s/2}\,{}_sY_{\ell m}]\ .$$

The interest of these operators is that they allow us to construct the spin weighted spherical harmonics directly from the spin-0 harmonics and, inversely, we can use them to build spin-0 quantities from spin weighted harmonics. One can actually show (e.g., by using Eq. (A4.52)) that

$$\eth\,({}_sY_{\ell m}) = \sqrt{(\ell-s)(\ell+s+1)}\,{}_{s+1}Y_{\ell m}\ , \tag{A4.62}$$

$$\eth^*\,({}_sY_{\ell m}) = -\sqrt{(\ell+s)(\ell-s+1)}\,{}_{s-1}Y_{\ell m}\ , \tag{A4.63}$$

and therefore

$$\eth^2\,({}_{-2}Y_{\ell m}) = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,Y_{\ell m}\ , \tag{A4.64}$$

$$(\eth^*)^2\,({}_{2}Y_{\ell m}) = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,Y_{\ell m}\ . \tag{A4.65}$$

On the other hand, the spin-2 spherical harmonics can be obtained from the ordinary spherical harmonics by acting twice with the differential operators $\eth$ or $\eth^*$,

$$(\eth)^2 Y_{\ell m} = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,{}_{2}Y_{\ell m}\ , \tag{A4.66}$$

$$(\eth^*)^2 Y_{\ell m} = \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\,{}_{-2}Y_{\ell m}\ . \tag{A4.67}$$

To see that $\eth\ {}_sf$ has spin $s+1$ and $\eth^*\ {}_sf$ has spin $s-1$, for an arbitrary component ${}_sf$ with spin weight s, we show that $\eth$ is proportional to a covariant derivative in direction $g^{+-}\mathbf{e}_-$ and correspondingly $\eth^* \propto g^{-+}\nabla_{\mathbf{e}_+}$. Since $\eth^*$ is the adjoint of $-\eth$, it is sufficient if we show the first identity. For $s=0$, using $\mathbf{e}_+ = \frac{1}{\sqrt{2}}(\mathbf{e}_\vartheta - i\,\mathbf{e}_\varphi) = \frac{1}{\sqrt{2}}(\partial_\vartheta - i\,\frac{1}{\sin\vartheta}\partial_\varphi)$, we obtain $\eth^* f = -\sqrt{2}\mathbf{e}_+ f$. For $s \neq 0$ tensor fields we have to compute the Christoffel symbols in order to determine the covariant derivatives. In terms of the helicity basis, the canonical metric on the 2-sphere takes the form $ds^2 = \theta^+\theta^-$, where $\theta^\pm$ denote the 1-forms dual to the vector fields $\mathbf{e}_\pm$ defined by $\theta^+(\mathbf{e}_+) = \theta^-(\mathbf{e}_-) = 1$ and $\theta^-(\mathbf{e}_+) = \theta^+(\mathbf{e}_-) = 0$. Hence

$$\theta^\pm = \frac{1}{\sqrt{2}}\,(d\vartheta \pm i\,\sin\vartheta\,d\varphi)\ .$$

Therefore, the metric components are simply $g_{++} = g_{--} = 0$ and $g_{-+} = g_{+-} = 1$. A careful evaluation of the Christoffel symbols defined by $\nabla_{\mathbf{e}_k}\mathbf{e}_j = \Gamma^i_{kj}\mathbf{e}_i$ in the helicity basis gives[2]

$$\Gamma^+_{-+} = \Gamma^-_{+-} = -\Gamma^+_{++} = -\Gamma^-_{--} = -\frac{1}{\sqrt{2}}\frac{\cos\vartheta}{\sin\vartheta} = -\frac{1}{\sqrt{2}}\mathrm{ctg}\vartheta\ . \tag{A4.68}$$

All other Christoffel symbols vanish. With this we find for the covariant derivatives of the spin-s components of a tensor

$$\begin{aligned} T^{+\cdots+;+} = T^{+\cdots+}_{\ ;-} &= \mathbf{e}_-(T^{+\cdots+}) - \frac{s}{\sqrt{2}}\mathrm{ctg}\vartheta\,T^{+\cdots+}\ , \\ &= \frac{1}{\sqrt{2}}\left(\partial_\vartheta + \frac{i}{\sin\vartheta}\partial_\varphi\right)(T^{+\cdots+}) - \frac{s}{\sqrt{2}}\mathrm{ctg}\vartheta\,T^{+\cdots+}\ , \\ &= \frac{1}{\sqrt{2}}\left[\partial_\vartheta - s\,\mathrm{ctg}\vartheta + \frac{i}{\sin\vartheta}\partial_\varphi\right]T^{+\cdots+} \\ &= \frac{-1}{\sqrt{2}}\eth T^{+\cdots+}\ . \end{aligned}$$

[2] These Christoffel symbols are most easily calculated using the Cartan formalism which can be found in most modern books on general relativity, e.g. Straumann (2004).

In the same way one obtains

$$T^{+\cdots+;-} = \frac{-1}{\sqrt{2}}\eth^* T^{+\cdots+} \,.$$

In other words,

$$\eth = -\sqrt{2}\,\nabla_{\mathbf{e}_-} \text{ and } \quad \eth^* = -\sqrt{2}\,\nabla_{\mathbf{e}_+} \,. \tag{A4.69}$$

Since for an arbitrary rank-s tensor field the component $T^{+\cdots+}$ has helicity s and $T^{+\cdots+;+}$ has helicity $s+1$ while $T^{+\cdots+;-}$ has helicity $s-1$, this shows that $\eth$ and $\eth^*$ are spin raising and lowering operators. Correspondingly one obtains $\frac{-1}{\sqrt{2}}\eth^* T^{-\cdots-} = T^{-\cdots-;-}$ and $\frac{-1}{\sqrt{2}}\eth T^{-\cdots-} = T^{-\cdots-;+}$, which have spin weight $-s-1$ and $-s+1$ respectively.

The spin-2 spherical harmonics are therefore just the doubly covariant derivatives of the usual spherical harmonics,

$${}_{-2}Y_{\ell m} = 2\sqrt{\frac{(\ell-2)!}{(\ell+2)!}}\nabla_{\mathbf{e}_+}\nabla_{\mathbf{e}_+} Y_{\ell m} \,, \tag{A4.70}$$

$${}_{+2}Y_{\ell m} = 2\sqrt{\frac{(\ell-2)!}{(\ell+2)!}}\nabla_{\mathbf{e}_-}\nabla_{\mathbf{e}_-} Y_{\ell m} \,. \tag{A4.71}$$

Finally, we want to interpret the spin-0 quantities $\eth\eth T^{--}$ and $\eth^*\eth^* T^{++}$ of a symmetric traceless spin-2 tensor $T = T^{++}\mathbf{e}_+ \otimes \mathbf{e}_+ + T^{--}\mathbf{e}_- \otimes \mathbf{e}_-$. In Exercise A4.22 we find that for a vector with components V^+ and V^-, the divergence and curl are given by $V^+{}_{;-} + V^-{}_{;+} = V^i{}_{;i} = \operatorname{div} V$ and $V^+{}_{;-} - V^-{}_{;+} = -i\epsilon_{ij}V^{i;j} = -i\,\operatorname{rot} V$. Here ϵ_{ij} is the totally anti-symmetric tensor in two dimensions,

$$\epsilon_{\vartheta\vartheta} = \epsilon_{\varphi\varphi} = 0, \qquad \epsilon_{\vartheta\varphi} = -\epsilon_{\varphi\vartheta} = \sqrt{\det g} = \sin\vartheta \,. \tag{A4.72}$$

In the same way one also finds

$$\begin{aligned}\frac{1}{2}\left(\eth^{*2}T^{++} + \eth^2 T^{--}\right) &= T^{++;--} + T^{--;++} \\ &= T^{ij}{}_{;ij} = \operatorname{div}(\operatorname{div}(T)) \,, \end{aligned} \tag{A4.73}$$

$$\begin{aligned}\frac{1}{2}\left(\eth^{*2}T^{++} - \eth^2 T^{--}\right) &= T^{++;--} - T^{--;++} \\ &= -\epsilon_{ik}\epsilon_{jm}T^{ij;km} = -\operatorname{rot}(\operatorname{rot}(T)) \,. \end{aligned} \tag{A4.74}$$

Hence the sum of the two scalars $T^{++;--}$ and $T^{--;++}$ gives the double divergence while their difference gives the double curl of T. Note that in two dimensions the curl is a (pseudo-)scalar. It is a three-dimensional (pseudo-)vector with a purely radial component which as a field on the tangent space of the sphere has scalar character.

Exercise A4.2.1: Tensor components in the helicity basis
Show that the components T^{ij} of a 2-tensor on the sphere are related to the components in the helicity basis $\mathbf{e}_+$ and $\mathbf{e}_-$ via

$$\begin{aligned} T^{\pm,\pm} &= \frac{1}{2}\left(T^{11} - T^{22} \mp i(T^{12} + T^{21})\right) \,, \\ T^{\pm,\mp} &= \frac{1}{2}\left(T^{11} + T^{22} \pm i(T^{12} - T^{21}\right) \,. \end{aligned}$$

In particular, if T is symmetric $T^{+-} = T^{-+} = \frac{1}{2}\mathrm{tr}\, T$.

Exercise A4.2.2: Divergence and curl in the helicity basis
Show that for a vector field on the sphere given by

$$V = V^+ e_+ + V^- e_- = V^\vartheta \partial_\vartheta + V^\varphi \partial_\varphi \ ,$$

we have

$$V^{+;-} + V^{-;+} = V^\vartheta_{,\vartheta} + V^\varphi_{,\varphi} + \mathrm{ctg}\vartheta\, V^\vartheta = \mathrm{divV} \tag{A4.75}$$

$$\begin{aligned} V^{+;-} - V^{-;+} &= -i\left(\frac{1}{\sin\vartheta} V^\vartheta_{,\varphi} - \sin\vartheta\, V^\varphi_{,\vartheta} - 2\cos\vartheta\, V^\varphi\right) \ , \\ &= \frac{-i}{\sin\vartheta}\left(V^\vartheta_{,\varphi} - (\sin^2\vartheta\, V^\varphi)_{,\vartheta}\right) = -i\,\mathrm{rot}V \ . \end{aligned} \tag{A4.76}$$

Solution
We first calculate the Christoffel symbols with respect to the coordinate basis (ϑ, φ). Using $ds^2 = d\vartheta^2 + \sin^2\vartheta\, d\varphi^2$ quickly gives

$$\Gamma^\vartheta_{\varphi\varphi} = -\cos\vartheta\,\sin\vartheta \ , \quad \Gamma^\varphi_{\varphi\vartheta} = \Gamma^\varphi_{\vartheta\varphi} = \mathrm{ctg}\,\vartheta \ ,$$

and zero for all other Christoffel symbols, so that

$$V^\vartheta_{;\vartheta} = V^\vartheta_{,\vartheta} \ , \tag{A4.77}$$

$$V^\vartheta_{;\varphi} = V^\vartheta_{,\varphi} - \cos\vartheta\,\sin\vartheta\, V^\varphi \ , \tag{A4.78}$$

$$V^\varphi_{;\varphi} = V^\varphi_{,\varphi} + \mathrm{ctg}\vartheta\, V^\vartheta \ , \tag{A4.79}$$

$$V^\varphi_{;\vartheta} = V^\varphi_{,\vartheta} + \mathrm{ctg}\vartheta\, V^\varphi \ . \tag{A4.80}$$

so that

$$\mathrm{div}V = V^\vartheta_{;\vartheta} + V^\varphi_{;\varphi} = V^\vartheta_{,\vartheta} + V^\varphi_{,\varphi} + \mathrm{ctg}\vartheta\, V^\vartheta \ , \tag{A4.81}$$

$$\begin{aligned} \mathrm{rot}V &= \sqrt{\det g}(V^{\vartheta;\varphi} - V^{\varphi;\vartheta}) = \sin\vartheta\left(\frac{1}{\sin^2\vartheta} V^\vartheta_{;\varphi} - V^\varphi_{;\vartheta}\right) \\ &= \frac{1}{\sin\vartheta} V^\vartheta_{,\varphi} - \sin\vartheta\, V^\varphi_{,\vartheta} - 2\cos\vartheta\, V^\varphi \ . \end{aligned} \tag{A4.82}$$

On the other hand, we have

$$V^{+;-} \pm V^{-;+} = V^+_{;+} \pm V^-_{;-} = V^+_{,+} \pm V^-_{,-} + \Gamma^+_{++} V^+ \pm \Gamma^-_{--} V^- \ . \tag{A4.83}$$

Inserting $V^\pm = \frac{1}{\sqrt{2}}(V^\vartheta \pm i\,\sin\vartheta\, V^\varphi)$ and $V^\bullet_{,\pm} = \frac{1}{\sqrt{2}}(\partial_\vartheta \mp \frac{i}{\sin\vartheta}\partial_\varphi)V^\bullet$ together with $\Gamma^+_{++} = \Gamma^-_{--} = \frac{1}{\sqrt{2}}\mathrm{ctg}\,\vartheta$ we find the above result.

Exercise A4.2.3: The Laplacian in the helicity basis

In the helicity basis the Laplacian on the sphere is given by

$$\Delta = g^{ab}\nabla_b\nabla_a = g^{-+}\nabla_-\nabla_+ + g^{+-}\nabla_+\nabla_- = \nabla_-\nabla_+ + \nabla_+\nabla_- = \frac{1}{2}\left(\eth\eth^* + \eth^*\eth\right) .$$

Apply this formula to the spherical harmonics $Y_{\ell m}$ to show that $\Delta Y_{\ell m} = -\ell(\ell+1)Y_{\ell m}$.

A4.3 Bessel functions

A4.3.1 Bessel functions of integer order

The Bessel functions $J_\nu(x)$ and $Y_\nu(x)$ are real solutions to the differential equation

$$x^2\frac{d^2f}{dx^2} + x\frac{df}{dx} + (x^2 - \nu^2)f = 0 . \tag{A4.84}$$

We only consider $\nu \in \mathbb{R}$, hence we may consider $\nu \geq 0$. Actually J_ν and $J_{-\nu}$ satisfy the same differential equation but one defines

$$J_{-\nu} = \cos(\nu\pi)J_\nu - \sin(\nu\pi)Y_\nu .$$

With this definition the Bessel functions are also analytic in the order ν. If $\nu \geq 0$, J_ν is regular at $x = 0$, $J_\nu(x) \propto x^\nu$ for $|x| \ll \nu$ and Y_ν diverges, $Y_\nu(x) \propto x^{-\nu}$ for $|x| \ll \nu$. For large values of $|x|$ both functions oscillate with a period of approximately 2π and decay like $1/\sqrt{x}$. We sometimes also use the Hankel functions defined by

$$H_\nu^{(1)} = J_\nu + iY_\nu , \quad H_\nu^{(2)} = J_\nu - iY_\nu . \tag{A4.85}$$

All of these functions satisfy the recurrence relations

$$F_{\nu-1} + F_{\nu+1} = \frac{2\nu}{x}F_\nu , \tag{A4.86}$$

$$F_{\nu-1} - F_{\nu+1} = 2F_\nu' , \tag{A4.87}$$

$$F_{\nu-1} - \frac{\nu}{x}F_\nu = F_\nu' , \tag{A4.88}$$

$$-F_{\nu+1} + \frac{\nu}{x}F_\nu = F_\nu' . \tag{A4.89}$$

The Bessel functions are well defined in the complex plane (with suitably chosen cuts) even for complex values ν. The Bessel functions J_n, $n \in \mathbb{N}$ can be represented as the integral

$$J_n(x) = \frac{(-i)^n}{\pi}\int_0^\pi e^{ix\cos\theta}\cos(n\theta)\,d\theta . \tag{A4.90}$$

With this one finds the useful expansion

$$e^{iy\cos\phi} = J_0(y) + 2\sum_{n=1}^{\infty} i^n J_n(y)\cos(n\phi) = \sum_{n=-\infty}^{\infty} i^n J_n(y)e^{in\phi} . \tag{A4.91}$$

We shall also employ the modified Bessel functions which are defined by

$$I_\nu(x) = (-i)^\nu J_\nu(ix) , \tag{A4.92}$$

$$K_\nu(x) = \frac{i\pi}{2}(i)^\nu H_\nu^{(1)}(ix) = -\frac{i\pi}{2}(-i)^\nu H_\nu^{(1)}(-ix) . \tag{A4.93}$$

A4.3.2 Spherical Bessel functions

The spherical Bessel (Hankel) functions are Bessel (Hankel) functions of half-integer order,

$$j_n(x) = \sqrt{\frac{\pi}{2x}} J_{n+1/2}(x) \,, \tag{A4.94}$$

$$y_n(x) = \sqrt{\frac{\pi}{2x}} Y_{n+1/2}(x) \,, \tag{A4.95}$$

$$h_n^{(1)} = j_n + i y_n \,, \tag{A4.96}$$

$$h_n^{(2)} = j_n - i y_n \,. \tag{A4.97}$$

They are solutions of the differential equation

$$x^2 \frac{d^2 f}{dx^2} + 2x \frac{df}{dx} + (x^2 - n(n+1)) f = 0 \,. \tag{A4.98}$$

The spherical Bessel/Hankel functions satisfy the recurrence relations

$$\frac{f_n}{x} = \frac{1}{2n+1} \left(f_{n-1} + f_{n+1} \right) \,, \tag{A4.99}$$

$$f_n' = \frac{1}{2n+1} \left(n f_{n-1} - (n+1) f_{n+1} \right) \,. \tag{A4.100}$$

Expressing the three-dimensional Laplace operator in polar coordinates, and observing that the spherical part of the Laplacian applied to a spherical harmonic function gives $\Delta_{\vartheta\varphi} Y_{\ell m} = -\ell(\ell+1) Y_{\ell m}$ we find that $j_\ell(rk) Y_{\ell m}(\hat{\mathbf{x}})$ as well as $y_\ell(rk) Y_{\ell m}(\hat{\mathbf{x}})$ is a solution of

$$\left(\Delta + k^2 \right) f = 0$$

for arbitrary values of ℓ and $-\ell \le m \le \ell$. Only the j_ℓs are regular at $r = 0$. On the other hand, the exponential function, $e^{i\mathbf{x}\cdot\mathbf{k}}$, for arbitrary $\mathbf{k}$ with modulus $|\mathbf{k}| = k$ also solves this equation. Since the spherical harmonics form a complete system of functions on the sphere, there must exist an expansion

$$e^{i\mathbf{x}\cdot\mathbf{k}} = e^{ikr\hat{\mathbf{x}}\cdot\hat{\mathbf{k}}} = \sum_{\ell m} c_{\ell m} j_{\ell m}(rk) Y_{\ell m}(\hat{\mathbf{x}}) \,.$$

This represents the decomposition of the exponential into its contributions of orbital angular momentum ℓ. To determine the coefficients $c_{\ell m}$, we choose the z-axis in the direction of $\mathbf{k}$, so that the function is independent of φ and only the terms with $m = 0$ contribute. Setting $\mu = \cos\vartheta$ this yields

$$e^{irk\mu} = \sum_\ell c_{\ell 0} j_\ell(rk) Y_{\ell 0}(\mu) = \sum_\ell c_\ell j_\ell(rk) P_\ell(\mu) \,,$$

where we have set $c_\ell = \frac{2\ell+1}{4\pi} c_{\ell 0}$ and made use of Eq. (A4.35). The coefficients c_ℓ are now obtained by taking the nth derivative with respect to rk, multiplying with P_n and integrating over μ (for more details see Arfken & Weber, 2001). One finds $c_\ell = i^\ell (2\ell+1)$ so that

$$e^{i\mathbf{k}\cdot\mathbf{n}r} = \sum_{\ell=0}^{\infty} (2\ell+1) i^\ell j_\ell(kr) P_\ell(\mu) \,. \tag{A4.101}$$

An important Bessel function integral which we use especially in Chapter 4 is

$$I(p,\ell) \equiv \frac{2}{\pi}\int_0^\infty dx\, x^p j_\ell^2(x) = \int_0^\infty x^{p-1} J^2_{\ell+1/2}(x)$$
$$= \frac{\Gamma(1-p)\Gamma\left(\ell+\frac{p+1}{2}\right)}{2^{1-p}\Gamma^2(1-p/2)\Gamma\left(\ell+\frac{3-p}{2}\right)}\,. \tag{A4.102}$$

Here Γ denotes the Gamma function, i.e. $\Gamma(n) = (n-1)!$ for positive integers. The integral is well defined for $1 > p > -2\ell - 1$. Of special interest is the case $p = -1$ which occurs in the calculation of the CMB spectrum for scale-invariant fluctuations,

$$I(-1,\ell) = \frac{1}{\pi\,\ell(\ell+1)}\,. \tag{A4.103}$$

Appendix 5

Entropy production and heat flux

Here we show that the perturbation variable Γ defined in Eq. (2.82) is related to the divergence of the entropy flux. We consider a system which deviates slightly from thermal equilibrium.

A5.1 Thermal equilibrium

We first recollect some important relations in thermal equilibrium. We consider an arbitrary mix of different (relativistic and non-relativistic) particles which may or may not be conserved. The only *total* thermodynamical quantities then are temperature T, entropy S, energy E, pressure P and volume V. We shall also use the densities $s = dS/dV$ and $\rho = dE/dV$. Certain conserved species may have a chemical potential, but we are not interested in this 'fine structure' here. The corresponding treatment for one conserved particle species can be found in Straumann (1984), Appendix B.

We start with the Gibbs relation

$$T\,dS = dE + P\,dV\,, \quad \text{or} \quad T\frac{dS}{dV} = Ts = \rho + P\,. \tag{A5.1}$$

S and E are extensive quantities. Locally they are simply given by $S = sV$ and $E = \rho V$. Inserting this in the Gibbs relation we obtain

$$TV\,ds + Ts\,dV = V\,d\rho + \rho\,dV + P\,dV \quad \text{hence} \quad T\,ds = d\rho\,. \tag{A5.2}$$

Defining the entropy 4-velocity field by U^μ. The entropy flux is then given by $S^\mu = sU^\mu = T^{-1}(\rho + P)U^\mu$. In thermal equilibrium the entropy velocity coincides with the energy flux $u^\mu = U^\mu$, so that $T^{\mu\nu}U_\mu = -\rho U^\nu$. In thermal equilibrium we therefore have

$$S^\mu = -\frac{1}{T}U_\nu T^{\mu\nu} + \frac{P}{T}U^\mu\,. \tag{A5.3}$$

In a FL background $(U^\mu) = (u^\mu) = a^{-1}(1, \mathbf{0})$ with $U^\mu{}_{;\mu} = 3\dot{a}/a^2$, so that entropy conservation becomes $0 = S^\mu{}_{;\mu} = a^{-1}\dot{s} + 3(\dot{a}/a^2)s$ which results in the well known law of adiabatic expansion, $\dot{s} = -3(\dot{a}/a)s$. Furthermore, with Eq. (A5.2) small variations of the entropy flux at fixed velocity field U^μ are given by

$$dS^\mu = U^\mu\,ds = \frac{1}{T}U^\mu\,d\rho = -\frac{1}{T}U_\nu\,dT^{\mu\nu}\,. \tag{A5.4}$$

A5.2 Small departures from thermal equilibrium

We now proceed to the study of small deviations from equilibrium. There is some arbitrariness in fitting the actual state with an equilibrium state plus small deviations. Following Israel & Stewart (1980), we approximate the actual state with the thermal equilibrium at the same energy density ρ and entropy velocity field U^μ. We neglect all second-order quantities, taking into account only first-order deviations from thermal equilibrium and/or from the FL background. We specify the deviation of the energy–momentum tensor from thermal equilibrium, $\delta T^{\mu\nu}$, by the following ansatz

$$T^{\mu\nu} = (\rho + P_{\rm eq})U^\mu U^\nu + P_{\rm eq} g^{\mu\nu} + \delta T^{\mu\nu} \ . \tag{A5.5}$$

Here $P_{\rm eq}$ is the pressure of the thermal equilibrium state with energy density ρ. Setting $\rho = \bar\rho + \delta\rho$ we therefore have $P_{\rm eq} = \bar P + \delta P$ with $\delta P = c_s^2 \delta\rho$.

On the other hand, the energy flux 4-velocity u^μ is defined by (2.62) as the time-like eigenvector of the energy–momentum tensor and $T^{\mu\nu}$ can also be written in the form

$$T^{\mu\nu} = (\rho + P)u^\mu u^\nu + P g^{\mu\nu} + \Pi^{\mu\nu} = \rho u^\mu u^\nu + \tau^{\mu\nu} \ , \tag{A5.6}$$

where τ is given in Eq. (2.66),

$$\tau^{\mu\nu} = P(u^\mu u^\nu + g^{\mu\nu}) + \Pi^{\mu\nu} \ , \quad \Pi^\lambda_\lambda = 0 \ . \tag{A5.7}$$

The tensor $\Pi^{\mu\nu}$ is orthogonal to u^μ, $\Pi^{\mu\nu}u_\nu = 0$. Defining Q^μ by

$$u^\mu = U^\mu + Q^\mu \ ,$$

we can rewrite (A5.6) in the following manner:

$$\begin{aligned} T^{\mu\nu} &= (\rho + P)U^\mu U^\nu + P g^{\mu\nu} + U^\mu q^\nu + U^\nu q^\mu + \Pi^{\mu\nu} \\ &= (\rho + P_{\rm eq})U^\mu U^\nu + P_{\rm eq} g^{\mu\nu} + U^\mu q^\nu + U^\nu q^\mu \\ &\quad + (P - P_{\rm eq})(U^\mu U^\nu - g^{\mu\nu}) + \Pi^{\mu\nu} \ , \end{aligned} \tag{A5.8}$$

where we have introduced

$$q^\mu = (\rho + p)Q^\mu \ . \tag{A5.9}$$

$\Pi_{\mu\nu}$, $\delta T_{\mu\nu}$, Q^μ and therefore also q^μ vanish in the background, they are of first order. Since $u^2 = U^2 = -1$, we have to first order $q \cdot U = 0$, $q \cdot u = 0$.

Identifying $\delta T^{\mu\nu}$ by comparing Eq. (A5.8) with the definition given in Eq. (A5.5), we obtain to first order

$$\delta T^{\mu\nu} = U^\mu q^\nu + U^\nu q^\mu + (P - P_{\rm eq})(U^\mu U^\nu - g^{\mu\nu}) + \Pi^{\mu\nu} \ , \tag{A5.10}$$

and

$$\delta T^{\mu\nu} U_\mu = -q^\nu - \Pi^{\mu\nu} Q_\mu = -q^\nu \ ,$$

since $\Pi^{\mu\nu}$ and Q_μ are both first order and normal to U^μ. With Eq. (A5.4) the perturbed entropy flux $S^\mu = S^\mu_{\rm eq} + \delta S^\mu$ becomes

$$S^\mu = sU^\mu - \frac{1}{T}\delta T^{\mu\nu}U_\nu = sU^\mu + \frac{1}{T}q^\mu \ . \tag{A5.11}$$

This equation shows that q^μ represents the heat flux.

From $P = \bar P(1 + \pi_L)$ and $P_{\rm eq} = \bar P(1 + \frac{c_s^2}{w}\delta)$, $\delta = \delta\rho/\bar\rho$ we find with Eq. (2.82)

$$P - P_{\rm eq} = \bar P \left(\pi_L - \frac{c_s^2}{w}\delta\right) = \bar P \Gamma \ . \tag{A5.12}$$

Taking the divergence of Eq. (A5.11) we find

$$S^{\mu}{}_{;\mu} = s_{,\mu}U^{\mu} + sU^{\mu}{}_{;\mu} + \frac{T_{,\mu}}{T^2}\delta T^{\mu\nu}U_{\nu} - \frac{1}{T}\delta T^{\mu\nu}{}_{;\mu}U^{\nu} - \frac{1}{T}\delta T^{\mu\nu}U_{(\nu;\mu)} \,, \tag{A5.13}$$

where $(\nu;\mu)$ denotes symmetrization, $U_{(\nu;\mu)} = \frac{1}{2}(U_{\mu;\nu} + U_{\nu;\mu})$. In the last term we have used the fact that $\delta T^{\mu\nu}$ is symmetric. To evaluate the third term on the r.h.s. we make use of energy–momentum conservation in the form

$$0 = U_{\nu}T^{\mu\nu}{}_{;\mu} = U_{\nu}[(\rho + P_{\rm eq})U^{\mu}U^{\nu} + P_{\rm eq}g^{\mu\nu}]_{;\mu} + U_{\nu}\delta T^{\mu\nu}{}_{;\mu} \,.$$

Expanding the derivative of the square bracket leads to

$$(\rho + P_{\rm eq})U^{\mu}{}_{;\mu} + \rho_{,\mu}U^{\mu} = \delta T^{\mu\nu}{}_{;\mu}U_{\nu} \,. \tag{A5.14}$$

With Eq. (A5.1), the first term on the l.h.s. of Eq. (A5.14) canals the second term on the r.h.s. of Eq. (A5.13), and Eq. (A5.2) implies $s_{,\mu} = T^{-1}\rho_{,\mu}$, so that the second term of Eq. (A5.14) canals the first term in Eq. (A5.13). The fourth term on the r.h.s. of Eq. (A5.13) therefore cancels the first two and we are left with

$$S^{\mu}{}_{;\mu} = \frac{T_{,\mu}}{T^2}q^{\mu} - \frac{1}{T}\delta T^{\mu\nu}U_{(\nu;\mu)} \,. \tag{A5.15}$$

To evaluate $\delta T^{\mu\nu}U_{(\nu;\mu)}$ we define the projector $h_{\mu}{}^{\nu}$ onto the three-dimensional subspace of tangent space normal to U^{μ}:

$$h_{\mu}{}^{\nu} = U^{\mu}U_{\nu} + \delta_{\mu}{}^{\nu} \,,$$

and the acceleration

$$a_{\mu} = U^{\nu}U_{\mu;\nu} \,.$$

With this it is easy to show that

$$U_{\mu;\nu} = h_{\mu}{}^{\alpha}h_{\nu}{}^{\beta}U_{\alpha;\beta} - a_{\mu}U_{\nu} \,.$$

Defining the expansion tensor $\theta_{\mu\nu} = \theta_{\nu\mu} = h_{\mu}{}^{\alpha}h_{\nu}{}^{\beta}U_{(\alpha;\beta)}$, Eq. (A5.15) now becomes

$$S^{\mu}{}_{;\mu} = -\frac{1}{T}\left(\frac{T_{,\mu}}{T} + a_{\mu}\right)q^{\mu} - \frac{1}{T}\,\delta T^{\mu\nu}\theta_{\nu\mu} \,. \tag{A5.16}$$

The acceleration a_{μ} is of first order, and to lowest order $T_{,\mu}/T \propto U_{\mu}$ so that $(\frac{T_{,\mu}}{T} + a_{\mu})q^{\mu}$ vanishes to first order. Furthermore, $\delta T^{\mu\nu}$ is of first order. To determine the divergence of the entropy flux to first order, it is therefore sufficient to determine $\theta_{\nu\mu}$ to zeroth order. But to zeroth order

$$h_0^0 = h_i^0 = h_0^i = 0 \quad \text{and} \quad h_i^j = \delta_i^j \,.$$

Furthermore, $(U_{\mu}) = -a(1, 0)$ so that

$$U_{i;j} = -\dot{a}\delta_{ij} \quad \text{and} \quad U_{0;0} = U_{0;j} = U_{i;0} = 0 \,.$$

Inserting this in the definition of $\theta_{\mu\nu}$, we obtain

$$\theta_{ij} = \frac{\dot{a}}{a^2}g_{ij} \quad \text{and} \quad \theta_{00} = \theta_{i0} = \theta_{0j} = 0 \,. \tag{A5.17}$$

With Eq. (A5.16) the divergence of the entropy flux then becomes to first order

$$S^{\mu}{}_{;\mu} = \frac{1}{T}\,\delta T^{\mu\nu}\theta_{\nu\mu} = \frac{1}{T}\,\delta T^{ij}\theta_{ij} = 3\frac{\dot{a}}{a^2}\frac{P - P_{\rm eq}}{T} = 3\frac{\dot{a}}{a^2}\frac{\bar{P}}{T}\Gamma\ . \tag{A5.18}$$

For the last equals sign we have used Eq. (A5.12). This demonstrates that the entropy production rate is proportional to Γ.

A5.3 Phenomenological coefficients

Finally, for completeness, we want to introduce the heat conductivity coefficient as well as bulk and shear viscosities. We now no longer assume the energy–momentum tensor to be nearly homogeneous and isotropic, but we allow only for small departures from thermal equilibrium which we take into account to first order only. It is then easy to check that Eq. (A5.16) is still valid. We now define $\eta^{\mu\nu}$ by

$$\delta T^{\mu\nu} = \eta^{\mu\nu} + U^{\mu}q^{\nu} + q^{\mu}U^{\nu}, \quad \eta \equiv \eta^{\mu}_{\mu}\ . \tag{A5.19}$$

From Eq. (A5.10) we have $U_{\mu}\eta^{\mu\nu} = 0$, hence $h^{\mu}_{\alpha}h^{\nu}_{\beta}\eta^{\alpha\beta} = h^{\mu}_{\alpha}\eta^{\alpha\nu} = \eta^{\mu\nu}$. Therefore $\eta^{\mu\nu}$ 'lies' in the hypersurface of tangent space normal to U^{μ},

$$S^{\mu}{}_{;\mu} = -\frac{1}{T^2}q^{\mu}h^{\nu}_{\mu}\left(T_{,\nu} + Ta_{\nu}\right) - \frac{1}{T}\eta^{\mu\nu}\theta_{\mu\nu}\ . \tag{A5.20}$$

Let us also introduce the traceless part of $\eta^{\mu\nu}$,

$$\hat{\eta}^{\mu\nu} = \eta^{\mu\nu} - \frac{1}{3}\eta\, h^{\mu\nu}\ .$$

With the shear tensor defined by

$$\sigma_{\mu\nu} = \theta_{\nu\mu} - \frac{1}{3}\theta h_{\nu\mu}\ , \quad \theta = \theta^{\mu}_{\mu}\ ,$$

Eq. (A5.16) can now be written as

$$S^{\mu}{}_{;\mu} = -\frac{1}{T^2}q^{\mu}h^{\nu}_{\mu}\left(T_{,\mu} + Ta_{\mu}\right) - \frac{1}{T}\hat{\eta}^{\mu\nu}\sigma_{\mu\nu} - \frac{1}{3T}\theta\eta\ . \tag{A5.21}$$

The three terms in Eq. (A5.21) are independent. The second law of thermodynamics, $S^{\mu}{}_{;\mu} \geq 0$, requires that each of them be non-negative. This is only possble if

$$q^{\mu} = -\chi h^{\mu\nu}\left(T_{,\nu} + Ta_{\nu}\right)\ , \quad \chi \geq 0 \tag{A5.22}$$

$$\hat{\eta}^{\mu\nu} = -2\zeta\sigma^{\mu\nu}\ , \quad \zeta \geq 0 \tag{A5.23}$$

$$\eta = -3\xi\theta\ , \quad \xi \geq 0\ . \tag{A5.24}$$

The quantities ζ and ξ are called the shear and bulk viscosity respectively and χ is the heat conductivity coefficient. For small deviations from thermal equilibrium, the second law of thermodynamics requires that $\delta T^{\mu\nu}$ be fully determined by these three coefficients and be of the form

$$\delta T^{\mu\nu} = -\chi\left[U^{\mu}h^{\nu\lambda}\left(T_{,\lambda} + Ta_{\lambda}\right) + U^{\nu}h^{\mu\lambda}\left(T_{,\lambda} + Ta_{\lambda}\right)\right] - 2\zeta\sigma^{\mu\nu} - \xi\theta h^{\mu\nu}\ . \tag{A5.25}$$

Eq. (A5.22) is the relativistic generalization of Fourier's law, $\mathbf{q} = -\chi \nabla T$. In terms of the phenomenological coefficients χ, ζ and ξ the entropy production is given by

$$S^{\mu}{}_{;\mu} = \frac{1}{\chi T^2} q^{\mu} q_{\mu} + \frac{2\zeta}{T} \sigma^{\mu\nu} \sigma_{\mu\nu} + \frac{\xi}{T} \theta^2 \geq 0 \,. \tag{A5.26}$$

Each of these contributions is non-negative since q^{μ} is a spatial vector and $\sigma^{\mu\nu}$ is a spatial tensor. They are understood as entropy production due to heat flux, shear viscosity and bulk viscosity respectively. Only the third term, the bulk viscosity, contributes to first order in deviations from homogeneity and isotropy, since in this case q^{μ} and $\sigma^{\mu\nu}$ are small.

Appendix 6

Mixtures

In this appendix we derive Eqs. (2.136) and (2.137). Let us first recall the definitions of the difference variables,

$$S_{\alpha\beta} = \left[\frac{D_{g\alpha}}{1+w_\alpha} - \frac{D_{g\beta}}{1+w_\beta}\right] , \tag{A6.1}$$

$$V_{\alpha\beta} = V_\alpha - V_\beta , \tag{A6.2}$$

$$\Gamma_{\alpha\beta} = \frac{w_\alpha}{1+w_\alpha}\Gamma_\alpha - \frac{w_\beta}{1+w_\beta}\Gamma_\beta , \tag{A6.3}$$

$$\Pi_{\alpha\beta} = \frac{w_\alpha}{1+w_\alpha}\Pi_\alpha - \frac{w_\beta}{1+w_\beta}\Pi_\beta . \tag{A6.4}$$

We now calculate the derivative of one of the terms in $S_{\alpha\beta}$.

$$\begin{aligned}\left(\frac{D_{g\alpha}}{1+w_\alpha}\right)^{\cdot} &= \frac{\dot{D}_{g\alpha}}{1+w_\alpha} - \frac{\dot{w}_\alpha}{1+w_\alpha}\frac{D_{g\alpha}}{1+w_\alpha} \\ &= (1+w_\alpha)^{-1}\left[\dot{D}_{g\alpha} + 3\mathcal{H}(c_\alpha^2 - w_\alpha)D_{g\alpha}\right] \\ &= -kV_\alpha - 3\mathcal{H}\frac{w_\alpha}{1+w_\alpha}\Gamma_\alpha .\end{aligned} \tag{A6.5}$$

For the second equals sign we have used $\dot{w}_\alpha = -3\mathcal{H}(c_\alpha^2 - w_\alpha)(1+w_\alpha)$ and for the last equals sign we have inserted Eq. (2.130). With the definition (A6.1) we now simply obtain Eq. (2.136):

$$\dot{S}_{\alpha\beta} = -kV_{\alpha\beta} - 3\mathcal{H}\Gamma_{\alpha\beta} . \tag{A6.6}$$

To find a differential equation for $V_{\alpha\beta}$ we first take the difference of Eq. (2.130) for the component α with the same equation for component β.
This leads to

$$\begin{aligned}\dot{V}_{\alpha\beta} + \mathcal{H}V_{\alpha\beta} - 3\mathcal{H}(c_\alpha^2 V_\alpha - c_\beta^2 V_\beta) = 3k(c_\alpha^2 - c_\beta^2)\Phi + kc_\alpha^2\frac{D_{g\alpha}}{1+w_\alpha} \\ -kc_\beta^2\frac{D_{g\beta}}{1+w_\beta} + k\Gamma_{\alpha\beta} - \frac{2k}{3}\left(1 - \frac{3K}{k^2}\right)\Pi_{\alpha\beta} .\end{aligned} \tag{A6.7}$$

We now write

$$\begin{aligned}\frac{D_{g\alpha}}{1+w_\alpha} &= \frac{D_g}{1+w} + \frac{D_{g\alpha}}{1+w_\alpha} - \frac{D_g}{1+w} \\ &= \frac{D_g}{1+w} + \sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P}\left(\frac{D_{g\alpha}}{1+w_\alpha} - \frac{D_{g\gamma}}{1+w_\gamma}\right), \\ &= \frac{D_g}{1+w} + \sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P} S_{\alpha\gamma}\,. \end{aligned} \tag{A6.8}$$

Inserting this and $\frac{kD_g}{1+w} = \frac{kD}{1+w} - 3k\Phi - 3\mathcal{H}V$ in Eq. (A6.7) yields

$$\begin{aligned}\dot{V}_{\alpha\beta} + \mathcal{H}V_{\alpha\beta} - 3\mathcal{H}\big[c_\alpha^2(V_\alpha - V) - c_\beta^2(V_\beta - V)\big] &= \frac{k\big(c_\alpha^2 - c_\beta^2\big)}{1+w} D \\ + k\sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P}\big(c_\alpha^2 S_{\alpha\gamma} - c_\beta^2 S_{\beta\gamma}\big) + k\Gamma_{\alpha\beta} - \frac{2k}{3}\left(1 - \frac{3K}{k^2}\right)\Pi_{\alpha\beta}\,. \end{aligned} \tag{A6.9}$$

Note also that $V_\alpha - V = \sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P}(V_\alpha - V_\gamma) = \sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P} V_{\alpha\gamma}$. Furthermore, $(\rho_\gamma + P_\gamma)S_{\alpha\gamma} = \frac{1}{2}(\rho_\gamma + P_\gamma)[S_{\alpha\gamma} + S_{\beta\gamma}] + \frac{1}{2}(\rho_\gamma + P_\gamma)[S_{\alpha\gamma} - S_{\beta\gamma}]$. Using $S_{\alpha\gamma} - S_{\beta\gamma} = S_{\alpha\beta}$ this gives

$$\sum_\gamma (\rho_\gamma + P_\gamma)S_{\alpha\gamma} = \frac{1}{2}\sum_\gamma (\rho_\gamma + P_\gamma)[S_{\alpha\gamma} + S_{\beta\gamma}] + \frac{1}{2}(\rho + P)S_{\alpha\beta}\,.$$

The same identity holds with $S_{..}$ replaced by $V_{..}$. Inserting this in Eq. (A6.9) we finally obtain Eq. (2.137):

$$\begin{aligned}\dot{V}_{\alpha\beta} + \mathcal{H}V_{\alpha\beta} - \frac{3}{2}\mathcal{H}\big(c_\alpha^2 + c_\beta^2\big)V_{\alpha\beta} - \frac{3}{2}\mathcal{H}\big(c_\alpha^2 - c_\beta^2\big)\sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P}\big(V_{\alpha\gamma} + V_{\beta\gamma}\big) \\ = k\left[\frac{c_\alpha^2 - c_\beta^2}{1+w} D + \frac{c_\alpha^2 + c_\beta^2}{2} S_{\alpha\beta} + \frac{c_\alpha^2 - c_\beta^2}{2}\sum_\gamma \frac{\rho_\gamma + P_\gamma}{\rho + P}\big(S_{\alpha\gamma} + S_{\beta\gamma}\big)\right. \\ \left. + \Gamma_{\alpha\beta} - \frac{3}{2}\left(1 - \frac{3K}{k^2}\right)\Pi_{\alpha\beta}\right]. \end{aligned} \tag{A6.10}$$

Appendix 7

Statistical utensils

A7.1 Gaussian random variables

A7.1.1 Introduction

A random variable is a real function X on a probability space $(\Omega, d\mu)$. The set Ω is a measurable space with normalized measure μ, i.e., $\int_\Omega d\mu = 1$. The integral

$$\int_\Omega X \, d\mu = \langle X \rangle \ ,$$

is called the expectation value or simply the mean of X.

$$\int_\Omega (X - \langle X \rangle)^2 \, d\mu = \langle X^2 \rangle - \langle X \rangle^2 \ ,$$

is called the variance of X and its (positive) square root is the 'standard deviation'. If $\langle X \rangle = 0$, we call X a fluctuation. We are mainly interested in fluctuations. We sometimes call Ω the 'space of realizations' or the 'ensemble'. A random variable is strongly continuous if the derivative of its probability distribution $d\mu/dX \equiv p$ is an integrable function[1] on $\mathbb{R}$. Then we can write

$$\langle X \rangle = \int_\Omega X \, d\mu = \int_{\mathbb{R}} x p(x) \, dx \ . \tag{A7.1}$$

The probability distribution satisfies the normalization condition

$$\int_{\mathbb{R}} p(x) \, dx = 1 \ .$$

The distribution function $p(x)$, also called the probability density, fully determines the random variable.

Definition A7.1: *A random variable with probability distribution*

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-x_0)^2/2\sigma^2} \ ,$$

is called a Gaussian random variable (normal distribution) with mean x_0 and variance σ^2.

[1] In the more general case, p is a distribution in the sense of Schwartz, i.e., a functional on some space of functions on $\mathbb{R}$.

The Gaussian distribution with mean 0 and variance 1 *is called the standard normal distribution.*

The cumulants of a random variable with mean x_0 are defined by

$$V_n \equiv \int_{-\infty}^{\infty} (x - x_0)^n \, p(x) \, dx \ . \tag{A7.2}$$

The cumulants of a Gaussian random variable are given by

$$V_0 = 1, \, V_n = \begin{cases} 0 & \text{if } n \quad \text{is odd} \\ \sigma^n (n-1)!! & \text{if } n > 0, \quad \text{is even} \ , \end{cases} \tag{A7.3}$$

where the double factorial of a number is defined by $m!! = m(m-2)(m-4)\cdots$. This statement is evident for n odd. The proof of the even case is easiest done by induction and is left as an exercise.

A7.1.2 The central limit theorem

Let X_i be independent random variables with means x_i and variances σ_i. Independence means that $\langle (X_i - x_i)(X_j - x_j) \rangle = \delta_{ij} \sigma_i^2$. Then the sums

$$S_n = \frac{\sum_{i=1}^n X_i - x_i}{\sqrt{n} \sum_{i=1}^n \sigma_i} \ ,$$

converge (weakly) to the standard normal distribution. A proof of this important theorem can be found in most texts on probability theory.[2]

In physics it means that an experimental error which comes from many independent sources is often close to Gaussian.

For the CMB its main relevance is that the observed C_ℓs, which are given by the average $C_\ell^o = \frac{1}{2\ell+1} \sum_{-\ell}^{\ell} |a_{\ell m}|^2$ even though by themselves not Gaussian, tend to Gaussian variables with mean C_ℓ and variance C_ℓ^2/ℓ for large ℓ. As we have seen in Section 6.4.2, the variance of the variable $|a_{\ell m}|^2$ is $2C_\ell^2$, therefore the central limit theorem implies that

$$\frac{\sqrt{2}}{\sqrt{(2\ell+1)} C_\ell} \sum_{-\ell}^{\ell} (|a_{\ell m}|^2 - C_\ell) \ ,$$

converges to the standard normal distribution. Hence C_ℓ^o converges to a Gaussian distribution with mean C_ℓ and variance C_ℓ^2/ℓ which becomes small with increasing ℓ.

A7.1.3 A collection of Gaussian random variables

A collection $X_1 \cdots X_N$ of random variables is called Gaussian if their joint probability density is given by

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det(C)}} \exp\left(-\frac{1}{2} \mathbf{x}^T C^{-1} \mathbf{x} \right) \ , \qquad \mathbf{x} \in \mathbb{R}^N \ , \tag{A7.4}$$

where C is a real, positive-definite, symmetric $N \times N$ matrix.

[2] There is a technical condition which has to be fulfilled for the theorem to hold. It ensures that the series is not dominated by one (or a small number) of variables. This is certainly fulfilled if the variables X_i are equally distributed. But the central limit theorem applies in much more general cases.

First we show that

$$\langle X_i X_j \rangle \equiv \frac{1}{\sqrt{(2\pi)^N \det(C)}} \int x_i x_j \exp\left(-\frac{1}{2}\mathbf{x}^T C^{-1}\mathbf{x}\right) dx^N = C_{ij} \ . \tag{A7.5}$$

To prove this, we use the fact that symmetric matrices can be diagonalized. In other words, there exists an orthogonal matrix S, $SS^T = 1$ so that

$$S^T C^{-1} S = D = \begin{pmatrix} \lambda_1 & 0 & \cdots & \\ 0 & \lambda_2 & 0 & \cdots \\ & & \vdots & \\ 0 & \cdots & 0 & \lambda_N \end{pmatrix} . \tag{A7.6}$$

Since $|\det S| = 1$, $1/\det C = \det D = \lambda_1\lambda_2\cdots\lambda_N$ and for $y = Sx$ we have $d^N x = d^N y$. This yields

$$\begin{aligned} \langle X_i X_j \rangle &= S^T_{ik} S^T_{jl} \frac{\sqrt{\Pi_{n=1}^N \lambda_n}}{(2\pi)^{N/2}} \int y_k y_l \exp\left(-\frac{1}{2}\sum_{n=1}^N y_n^2 \lambda_n\right) dx^N \\ &= S^T_{ik} (\lambda_k)^{-1} \delta_{kl} S_{lj} = (S^T D^{-1} S)_{ij} = C_{ij} \ . \end{aligned} \tag{A7.7}$$

Here the sum over repeated indices is understood and we have made use of Eq. (A7.3) for the case $n = 1$ if $k \neq l$ and $n = 2$ if $k = l$. For obvious reasons, the matrix C is called the correlation matrix.

It is easy to see that arbitrary linear combinations of a collection of Gaussian random variables result again in a collection of Gaussian random variables, whereas powers of Gaussian random variables are not Gaussian. Actually, the sum of the squares of n independent Gaussian random variables with the same distribution results in a χ^2-distributed variable with n degrees of freedom.

A7.1.4 Wick's theorem

Wick's theorem provides a general formula for the n-point correlator of a collection of Gaussian variables. In this sense it is a generalization of Eq. (A7.3). It is clear that for n odd the result vanishes. For even ns we obtain the n-point correlator by summing all the possible products of 2-point correlators made from the variables $X_{i_1}, \ldots, X_{i_n}$,

$$\langle X_{i_1} \ldots X_{i_{2n}} \rangle = \sum_{\{j_1,\ldots j_{2n}\}=\{i_1,\ldots i_{2n}\}} C_{j_1\, j_2} \cdots C_{j_{2n-1}\, j_{2n}} \ . \tag{A7.8}$$

Here the sum is not over all permutations, but only over those which give rise to different pairs. Since $C_{ij} = C_{ji}$ we could also simply sum over all permutations of $(i_1, \ldots, i_{2n})$ and divide by $2^n n!$, since for each collection into pairs, there are $2^n n!$ permutations which give rise to the same pairs. The factor 2^n stems from the fact that in each of the pairs we can interchange the factors and $n!$ permutations simply interchange some of the pairs. For example, $2n = 4$ admits $4!/2^2 2 = 3$ different pairings, namely $\langle X_1 X_2\rangle\langle X_3 X_4\rangle$, $\langle X_1 X_3\rangle\langle X_2 X_4\rangle$ and $\langle X_1 X_4\rangle\langle X_3 X_2\rangle$.

This theorem is extremely important not only in probability theory but also in field theory. It means that for a collection of Gaussian random variables, all n-point correlators are determined by the 2-point correlator alone. Since its proof is not easily found in texts on probability but more often in the infinite-dimensional context of field theory where it is more complicated, we present it here for those who are interested.

Proof: We consider only the case of a diagonal covariance matrix C. The general case can then be obtained by diagonalization of C in the same way as in Eq. (A7.7). We show Eq. (A7.8) by induction. The case $n = 1$ is nothing other than Eq. (A7.7). For the step from n to $n + 1$ we use the fact that for an exponentially decaying function $\int_{-\infty}^{\infty} dx\, \frac{df}{dx} = 0$, hence

$$\begin{aligned}
0 &= \int dx^N \frac{d^2}{x^j x^k} \left(x_{i_1} x_{i_2} \cdots x_{i_{2n}} e^{-\sum_{m=1}^N x_m^2 \lambda_m / 2} \right) \\
&= \int dx^N \left(\sum_{r \neq s=1}^{2n} \delta_{i_r j} \delta_{i_s k} x_{i_1} \cdots \check{x}_{i_r} \cdots \check{x}_{i_s} \cdots x_{i_{2n}} - \lambda_j \delta_{jk} x_{i_1} \cdots x_{i_{2n}} \right. \\
&\qquad + \lambda_j \lambda_k x_j x_k x_{i_1} \cdots x_{i_{2n}} - \sum_r \lambda_j x_j \delta_{k i_r} x_{i_1} \cdots \check{x}_{i_r} \cdots x_{i_{2n}} \\
&\qquad \left. - \sum_r \lambda_k x_k \delta_{j i_r} x_{i_1} \cdots \check{x}_{i_r} \cdots x_{i_{2n}} \right) e^{-\sum_{m=1}^N x_m^2 \lambda_m / 2} .
\end{aligned}$$

Here a check over a variable x_m means that this variable is omitted in the product. In the above integral, all except the term proportional to $\lambda_j \lambda_k$ are $2n$-point correlators. We can therefore express the $(2n + 2)$-point correlator proportional to $\lambda_j \lambda_k$ in terms of $2n$-point correlators. Furthermore, we shall use the fact that $\lambda_k^{-1} \delta_{jk} = \langle X_j X_k \rangle$. The above equation therefore gives

$$\begin{aligned}
\langle X_j X_k X_{i_1} \cdots X_{i_{2n}} \rangle = &- \sum_{r \neq s=1}^{2n} \langle X_{i_r} X_j \rangle \langle X_{i_s} X_k \rangle \langle X_{i_1} \cdots \check{X}_{i_r} \cdots \check{X}_{i_s} \cdots X_{i_{2n}} \rangle \\
&+ \langle X_j X_k \rangle \langle X_{i_1} \cdots X_{i_{2n}} \rangle \\
&+ \sum_r \langle X_k X_{i_r} \rangle \langle X_j X_{i_1} \cdots \check{X}_{i_r} \cdots X_{i_{2n}} \rangle \\
&+ \sum_r \langle X_j X_{i_r} \rangle \langle X_k X_{i_1} \cdots \check{X}_{i_r} \cdots X_{i_{2n}} \rangle .
\end{aligned}$$

Wick's theorem for $2n$ implies that

$$\langle X_k X_{i_1} \cdots \check{X}_{i_r} \cdots X_{i_{2n}} \rangle = \sum_s \langle X_k X_{i_s} \rangle \langle X_{i_1} \cdots \check{X}_{i_r} \cdots \check{X}_{i_s} \cdots X_{i_{2n}} \rangle .$$

Therefore, the last sum cancels with the double sum and we end up with

$$\begin{aligned}
\langle X_j X_k X_{i_1} \cdots X_{i_{2n}} \rangle = &\langle X_j X_k \rangle \langle X_{i_1} \cdots X_{i_{2n}} \rangle \\
&+ \sum_r \langle X_k X_{i_r} \rangle \langle X_j X_{i_1} \cdots \check{X}_{i_r} \cdots X_{i_{2n}} \rangle .
\end{aligned} \tag{A7.9}$$

But since the $2n$-point correlators $\langle X_{i_1} \cdots X_{i_{2n}} \rangle$ and $\langle X_j X_{i_1} \cdots \check{X}_{i_r} \cdots X_{i_{2n}} \rangle$ are the sum of all possible products of 2-point correlators, this represents simply the sum of all possible products of 2-point correlators of X_j, X_k, $X_{i_1}, \ldots, X_{i_{2n}}$, hence Wick's theorem is proven.

A7.2 Random fields

A random field is an application $X : S \to \{\text{random variables}\} : \mathbf{n} \mapsto X(\mathbf{n})$ which assigns to each point $\mathbf{n}$ in the space S a random variable $X(\mathbf{n})$. Here the space S can be $\mathbb{R}^n$, the sphere or some other space. We think mainly of S being the CMB sky, hence the

sphere on which our random fields are for example the temperature fluctuations $\Delta T(\mathbf{n})$ or the polarization. Another example is three-dimensional-space (either $\mathbb{R}^3$, the 3-sphere or the 3-pseudo-sphere) where, for example, the density and velocity fluctuations are random fields of interest to us.

Definition A7.2: *A random field is called Gaussian if arbitrary collections* $X(\mathbf{n}_1), \ldots, X(\mathbf{n}_N)$ *are Gaussian random variables. The correlator*

$$\langle X(\mathbf{n}_1)X(\mathbf{n}_2)\rangle = C(\mathbf{n}_1, \mathbf{n}_2)$$

is called the correlation function or the 2-point function.

For Gaussian random fields, the n-point function is given by the sum of all possible different products of 2-point functions.[3]

A7.2.1 Statistical homogeneity and isotropy

A random field respects a symmetry group G of the space S if the correlation function is invariant under transformations $\mathbf{n} \mapsto R\mathbf{n}$ for all $R \in G$. In other words

$$C(R\mathbf{n}_1, R\mathbf{n}_2) = C(\mathbf{n}_1, \mathbf{n}_2)\,.$$

For this it is not necessary that $X(\mathbf{n}) = X(R\mathbf{n})$, but the transformed variable must have identical statistical properties. In cosmology, we expect the CMB sky to be statistically isotropic, i.e., invariant under rotations. This means that the correlation function is a function only of the scalar product $\mu = \mathbf{n}_1 \cdot \mathbf{n}_2$. Therefore, we can expand it in terms of Legendre polynomials,

$$C(\mathbf{n}_1, \mathbf{n}_2) = C(\mu) = \frac{1}{4\pi}\sum_\ell (2\ell+1) C_\ell P_\ell(\mu)\,. \tag{A7.10}$$

We shall show now, that when expanding a statistically isotropic random variable on the CMB sky in spherical harmonics,

$$X(\mathbf{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\mathbf{n}) \tag{A7.11}$$

the coefficients $a_{\ell m}$ satisfy

$$\langle a_{\ell_1 m_1} \bar{a}_{\ell_2 m_2}\rangle = \delta_{m_1 m_2}\delta_{\ell_1 \ell_2} C_{\ell_1}\,. \tag{A7.12}$$

To see this we write the correlation function as

$$\sum_{\ell m} C_\ell Y_{\ell m}(\mathbf{n}_1)\bar{Y}_{\ell m}(\mathbf{n}_2) = \sum_{\ell m \ell' m'} \langle a_{\ell m}\bar{a}_{\ell' m'}\rangle Y_{\ell m}(\mathbf{n}_1)\bar{Y}_{\ell' m'}(\mathbf{n}_2)\,. \tag{A7.13}$$

For the left-hand side of this equation we have used the addition theorem for spherical harmonics (see Appendix A4.2.3) to replace the P_ℓs in Eq. (A7.10) and for the left-hand side we simply used the expansion (A7.11). Multiplying this equation by $\bar{Y}_{\ell_1 m_1}(\mathbf{n}_1) Y_{\ell_2 m_2}(\mathbf{n}_2)$ and integrating over $\mathbf{n}_1$ and $\mathbf{n}_2$ we obtain Eq. (A7.12).

Therefore, for statistically isotropic random fields, the expansion in terms of spherical harmonics diagonalizes the correlation function. This is why it is so useful to determine

[3] Physicists are often more familiar with quantum field theory. In (Euclidean) quantum field theory the 2-point function is called the propagator and the theory is Gaussian if and only if it is trivial. Only in the absence of interactions are all n-point functions determined by the propagators and the so-called 'connected part', which is the n-point function subtracted by the Gaussian result, vanishes.

the $a_{\ell m}$s measured by an experiment. If the underlying fluctuations are Gaussian, these are independent Gaussian variables.

Let us now turn to random fields on three-dimensional space. For simplicity we consider Euclidean space. For a random field $X(\mathbf{x})$ we expect its statistical properties to be independent of translations and rotations. Therefore, the correlation function $C(\mathbf{x}_1, \mathbf{x}_2)$ depends only on the distance $|\mathbf{x}_1 - \mathbf{x}_2| \equiv r$. We now show that for statistically homogeneous random fields, the power spectrum is simply the Fourier transform of the correlation function. Statistical isotropy implies that the latter depends only on the modulus of $\mathbf{k}$. In Chapter 2 we have defined the power spectrum of a statistically homogeneous and isotropic random field X by

$$\langle X(\mathbf{k})\bar{X}(\mathbf{k}')\rangle = (2\pi)^3\delta(\mathbf{k}-\mathbf{k}')P_X(k)\,. \tag{A7.14}$$

We shall now see that $\langle X(\mathbf{k})\bar{X}(\mathbf{k}')\rangle$ is indeed of this form and that

$$P_X(k) = \hat{C}(k)\,. \tag{A7.15}$$

We simply insert the definition

$$\begin{aligned}
\langle X(\mathbf{k})\bar{X}(\mathbf{k}')\rangle &= \int d^3x\, d^3x'\, \langle X(\mathbf{x})X(\mathbf{x}')\rangle e^{i(\mathbf{k}\cdot\mathbf{x}-\mathbf{k}'\cdot\mathbf{x}')} \\
&= \int d^3x\, d^3x'\, \langle C(\mathbf{x}-\mathbf{x}')\rangle e^{i(\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')-(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}')} \\
&= \int d^3z\, d^3x'\, \langle C(\mathbf{z})\rangle e^{i(\mathbf{k}\cdot\mathbf{z}-(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}')} \\
&= (2\pi)^3\,\delta(\mathbf{k}-\mathbf{k}')\hat{C}(k)\,.
\end{aligned}$$

For the third equals sign we made the variable transform $\mathbf{z} = \mathbf{x} - \mathbf{x}'$ and for the last equals sign we used the fact that the integral of $e^{i\mathbf{k}\cdot\mathbf{x}}$ is a delta function. This proves the ansatz (A7.14) and Eq. (A7.15). This shows that for statistically homogeneous fields, the Fourier coefficients $X(\mathbf{k})$ are independent random variables and are therefore especially useful for statistical analysis and parameter estimation.

Appendix 8

Approximation for the tensor C_ℓ spectrum

In this appendix we derive in a self-consistent way the approximation (2.256) for the C_ℓ power spectrum of tensor perturbations in a FL background with vanishing curvature, $K = 0$. We start with the power spectrum of tensor perturbations of the metric,

$$\left\langle H_{ij}^{(T)}(\mathbf{k}, t) H_{mn}^{(T)*}(\mathbf{k}', t') \right\rangle = (2\pi)^3 P_{ijmn}(\mathbf{k}, t, t') \delta^3(\mathbf{k} - \mathbf{k}') \, .$$

$H_{ij}^{(T)}(\mathbf{k}, t)$ is symmetric, traceless and transverse, and since its Fourier transform is real we have $H_{ij}^{(T)*}(\mathbf{k}, t) = H_{ij}^{(T)}(-\mathbf{k}, t)$, hence $P_{ijmn}(\mathbf{k}, t, t') = P_{mnij}(-\mathbf{k}, t', t)$.

We define the projection tensor onto the plane normal to $\mathbf{k}$,

$$P_{ij} = \delta_{ij} - k^{-2} k_i k_j \, . \tag{A8.1}$$

The most generic tensor P_{ijmn} which has the above properties, is isotropic[1] and is also invariant under parity, $P_{ijmn}(\mathbf{k}, t, t') = P_{ijmn}(-\mathbf{k}, t, t')$, is to be of the form[2]

$$P_{ijmn}(\mathbf{k}, t, t') = \mathcal{H}(k, t, t') \mathcal{M}_{ijmn}(\mathbf{k}) \quad \text{with} \tag{A8.2}$$

$$\begin{aligned} \mathcal{M}_{ijmn}(\mathbf{k}) &\equiv P_{im} P_{jn} + P_{in} P_{jm} - P_{ij} P_{mn} \\ &= [\delta_{im}\delta_{jn} + \delta_{in}\delta_{jm} - \delta_{ij}\delta_{mn} + k^{-2}(\delta_{ij} k_m k_n \\ &\quad + \delta_{mn} k_i k_j - \delta_{im} k_j k_n - \delta_{in} k_m k_j - \delta_{jm} k_i k_n - \delta_{jn} k_m k_i) \\ &\quad + k^{-4} k_i k_j k_m k_n] \, . \end{aligned} \tag{A8.3}$$

The Fourier transform on Eq. (2.234) gives

$$\left(\frac{\Delta T(\mathbf{n}, \mathbf{k})}{T} \right)^{(T)} = - \int_i^f dt \, \exp(i \mathbf{k} \cdot \mathbf{n} (t_0 - t)) \, \partial_t H_{ij}(t, \mathbf{k}) n^i n^j \, . \tag{A8.4}$$

[1] This means that only $\mathbf{k}$ and invariant tensors like δ_{ij} or ϵ_{ijm} enter its construction. No external given vector or tensor which is not invariant under rotations is allowed.

[2] If we do not require parity invariance, an additional term proportional to $\mathcal{A}_{ijlm} = k^{-1} k_q \left(P_{jm} \epsilon_{ilq} + P_{il} \epsilon_{jmq} + P_{im} \epsilon_{jlq} + P_{jl} \epsilon_{imq} \right)$ can be added. But this term changes sign under parity and can be shown to be proportional to the difference of the amplitudes of the two polarization states (Caprini *et al.*, 2004). When it is present, parity violating terms (like corrolators between the temperature anisotropy and B-polarization, (see Caprini *et al.*, 2004 and Chapter 5) appear. We neglect this possibility in this appendix.

With this and Eq. (A8.2) we obtain

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{n},\mathbf{x})\frac{\Delta T}{T}(\mathbf{n}',\mathbf{x})\right\rangle \\
&\equiv \left(\frac{1}{2\pi}\right)^6 \int d^3k\, d^3k' \left\langle \frac{\Delta T}{T}(\mathbf{n},\mathbf{k})\frac{\Delta T}{T}(\mathbf{n}',\mathbf{k}')\right\rangle \exp(-i\mathbf{x}(\mathbf{k}-\mathbf{k}')) \\
&= \left(\frac{1}{2\pi}\right)^3 \int k^2\, dk\, d\Omega_{\hat{\mathbf{k}}} \int_{t_{\rm dec}}^{t_0} dt \int_{t_{\rm dec}}^{t_0} dt' \exp(i\mathbf{k}\cdot\mathbf{n}(t_0-t)) \exp(-i\mathbf{k}\cdot\mathbf{n}'(t_0-t')) \\
&\quad \times \frac{\partial^2}{\partial t\,\partial t'} P_{ijlm}(\mathbf{k},t,t') n_i n_j n'_l n'_m \,. \qquad \text{(A8.5)}
\end{aligned}$$

Here $d\Omega_{\hat{\mathbf{k}}}$ denotes the integral over directions in $\mathbf{k}$ space and we made use of the δ^3-function to get rid of the integral over d^3k'.

We now introduce the form (A8.2) of P_{ijlm}. We further assume that the perturbations have been created at some early epoch, e.g. during an inflationary phase, after which they evolved deterministically. The function $\mathcal{H}(k,t,t')$ is thus a product of the form

$$\mathcal{H}(k,t,t') = H(k,t)\cdot H^*(k,t')\,, \qquad \text{(A8.6)}$$

where $H(k,t)$ is the growing mode solution of Eq. (2.108) with the correct initial spectrum, $\langle H_{ij}(\mathbf{k},t_{\rm in}) H^*_{ij}(\mathbf{k}',t_{\rm in})\rangle = 4|H(k,t_{\rm in})|^2\delta^2(\mathbf{k}-\mathbf{k}')$ and $\langle \dot{H}_{ij}(\mathbf{k},t_{\rm in})\dot{H}^*_{ij}(\mathbf{k}',t_{\rm in})\rangle = 4|\dot{H}(k,t_{\rm in})|^2\delta(\mathbf{k}-\mathbf{k}')$. Introducing this form of P_{ijlm} in Eq. (A8.5) yields

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{n})\frac{\Delta T}{T}(\mathbf{n}')\right\rangle \\
&= \left(\frac{1}{2\pi}\right)^3 \int k^2\, dk\, d\Omega_{\hat{\mathbf{k}}} \left[2(\mathbf{n}\cdot\mathbf{n}')^2 - 1 + \mu'^2 + \mu^2 - 4\mu\mu'(\mathbf{n}\cdot\mathbf{n}') + \mu^2\mu'^2\right] \\
&\quad \times \int_{t_{\rm dec}}^{t_0} dt \int_{t_{\rm dec}}^{t_0} dt' \left[\dot{H}(k,t)\dot{H}^*(k,t') \exp(ik\mu(t_0-t)) \exp(-ik\mu'(t_0-t'))\right] , \qquad \text{(A8.7)}
\end{aligned}$$

where $\mu = (\mathbf{n}\cdot\hat{\mathbf{k}})$, $\mu' = (\mathbf{n}'\cdot\hat{\mathbf{k}})$ and $\dot{H} = \partial_t H$. To proceed, we use the identity (Abramowitz & Stegun, 1970)

$$\exp(ik\mu(t_0-t)) = \sum_{r=0}^{\infty} (2r+1) i^r j_r(k(t_0-t)) P_r(\mu)\,. \qquad \text{(A8.8)}$$

Here j_r denotes the spherical Bessel function of order r and P_r is the Legendre polynomial of degree r.

Furthermore, we replace each factor of μ in Eq. (A8.7) by a derivative of the exponential $\exp(ik\mu(t_0-t))$ with respect to $k(t_0-t)$ and correspondingly with μ'. We then obtain

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{n})\frac{\Delta T}{T}(\mathbf{n}')\right\rangle \\
&= \left(\frac{1}{2\pi}\right)^3 \sum_{r,r'} (2r+1)(2r'+1) i^{(r-r')} \int k^2\, dk\, d\Omega_{\hat{\mathbf{k}}}\, P_r(\mu) P_{r'}(\mu') \\
&\quad \times \Big[2(\mathbf{n}\cdot\mathbf{n}')^2 \int dt\, dt'\, j_r(k(t_0-t)) j_{r'}(k(t_0-t'))\dot{H}(k,t)\dot{H}^*(k,t')
\end{aligned}$$

$$
\begin{aligned}
&- \int dt dt' [j_r(k(t_0-t)) j_{r'}(k(t_0-t')) + j_r''(k(t_0-t)) j_{r'}(k(t_0-t')) \\
&+ j_r(k(t_0-t)) j_{r'}''(k(t_0-t')) - j_r''(k(t_0-t)) j_{r'}''(k(t_0-t'))] \dot{H}(k,t) \dot{H}^*(k,t') \\
&- 4(\mathbf{n}\cdot\mathbf{n}') \int dt\, dt'\, j_r'(k(t_0-t)) j_{r'}'(k(t_0-t')) \dot{H}(k,t) \dot{H}^*(k,t') \Big] \,. \qquad \text{(A8.9)}
\end{aligned}
$$

The primes on the Bessel functions denote the derivative with respect to the argument of the function. In the expression (A8.9) only the Legendre polynomials, $P_r(\mu)$ and $P_{r'}(\mu')$ depend on the direction $\hat{\mathbf{k}}$. To perform the integration $d\Omega_{\hat{\mathbf{k}}}$, we use the addition theorem for spherical harmonics (see Appendix A4.2.3),

$$
P_r(\mu) = \frac{4\pi}{(2r+1)} \sum_{s=-r}^{r} Y_{rs}(\mathbf{n}) Y_{rs}^*(\hat{\mathbf{k}}) \,. \qquad \text{(A8.10)}
$$

The orthogonality of spherical harmonics (see Appendix A4.2.3) then yields

$$
\begin{aligned}
&(2r+1)(2r'+1) \int d\Omega_{\hat{\mathbf{k}}}\, P_r(\mu) P_{r'}(\mu') \\
&\quad = 16\pi^2 \delta_{rr'} \sum_{s=-r}^{r} Y_{rs}(\mathbf{n}) Y_{rs}^*(\mathbf{n}') \\
&\quad = (2r+1) 4\pi \delta_{rr'} P_r(\mathbf{n}\cdot\mathbf{n}') \,. \qquad \text{(A8.11)}
\end{aligned}
$$

In Eq. (A8.9) the integration over $d\Omega_{\hat{\mathbf{k}}}$ leads to terms of the form $(\mathbf{n}\cdot\mathbf{n}')P_r(\mathbf{n}\cdot\mathbf{n}')$ and $(\mathbf{n}\cdot\mathbf{n}')^2 P_r(\mathbf{n}\cdot\mathbf{n}')$. To reduce them, we use recursion relations for Legendre polynomials like

$$
x P_r(x) = \frac{r+1}{2r+1} P_{r+1} + \frac{r}{2r+1} P_{r-1} \,. \qquad \text{(A8.12)}
$$

Applying this and its iteration for $x^2 P_r(x)$, we obtain

$$
\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{n}) \frac{\Delta T}{T}^*(\mathbf{n}') \right\rangle \\
&\quad = \frac{1}{2\pi^2} \sum_r (2r+1) \int k^2 dk \int dt dt' \dot{H}(k,t) \dot{H}^*(k,t') \times \\
&\quad \Big\{ \Big[\frac{2(r+1)(r+2)}{(2r+1)(2r+3)} P_{r+2} + \frac{1}{(2r-1)(2r+3)} P_r + \frac{2r(r-1)}{(2r-1)(2r+1)} P_{r-2} \Big] \\
&\quad \times j_r(k(t_0-t)) j_r(k(t_0-t')) - P_r [j_r(k(t_0-t)) j_r''(k(t_0-t')) \\
&\quad + j_r(k(t_0-t')) j_r''(k(t_0-t)) - j_r''(k(t_0-t)) j_{r'}''(k(t_0-t'))] \\
&\quad - 4 \Big[\frac{r+1}{2r+1} P_{r+1} + \frac{r}{2r+1} P_{r-1} \Big] j_r'(k(t_0-t)) j_r'(k(t_0-t')) \Big\} \,, \qquad \text{(A8.13)}
\end{aligned}
$$

where the argument, $\mathbf{n}\cdot\mathbf{n}'$, of the Legendre polynomials has been suppressed. Also using the relation

$$
j_r' = -\frac{r+1}{2r+1} j_{r+1} + \frac{r}{2r+1} j_{r-1} \qquad \text{(A8.14)}
$$

for Bessel functions, and its iteration for j'', we can rewrite Eq. (A8.13) in terms of the Bessel functions j_{r-2} to j_{r+2}.

To proceed we use the definition of C_ℓ:

$$\left\langle \frac{\Delta T}{T}(\mathbf{n}) \cdot \frac{\Delta T}{T}(\mathbf{n}') \right\rangle_{(\mathbf{n}\cdot\mathbf{n}')=\cos\theta} = \frac{1}{4\pi} \Sigma (2\ell+1) C_\ell P_\ell(\cos\theta) \,. \tag{A8.15}$$

If we expand

$$\frac{\Delta T}{T}(\mathbf{n}) = \sum_{\ell,m} a_{\ell,m} Y_{\ell,m}(\mathbf{n}) \tag{A8.16}$$

and use the orthogonality of the spherical harmonics as well as the addition theorem, Eq. (A8.10), we find

$$C_\ell = \langle a_{\ell m} a^*_{\ell m} \rangle \,. \tag{A8.17}$$

We thus have to determine the correlators

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \int d\Omega_{\mathbf{n}} \int d\Omega_{\mathbf{n}'} \left\langle \frac{\Delta T}{T}^*(\mathbf{n}) \frac{\Delta T}{T}(\mathbf{n}') \right\rangle Y^*_{\ell m}(\mathbf{n}) Y_{\ell' m'}(\mathbf{n}') \,. \tag{A8.18}$$

Inserting our result (A8.13), we obtain the somewhat lengthy expression

$$\begin{aligned}
&\langle a_{\ell m} a^*_{\ell' m'} \rangle = \\
&\frac{2}{\pi} \delta_{\ell\ell'} \, \delta m m' \int dk\, k^2 \int dt\, dt'\, \dot{H}(k,t) \dot{H}^*(k,t') \\
&\times \Bigg\{ j_\ell(k(t_0-t)) j_\ell(k(t_0-t')) \\
&\times \left(\frac{1}{(2\ell-1)(2\ell+3)} + \frac{2(2\ell^2+2\ell-1)}{(2\ell-1)(2\ell+3)} + \frac{(2\ell^2+2\ell-1)^2}{(2\ell-1)^2(2\ell+3)^2} \right. \\
&\left. - \frac{4\ell^3}{(2\ell-1)^2(2\ell+1)} - \frac{4(\ell+1)^3}{(2\ell+1)(2\ell+3)^2} \right) \\
&- \left[j_\ell(k(t_0-t)) j_{\ell+2}(k(t_0-t')) + j_{\ell+2}(k(t_0-t)) j_\ell(k(t_0-t')) \right] \\
&\times \frac{1}{2\ell+1} \left(\frac{2(\ell+2)(\ell+1)(2\ell^2+2\ell-1)}{(2\ell-1)(2\ell+3)^2} + \frac{2(\ell+1)(\ell+2)}{(2\ell+3)} - \frac{8(\ell+1)^2(\ell+2)}{(2\ell+3)^2} \right) \\
&- \left[j_\ell(k(t_0-t)) j_{\ell-2}(k(t_0-t')) + j_{\ell-2}(k(t_0-t)) j_\ell(k(t_0-t')) \right] \\
&\times \frac{1}{2\ell+1} \left(\frac{2\ell(\ell-1)(2\ell^2+2\ell-1)}{(2\ell-1)^2(2\ell+3)} + \frac{2\ell(\ell-1)}{(2\ell-1)^2} - \frac{8\ell^2(\ell-1)}{(2\ell-1)^2} \right) \\
&+ j_{\ell+2}(k(t_0-t)) j_{\ell+2}(k(t_0-t')) \\
&\times \left(\frac{2(\ell+2)(\ell+1)}{(2\ell+1)(2\ell+3)} - \frac{4(\ell+1)(\ell+2)^2}{(2\ell+1)(2\ell+3)^2} + \frac{(\ell+1)^2(\ell+2)^2}{(2\ell+1)^2(2\ell+3)^2} \right) \\
&+ j_{\ell-2}(k(t_0-t)) j_{\ell-2}(k(t_0-t')) \\
&\times \left(\frac{2\ell(\ell-1)}{(2\ell-1)(2\ell+1)} - \frac{4\ell(\ell-1)^2}{(2\ell-1)^2(2\ell+1)} + \frac{\ell^2(\ell-1)^2}{(2\ell-1)^2(2\ell+1)^2} \right) \Bigg\} .
\end{aligned} \tag{A8.19}$$

An analysis of the coefficient of each term reveals that this expression is equivalent to

$$C_\ell^{(T)} = \frac{2}{\pi} \int dk\, k^2 \left| \int_{t_{\rm dec}}^{t_0} dt\, \dot{H}(t,k) \frac{j_\ell(k(t_0-t))}{(k(t_0-t))^2} \right|^2 \frac{(\ell+2)!}{(\ell-2)!}\,. \tag{A8.20}$$

To obtain this result we have used the identity

$$\begin{aligned} &\frac{j_{\ell+2}(k(t_0-t))}{(2\ell+3)(2\ell+1)} + \frac{2j_\ell(k(t_0-t))}{(2\ell+3)(2\ell-1)} + \frac{j_{\ell-2}(k(t_0-t))}{(2\ell+1)(2\ell-1)} \\ &\quad = \frac{j_\ell(k(t_0-t))}{(k(t_0-t))^2}\,. \end{aligned} \tag{A8.21}$$

Appendix 9

Boltzmann equation in a universe with curvature

In this appendix we discuss the changes of the Boltzmann equation in the case of non-vanishing curvature $K = H_0^2(\Omega_0 - 1) \neq 0$. We closely follow Abbott & Schaefer (1986) and Hu *et al*. (1998). We write the metric of the unperturbed FL universe as

$$ds^2 = a^2 \gamma_{\mu\nu}\, dx^\mu\, dx^\nu\ , \tag{A9.1}$$

with

$$\gamma_{00} = -1\ , \tag{A9.2}$$

$$\gamma_{0i} = 0\ , \tag{A9.3}$$

$$\gamma_{ij}\, dx^i\, dx^j = \frac{1}{|K|}\left(d\chi^2 + \sin_K^2(\chi)(d\theta^2 + \sin^2\theta\, d\phi^2)\right)\ . \tag{A9.4}$$

Here we have set

$$\sin_K \chi = \begin{cases} \sin\chi & \text{for} \quad K > 0 \\ \chi & \text{for} \quad K = 0 \\ \mathrm{sh}\chi & \text{for} \quad K < 0\ , \end{cases} \tag{A9.5}$$

where sh denotes the hyperbolic sine (correspondingly we shall denote the hyperbolic cosine by ch).

A9.1 The Boltzmann equation

We now derive the Boltzmann equation. The collision term is not affected by curvature since it is purely local. If we use a 'quasi-orthonormal' spatial basis, also the gravitational source term coming from $\delta\Gamma^i_{\alpha\beta} n^\alpha n^\beta \gamma_{ij} n^j v^2 (d\bar{f}/dv)$ is not modified. Here 'quasi-orthonormal means that $n^i n^j \gamma_{ij} = 1$ and $(p^\mu) = (p, pn^i)$. Again $v = pa$ denotes the redshift corrected photon energy, the only variable on which the background distribution function $\bar{f}$ depends. The only modification comes from the fact that on the left-hand side we have to add a term due to the unperturbed three-dimensional Christoffel symbols,

$$(\partial_t + n^i \partial_i)\mathcal{M} \to \left(\partial_t + n^i \partial_i - \bar{\Gamma}^i_{jl} n^j n^l \frac{\partial}{\partial n^i}\right)\mathcal{M}\ . \tag{A9.6}$$

We now show that this simply corresponds to replacing the partial derivatives ∂_i in the second term by covariant derivatives w.r.t. the spatial background metric γ_{ij}. To see this

we expand $\mathcal{M}$ in moments of n^i,

$$\mathcal{M}(t,\mathbf{x},\mathbf{n}) = \sum Q(t,\mathbf{x})^{(m)}_{i_1\cdots i_m} n^{i_1}\cdots n^{i_m} \,. \tag{A9.7}$$

Here the sum goes from 1 to 3 for each of the indices i_l and the number m of indices goes from zero to infinity. For uniqueness, we require that $Q_{i_1\cdots i_m}$ be a traceless totally symmetric tensor. Since $\gamma_{jl}n^j n^l = 1$, a trace in $Q^{(m)}_{i_1\cdots i_{m-2}jl}$ which contributes a term $Q^{(m-2)}_{i_1\cdots i_{m-2}}\gamma_{jl}n^{i_1}\cdots n^{i_{m-2}}n^j n^l = Q^{(m-2)}_{i_1\cdots i_{m-2}}$ and can be absorbed in $Q^{(m-2)}$. With this ansatz we find

$$\begin{aligned}
&\left(n^s\partial_s - \bar\Gamma^j_{rs}n^r n^s\frac{\partial}{\partial n^j}\right)\mathcal{M} \qquad\qquad (A9.8)\\
&\quad= \sum n^s\partial_s Q(t,\mathbf{x})^{(m)}_{i_1\cdots i_m} n^{i_1}\cdots n^{i_m} - \bar\Gamma^j_{rs}Q^{(m)}_{i_1\cdots i_m} n^r n^s(\delta_{ji_1}n^{i_2}\cdots n^{i_m}\\
&\qquad\qquad + \cdots + n^{i_1}\cdots n^{i_{m-1}}\delta_{ji_m})\\
&\quad= \sum n^s\left(\partial_s Q(t,\mathbf{x})^{(m)}_{i_1\cdots i_m} n^{i_1}\cdots n^{i_m} - \bar\Gamma^j_{rs}Q^{(m)}_{j\cdots i_m} n^r n^{i_2}\cdots n^{i_m}\right.\\
&\qquad\left. - \cdots - \bar\Gamma^j_{rs}Q^{(m)}_{i_1\cdots j} n^r n^{i_1}\cdots n^{i_{m-1}}\right)\\
&\quad= \sum n^s\left(\partial_s Q(t,\mathbf{x})^{(m)}_{i_1\cdots i_m} n^{i_1}\cdots n^{i_m} - \bar\Gamma^j_{i_1 s}Q^{(m)}_{j\cdots i_m} n^{i_1} n^{i_2}\cdots n^{i_m}\right.\\
&\qquad\left. - \cdots - \bar\Gamma^j_{i_m s}Q^{(m)}_{i_1\cdots j} n^{i_1}\cdots n^{i_m}\right)\\
&\quad= \sum n^s Q(t,\mathbf{x})^{(m)}_{i_1\cdots i_m|s} n^{i_1}\cdots n^{i_m} \equiv n^s\mathcal{M}_{|s}\,. \qquad\qquad (A9.9)
\end{aligned}$$

Note that the last expression really is a definition. It tells us how we have to interpret a covariant derivative for a function in momentum space. With this, the Boltzmann equation in spaces with non-vanishing curvature becomes simply

$$\partial_t\mathcal{V} + n^i\mathcal{V}_{|i} = C[\mathcal{V}] + \begin{pmatrix} G[h_{\mu\nu}]\\ 0\\ 0\end{pmatrix}\,, \tag{A9.10}$$

where $C[\mathcal{V}]$ denotes the collision term given in Eq. (5.29) and

$$\mathcal{V} = \begin{pmatrix}\mathcal{M}\\ \mathcal{E}+i\mathcal{B}\\ \mathcal{E}-i\mathcal{B}\end{pmatrix}\,. \tag{A9.11}$$

The gravitational term is

$$G[h_{\mu\nu}]\begin{cases} -n^i(\Psi+\Phi)_{|i} & \text{for scalar perturbations,}\\ -\sigma^{(V)}_{ij}n^i n^j & \text{for vector perturbations,}\\ -\dot H_{ij}n^i n^j & \text{for tensor perturbations.}\end{cases} \tag{A9.12}$$

A9.2 Line-of-sight integration

The homogeneous part of the Boltzmann equation (A9.10),

$$\partial_t\mathcal{V} + n^i\mathcal{V}_{|i} = 0\,, \tag{A9.13}$$

simply represents free streaming. If the source term can be neglected, the temperature fluctuations and the polarization are modified by photon free streaming, which in this case

means that the photons move along spatial geodesics of the unperturbed metric γ_{ij}. Unlike in flat space, $\mathbf{n}$ is not constant, but varies along a geodesic.

Let us denote by $\mathbf{y}(\mathbf{x}, \mathbf{n}(\lambda), \lambda)$ the (spatial) geodesic of the unperturbed metric γ_{ij} which arrives at $\mathbf{x}$ at time λ and then moves on in direction $-\mathbf{n}(\lambda)$, so that $\mathbf{x} = \mathbf{y}(\mathbf{x}, \mathbf{n}, \lambda)$. In flat space, $\mathbf{y}(\mathbf{x}, \mathbf{n}, \lambda)(t) = \mathbf{x} - (t - \lambda)\mathbf{n}$. The general solution to Eq. (A9.13) with initial condition $\mathcal{V}(t_{\rm in}, \mathbf{x}, \mathbf{n}) = \mathcal{V}_1(\mathbf{x}, \mathbf{n})$ is then simply,

$$\mathcal{V}(t, \mathbf{x}, \mathbf{n}) = \mathcal{V}_1(\mathbf{y}(\mathbf{x}, \mathbf{n}(t - t_{\rm in}), t - t_{\rm in}), \mathbf{n}(t - t_{\rm in})) \,. \tag{A9.14}$$

To verify this we use that both, $\mathbf{y}$ and $\mathbf{n}$ depend on time so that $\partial_t \mathcal{V}(t, \mathbf{x}, \mathbf{n}) = -n^i \partial_i \mathcal{V}(t, \mathbf{x}, \mathbf{n}) + \dot{n}^i \frac{\partial}{\partial n^i} \mathcal{V}(t, \mathbf{x}, \mathbf{n})$. But since n moves along a geodesic $\dot{n}^i = \bar{\Gamma}^i_{rs} n^r n^s$, so that we end up with

$$\partial_t \mathcal{V}(t, \mathbf{x}, \mathbf{n}) + n^i \partial_i \mathcal{V}(t, \mathbf{x}, \mathbf{n}) - \bar{\Gamma}^i_{rs} n^r n^s \frac{\partial}{\partial n^i} \mathcal{V}(t, \mathbf{x}, \mathbf{n}) = 0 \,,$$

the right-hand side of Eq. (A9.6) which is equivalent to Eq. (A9.13) according to our definition (A9.9). This observation allows us also to formally solve (A9.10) with a line-of-sight integration as in the flat case: for an arbitrary source term, $S(t, \mathbf{x}, \mathbf{n})$ on the right-hand side of Eq. (A9.13) the solution is given by

$$\mathcal{V}(t, \mathbf{x}, \mathbf{n}) = \mathcal{V}_1(\mathbf{y}(\mathbf{x}, \mathbf{n}, t), \mathbf{n}) + \int_{t_{\rm in}}^{t} dt' \, S(t', \mathbf{y}(\mathbf{x}(t'), \mathbf{n}(t'), t'), \mathbf{n}(t')). \tag{A9.15}$$

Here $\mathbf{x}(t')$ is the geodesic which ends at $\mathbf{x} = \mathbf{x}(t)$ at time t with velocity $-\mathbf{n}(t)$.

So far this is only a formal solution. In the case of the Boltzmann equation the source on the right-hand side depends on the left-hand side. Furthermore, the geodesics $\mathbf{y}(\mathbf{x}, \mathbf{n}(t'), t')$ are not given explicitly. However, this is not a serious problem, since they are the solutions to well known ordinary differential equations.

A9.3 Mode functions, radial functions

In flat space, we have expanded Eq. (A9.10) in terms of mode functions

$${}_sG_{\ell m}(\mathbf{x}, \mathbf{n}) = (-i)^\ell \sqrt{\frac{4\pi}{2\ell + 1}} \, {}_sY_{\ell m}(\mathbf{n}) \exp(i\mathbf{k} \cdot \mathbf{x}) \,, \quad K = 0 \,. \tag{A9.16}$$

Here we have to replace the exponentials by eigenfunctions of the Laplacian in curved space. The functions $Q^{(m)}_{i_1 \cdots i_{|m|}}$ with

$$(\Delta_K + k^2) Q^{(m)}_{i_1 \cdots i_{|m|}} = 0 \,, \tag{A9.17}$$

where $k^2 > (|m| + 1)|K|$ and $k^2 = (p(p + 2) + |m|)K$, $p \in \mathbb{N}$, from a complete set of basis functions for $K < 0$ and $K > 0$ respectively (see Vilenkin & Smorodinskii, 1964). Here, like in Chapter 2, $Q^{(m)}$ is a totally symmetric traceless tensor with helicity m and rank $|m|$.

We shall concentrate on the case $K > 0$ since it is really different from that flat space case as its k-modes are discrete. For convenience we set

$$q = \sqrt{p(p + 2) + 1 + |m|} \,, \quad K > 0 \tag{A9.18}$$

$$q = \sqrt{\frac{k^2}{|K|} - 1 - |m|} \,, \quad K < 0 \,, \tag{A9.19}$$

and we shall use this dimensionless number to denominate the mode functions. The functions $p = 0$ and $p = 1$ are not of interest for us. They represent a simple constant ($p = 0$) and ($p = 1$) a pure dipole contribution which is gauge dependent. We therefore consider $q = 3, 4, \ldots$ for $m = 0$, $K > 0$.

As we have done for the exponential, we want to expand the function $Q(\mathbf{x})$ in its orbital angular momentum. For a given mode function Q we orient the coordinate system such that the angular dependence is given by Y_{L0} alone. (In flat space this corresponds to choosing the z direction parallel to $\mathbf{k}$.) We can then write

$$ {}_sG_{\ell m} = \left(4\pi \sum_L \sqrt{\frac{2L+1}{2\ell+1}} i^{\ell-L} \phi_{q\,L}(\chi) Y_{L0}(\hat{\mathbf{x}}) \right) {}_sY_{\ell m} \,. \tag{A9.20} $$

Each angular momentum component $Q = \phi_{q\,L}(\chi) Y_{L0}(\hat{\mathbf{x}})$ satisfies

$$ \Delta Q = \gamma^{rs} Q_{|rs} - \gamma^{rs} \Gamma^1_{rs} Q_{,1} + \gamma^{11} Q_{,11} = -|K|(q^2 \mp 1) Q \,, $$

where 1 denotes the χ direction and rs stand for the ϑ and φ directions. In $q^2 \mp 1$, the minus sign is for $K > 0$ and the plus sign for $K < 0$. Denoting $\sin_K = \sin$ for $K > 0$ and $\sin_K = \mathrm{sh}$ for $K < 0$, we have

$$ \Gamma^1_{rs} = -\frac{\sin'_K(\chi)}{\sin_K(\chi)} \gamma_{rs} \,, $$

and $\gamma^{rs} Q_{|rs} = |K| \sin_K^{-2}(\chi) \Delta_{\vartheta,\varphi} Q = -|K| L(L+1) \sin_K^{-2} Q$. With this we obtain the following differential equation for $\phi_{q\,L}$:

$$ \frac{d^2 \phi_{q\,L}}{d\chi^2} + 2 \frac{\cos\chi}{\sin\chi} \frac{d\phi_{q\,L}}{d\chi} + \left(q^2 - 1 - \frac{L(L+1)}{\sin^2\chi} \right) \phi_{q\,L} = 0 \,. $$

In this form, the equation is valid for $K > 0$. For $K < 0$ one has to replace $\sin\chi$ by $\mathrm{sh}\chi$ and $\cos\chi$ by $\mathrm{ch}\chi$ as well as $q^2 - 1$ by $q^2 + 1$. The solutions to this equation, which are regular at $\chi = 0$, are

$$ \phi_{q\,L}(\chi) \propto \begin{cases} -q^{-2} (\sin\chi)^L \frac{d^{L+1}}{d(\cos\chi)^{L+1}} \cos(q\chi) \,, & \text{for } K > 0 \\ -q^{-2} (\mathrm{sh}\chi)^L \frac{d^{L+1}}{d(\mathrm{ch}\chi)^{L+1}} \cos(q\chi) \,, & \text{for } K < 0 \,. \end{cases} \tag{A9.21} $$

These functions can also be expressed in terms of associated Legendre functions, see Abramowitz & Stegun (1970),

$$ \phi_{q\,L}(\chi) \propto \begin{cases} (\sin\chi)^{-1/2} P_{\ell\,m}(\cos\chi) \,, \text{ with} \\ \quad \ell = -1/2 - q \,, \quad m = -\frac{1}{2} - L \,, \quad \text{for} K > 0 \\ (\mathrm{sh}\chi)^{-1/2} P_{\ell\,m}(\mathrm{ch}\chi) \,, \text{ with} \\ \quad \ell = -1/2 + iq \,, \quad m = -\frac{1}{2} - L \,, \quad \text{for } K < 0 \,. \end{cases} \tag{A9.22} $$

This is easily verified by deriving the differential equation for $(\sin\chi)^{1/2} \phi_{q\,L}(\chi)$ and comparing it with the one for associated Legendre functions given in Eq. (A4.15). The hyperspherical Bessel functions $\phi_{q\,L}$ satisfy the recurrence relations

$$ \begin{aligned} \frac{d}{d\chi} \phi_{q\,L} &= \frac{1}{2L+1} \left[L \sqrt{q^2 - L^2} \phi_{q\,L-1} - (L+1) \sqrt{q^2 - (L+1)^2} \phi_{q\,L+1} \right] , \\ \cot\chi \phi_{q\,L} &= \frac{1}{2L+1} \left[\sqrt{q^2 - L^2} \phi_{q\,L-1} + \sqrt{q^2 - (L+1)^2} \phi_{q\,L+1} \right] , \end{aligned} \tag{A9.23} $$

for $K > 0$. For negative curvature, the terms $q^2 - n^2$ have to be replaced by $q^2 + n^2$ and $\cot\chi$ by $\coth(\chi)$. These relations define the hyperspherical Bessel functions in terms of the first member

$$\phi_{q\,0}(\chi) = \begin{cases} \frac{\sin(q\chi)}{q\sin\chi} & \text{for } K > 0 \\ \frac{\mathrm{sh}(q\chi)}{q\,\mathrm{sh}\chi} & \text{for } K < 0\,. \end{cases} \tag{A9.24}$$

The normalization is chosen such that $\lim_{K\to 0}\phi_{q\,L}(\chi) = j_L(kr)$.

We do not need all the details of the mode functions ${}_sG_{\ell m}$, but we have to calculate $n^i[{}_sG_{\ell m}]_{|i}(\mathbf{x},\mathbf{n})$ which enters the Boltzmann equation. To obtain the Boltzmann hierarchy we have to express this derivative in terms of ${}_sG_{\ell+1\,m}$, ${}_sG_{\ell m}$ and ${}_sG_{\ell-1\,m}$. This is obtained most easily if we consider $\mathbf{n} = -\hat{\mathbf{x}}$ so that $\mathbf{x} = -\sqrt{|K|}\chi\mathbf{n}$. We then have

$$n^i[{}_sG_{\ell m}]_{|i}(-\sqrt{|K|}\chi\mathbf{n},\mathbf{n}) = -\sqrt{|K|}\frac{d}{d\chi}[{}_sG_{\ell m}]_{|i}(-\sqrt{|K|}\chi\mathbf{n},\mathbf{n})\,. \tag{A9.25}$$

To calculate this derivative we expand ${}_sG_{\ell m}(-\sqrt{|K|}\chi\mathbf{n},\mathbf{n})$ in its total angular momentum.

$$\begin{aligned} {}_sG_{\ell m}(-\sqrt{|K|}\chi\mathbf{n},\mathbf{n}) &= \left(4\pi\sum_L \sqrt{\frac{2L+1}{2\ell+1}}\,i^{\ell-L}\phi_{q\,L}(\chi)Y_{L0}(\mathbf{n})\right){}_sY_{\ell m}(\mathbf{n}) \\ &= \sum_j (-i)^j\sqrt{4\pi(2j+1)}\,{}_sf_j^{(\ell m)}(\chi)Y_{jm}(\mathbf{n})\,. \end{aligned} \tag{A9.26}$$

For $\ell = 0$, we immediately obtain

$${}_0f_j^{(00)}(\chi,q) \equiv \alpha_j^{(00)}(\chi) = \phi_{q\,j}(\chi)\,. \tag{A9.27}$$

The other coefficients can in principle be obtained with the help of the Clebsch–Gordan series for the products $Y_{L0}(\mathbf{n})\,{}_sY_{\ell m}(\mathbf{n})$ and the recurrence relations for the hyperspherical Bessel functions $\phi_{q\,L}$. This straightforward but cumbersome calculation has never appeared in print, and we do not want to break with this tradition here. It is much easier to use the fact that ${}_0G_{m\,m}$ are given by (see e.g. Thorne (1980) and Maggiore (2007))

$${}_0G_{m\,m} = n^{i_1}\cdots n^{i_m}Q^{(m)}_{i_1\cdots i_m}\,,$$

and

$${}_{\pm2}G_{2\,m} \propto \mathbf{e}^{\pm}_{i_1}\mathbf{e}^{\pm}_{i_2}Q^{(m)}_{i_1 i_2}\,,$$

for $0 \le |m| \le 2$. With this it is relatively easy to derive relations between the hyperspherical Bessel functions and the coefficients ${}_sf_j^{(\ell m)}(\chi)$. Most importantly

$$\begin{aligned} \alpha_j^{(11)}(\chi,q) &= \sqrt{\frac{j(j+1)}{2(q^2-1)}}\,\mathrm{sc}(\chi)\,\phi_{q\,j}(\chi)\,, \\ \alpha_j^{(22)}(\chi,q) &= \sqrt{\frac{3}{8}\frac{(j+2)(j^2-1)j}{(q^2-4)(q^2-1)}}\,\mathrm{sc}^2(\chi)\,\phi_{q\,j}(\chi)\,. \end{aligned} \tag{A9.28}$$

Similarly for

$${}_{\pm2}f_j^{(2m)} = \epsilon_j^{(m)} \pm i\beta_j^{(m)}\,,$$

one finds

$$\begin{aligned}
\epsilon_j^{(0)}(\chi,q) &= \sqrt{\frac{3}{8}\frac{(j+2)(j^2-1)j}{(q^2-4)(q^2-1)}}\,\mathrm{sc}^2(\chi)\phi_{qj}(\chi)\,,\\
\epsilon_j^{(1)}(\chi,q) &= \frac{1}{2}\sqrt{\frac{(j-1)(j+2)}{(q^2-4)(q^2-1)}}\,\mathrm{sc}\chi\left[\cot(\chi)\phi_{q\,j}(\chi)+\phi'_{qj}(\chi)\right]\,,\\
\epsilon_j^{(2)}(\chi,q) &= \frac{1}{4}\sqrt{\frac{1}{(q^2-4)(q^2-1)}}\left[\phi''_{q\,j}(\chi)\right.\\
&\quad \left.+4\cot(\chi)\phi'_{q\,j}(\chi)-\left(q^2+1-2\cot^2\chi\right)\phi_{q\,j}(\chi)\right]\,,
\end{aligned} \tag{A9.29}$$

and

$$\begin{aligned}
\beta_j^{(0)}(\chi,q) &= 0\,,\\
\beta_j^{(1)}(\chi,q) &= \frac{1}{2}\sqrt{\frac{(j-1)(j+2)q^2}{(q^2-4)(q^2-1)}}\,\mathrm{sc}(\chi)\phi_{q\,j}(\chi)\,,\\
\beta_j^{(2)}(\chi,q) &= \frac{1}{2}\sqrt{\frac{q^2}{(q^2-4)(q^2-1)}}\left[\phi'_{q\,j}(\chi)+2\cot(\chi)\phi_{q\,j}(\chi)\right]\,,
\end{aligned} \tag{A9.30}$$

for $m>0$. For $m<0$, $\beta_j^{(-m)}=-\beta_j^{(m)}$ while $\epsilon_j^{(m)}=\epsilon_j^{(-m)}$ and $\alpha_j^{(m)}=\alpha_j^{(-m)}$. Also here, the formulae presented are for positive curvature. For negative curvature all terms of the form q^2-n^2 have to be replaced by q^2+n^2, and the trigonometric functions have to be replaced by hyperbolic functions, e.g. $\mathrm{sh}\,\chi\equiv 1/\sin\chi$ becomes $\mathrm{sch}\chi\equiv 1/\mathrm{sh}\,\chi$. The overall normalization of the modes is chosen such that

$${}_sf_j^{(\ell m)}(0,q)=\frac{1}{2j+1}\delta_{j\ell}\,. \tag{A9.31}$$

From the recurrence relation for the hyperspherical Bessel functions we obtain the following relation for the coefficients ${}_sf_j^{(\ell m)}$ defined so far:

$$\frac{d}{d\chi}[{}_sf_j^{(\ell m)}]=\frac{q}{2j+1}\left[{}_s\theta_j^m\;{}_sf_{j-1}^{(\ell m)}-{}_s\theta_{j+1}^m\;{}_sf_{j+1}^{(\ell m)}\right]-i\frac{qms}{j(j+1)}\;{}_sf_j^{(\ell m)}\,, \tag{A9.32}$$

where

$${}_s\theta_j^m=\sqrt{\left[\frac{(j^2-m^2)(j^2-s^2)}{j^2}\right]\left(\frac{j^2}{q^2}\pm 1\right)}\,, \tag{A9.33}$$

here the $+$ sign is for negative curvature and the $-$ sign is for $K>0$.

Since the relation (A9.32) is independent of ℓ it is valid for all ℓs and can also be used to define ${}_sf_{j+1}^{(\ell m)}$ from ${}_sf_j^{(\ell m)}$ and ${}_sf_{j-1}^{(\ell m)}$.

With the expansion Eqs. (A9.26) and (A9.32) we can now write the derivative (A9.25) as

$$n^i[{}_sG_{\ell m}]_{|i}=-\frac{\sqrt{|K|}q}{2\ell+1}\left[{}_s\theta_\ell^m\;{}_sG_{\ell-1\,m}-{}_s\theta_{\ell+1}^m\;{}_sG_{\ell+1\,m}\right]+i\frac{\sqrt{|K|}qms}{\ell(\ell+1)}\;{}_sG_{\ell m}\,. \tag{A9.34}$$

Like in flat space, we expand the CMB anisotropy and polarization as

$$\mathcal{M}(t,\mathbf{x},\mathbf{n}) = \sum\!\!\!\!\!\!\int \frac{d^3q}{(2\pi)^3} \sum_\ell \sum_{m=-2}^{2} \mathcal{M}_\ell^{(m)}{}_0G_{\ell m}\ , \tag{A9.35}$$

$$(Q \pm iU)(t,\mathbf{x},\mathbf{n}) = \sum\!\!\!\!\!\!\int \frac{d^3q}{(2\pi)^3} \sum_\ell \sum_{m=-2}^{2} (\mathcal{E}_\ell^{(m)} \pm i\mathcal{B}_\ell^{(m)})_{\pm2}G_{\ell m}\ . \tag{A9.36}$$

The sign $\sum\!\!\!\!\!\!\int$ indicates that for positive curvature the integral over q has to be replaced by a sum.

For the coefficients $\mathcal{M}_\ell^{(m)}(t,q)$, $\mathcal{E}_\ell^{(m)}(t,q)$ and $\mathcal{B}_\ell^{(m)}(t,q)$ we now obtain the desired Boltzmann hierarchy

$$\begin{aligned}\dot{\mathcal{M}}_\ell^{(m)} - q\sqrt{|K|}\left[{}_0\theta_\ell^m \mathcal{M}_{\ell-1}^{(m)} - {}_0\theta_{\ell+1}^m \mathcal{M}_{\ell+1}^{(m)}\right] \\ = S_\ell^{(m)} + \dot\kappa\left[P_\ell^{(m)} - \mathcal{M}_\ell^{(m)}\right]\ ,\end{aligned} \tag{A9.37}$$

with

$$S_\ell^{(0)} = -k(\Psi+\Phi)\delta_{\ell1}\ , \tag{A9.38}$$

$$S_\ell^{(\pm1)} = -\frac{\sqrt{3}}{3k}\sqrt{k^2-2K}\,\sigma_\pm\,\delta_{\ell2}\ , \tag{A9.39}$$

$$S_\ell^{(\pm2)} = \frac{1}{\sqrt{3}}\dot H_{\pm2}\,\delta_{\ell2}\ , \tag{A9.40}$$

$$P_\ell^{(0)} = \mathcal{M}_0^{(0)}\delta_{\ell0} + V^{(b)}\delta_{\ell1} + \frac{1}{10}[\mathcal{M}_2^{(0)} - \sqrt{6}E_2^{(0)}]\delta_{\ell2}\ , \tag{A9.41}$$

$$P_\ell^{(\pm1)} = V_b^{(\pm1)}\delta_{\ell1} + \frac{1}{10}[\mathcal{M}_2^{(\pm1)} - \sqrt{6}E_2^{(\pm1)}]\delta_{\ell2}\ , \tag{A9.42}$$

$$P_\ell^{(\pm2)} = \frac{1}{10}[\mathcal{M}_2^{(\pm2)} - \sqrt{6}E_2^{(\pm2)}]\delta_{\ell2}\ . \tag{A9.43}$$

As in Chapter 5, the superscript (m) indicates scalar perturbations for $m=0$, vector perturbations for $m=\pm1$ and tensor perturbations for $m=\pm2$. For Eqs. (A9.38) and (A9.39) we made use of

$$-n^i[{}_0G_{00}]_{|i} = k\,{}_0G_{10} \quad \text{and}$$

$$n^i n^j Q_{i|j}^{(m)} = n^i[{}_0G_{1m}]_{|i} = \frac{\sqrt{3}}{3}\sqrt{k^2-2K}\ {}_0G_{2m}\ ,$$

for $m=\pm1$.

The Boltzmann hierarchy for E- and B-polarization becomes

$$\begin{aligned}\dot{\mathcal{E}}_\ell^{(m)} = q\sqrt{|K|}\left[\frac{{}_2\theta_\ell^m}{(2\ell-1)}\mathcal{E}_{\ell-1}^{(m)} - \frac{2m}{\ell(\ell+1)}B_\ell^{(m)} - \frac{{}_2\theta_{\ell+1}^m}{(2\ell+3)}\mathcal{E}_{\ell+1}^{(m)}\right] \\ -\dot\kappa\left(\mathcal{E}_\ell^{(m)} + \frac{\sqrt{6}}{10}\left[\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)}\right]\delta_{\ell,2}\right)\ ,\end{aligned} \tag{A9.44}$$

$$\begin{aligned}\dot{\mathcal{B}}_\ell^{(m)} = q\sqrt{|K|}\left[\frac{{}_2\theta_\ell^m}{(2\ell-1)}\mathcal{B}_{\ell-1}^{(m)} + \frac{2m}{\ell(\ell+1)}\mathcal{E}_\ell^{(m)} - \frac{{}_2\theta_{\ell+1}^m}{(2\ell+3)}\mathcal{B}_{\ell+1}^{(m)}\right] \\ -\dot\kappa\mathcal{B}_\ell^{(m)}.\end{aligned} \tag{A9.45}$$

A fast Boltzmann code (such as CMBfast) calculates only the lowest few (about ten) modes with the Boltzmann hierarchy and then uses the results $\ell = 0$, 1 and 2 as input for the integral solutions. These are obtained exactly like in the flat case (see Chapter 5) by replacing the flat radial functions by the ones obtained for curved spaces.

$$\frac{\mathcal{M}_\ell^{(m)}(t_0,q)}{\ell+1} = \int_0^{t_0} dt\, e^{-\kappa} \sum_{j=0}^{2} [S_j^{(m)} + \dot{\kappa}\, P_j^{(m)}] \alpha_\ell^{(jm)}(\sqrt{|K|}t, q)\,, \tag{A9.46}$$

$$\frac{\mathcal{E}_\ell^{(m)}(t_0,q)}{\ell+1} = \int_0^{t_0} dt\, e^{-\kappa} \dot{k} \frac{\sqrt{6}}{10} \left[\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)}\right] \epsilon_\ell^{(m)}(\sqrt{|K|}t, q)\,, \tag{A9.47}$$

$$\frac{\mathcal{B}_\ell^{(m)}(t_0,q)}{\ell+1} = \int_0^{t_0} dt\, e^{-\kappa} \dot{k} \frac{\sqrt{6}}{10} \left[\mathcal{M}_2^{(m)} - \sqrt{6}\mathcal{E}_2^{(m)}\right] \beta_\ell^{(m)}(\sqrt{|K|}t, q)\,. \tag{A9.48}$$

A9.4 The energy–momentum tensor

The perturbations of the energy–momentum tensor of radiation which enter the Einstein equations are obtained from their definitions by integration over the directions **n**,

$$D_g^{(r)} = 4\mathcal{M}_0^{(0)}\,, \tag{A9.49}$$

$$V_r^{(m)} = \mathcal{M}_1^{(m)}\,, \tag{A9.50}$$

$$\sqrt{1 - \frac{3K}{k^2}}\, \Pi_r^{(0)} = \frac{12}{5}\mathcal{M}_2^{(0)}\,, \tag{A9.51}$$

$$\sqrt{1 - \frac{2K}{k^2}}\, \Pi_r^{(1)} = \frac{8\sqrt{3}}{5}\mathcal{M}_2^{(1)}\,, \tag{A9.52}$$

$$\Pi_r^{(2)} = \frac{8}{5}\mathcal{M}_2^{(2)}\,. \tag{A9.53}$$

A9.5 Power spectra

In the derivation of the power spectra the only change w.r.t. flat space is that for positive curvature the integral over q has to be replaced by a sum. Note also that our variable q is dimensionless and therefore so are our amplitudes $X_\ell^{(m)}$

$$(2\ell+1)^2 C_\ell^{XY} = \frac{2}{\pi} \sum\!\!\!\!\!\!\int \frac{dq}{q} \sum_{m=-2}^{2} q^3 P_{\ell m}^{(XY)}\,, \tag{A9.54}$$

where X, Y are $\mathcal{M}$, $\mathcal{E}$ or $\mathcal{B}$ and the power spectra are defined like in the flat case,

$$\langle \mathcal{M}_\ell^{(m)}(q) \mathcal{M}_\ell^{(m)*}(q') \rangle \equiv (2\pi)^3 \delta_{q,q'} M_\ell^{(m)}(q)\,, \tag{A9.55}$$

$$\langle \mathcal{E}_\ell^{(m)}(q) \mathcal{E}_\ell^{(m)*}(q') \rangle \equiv (2\pi)^3 \delta_{q,q'} E_\ell^{(m)}(q)\,, \tag{A9.56}$$

$$\langle \mathcal{B}_\ell^{(m)}(q) \mathcal{B}_\ell^{(m)*}(q') \rangle \equiv (2\pi)^3 \delta_{q,q'} B_\ell^{(m)}(q)\,, \tag{A9.57}$$

$$\langle \mathcal{E}_\ell^{(m)}(q) \mathcal{M}_\ell^{(m)*}(q') \rangle \equiv (2\pi)^3 \delta_{q,q'} F_\ell^{(m)}(q)\,. \tag{A9.58}$$

The formulae here are written for positive curvature. For negative curvature the Kronecker δ becomes a Dirac δ-function like in flat space. Hence $P^{(\mathcal{M}\mathcal{M})}_{\ell m}(q) = M^{(m)}_{\ell}(q)$, $P^{(\mathcal{E}\mathcal{E})}_{\ell m}(q) = E^{(m)}_{\ell}(q)$, $P^{(\mathcal{B}\mathcal{B})}_{\ell m}(q) = B^{(m)}_{\ell}(q)$ and $P^{(\mathcal{E}\mathcal{M})}_{\ell m}(q) = F^{(m)}_{\ell}(q)$. Due to their different parity, $\mathcal{M}$ and $\mathcal{B}$ as well as $\mathcal{E}$ and $\mathcal{B}$ are uncorrelated.

Appendix 10

The solutions of some exercises

A10.1 Chapter 1

Exercise 1.4

In a dust universe with curvature and with a cosmological constant the Friedmann equation can be written in the form

$$\dot{a}^2 = a^2\left[-K + \frac{C}{a} + \frac{1}{3}\Lambda a^2\right] \equiv G(a). \tag{A10.1}$$

Here

$$C = \frac{8\pi G}{3}\rho_m a^3 = \Omega_m H_0^2 a_0^3 = \begin{cases} \frac{\Omega_m}{H_0|\Omega_k|^{3/2}} = \frac{2q_0}{H_0|1-2q_0|^{3/2}} & \text{if } \Omega_k \neq 0 \\ \Omega_m H_0^2 & \text{if } \Omega_k = 0 \\ & \text{and } a_0 = 1. \end{cases} \tag{A10.2}$$

If the curvature is negative and $\Lambda > 0$, G is strictly positive and we find an expanding solution for all times. At late times, curvature becomes negligible and the universe expands like $a \propto 1/|t| \propto \exp(\sqrt{\Lambda/3}\tau)$. If $\Lambda < 0$ the square bracket is decreasing and G has a zero, $G(a_c) = 0$. At this point expansion turns into contraction and the universe recollapes.

The case $K = 0$ can be solved explicitly leading to

$$a^3(\tau) = \begin{cases} \frac{3C}{2\Lambda}\left(\cosh(\sqrt{3\Lambda}t) - 1\right) & \Lambda > 0\,,\, a_{\min} = \left(3C/2\Lambda\right)^{1/3} \\ \frac{-3C}{2\Lambda}\left(1 - \cos\left(\sqrt{-3\Lambda}t\right)\right) & \Lambda < 0. \end{cases} \tag{A10.3}$$

The qualitative behaviour is like for $K < 0$.

The case $K > 0$ is most interesting. The function G can be written as $G(a) = aP(a)$, where P is a third-order polynomial which has one or three real roots. In the dashed region of Fig. A10.1, P has one real root, but for a negative value of a. Hence the universe expands forever. In the upper left region, high cosmological constant, the scale factor has a minimum. Such a universe has no big bang but comes out of a previous contracting phase. It is called a bouncing solution. For a value of $\Omega_m > 0.01$ one finds a maximum redshift $z_{\max} < 4$ for a bouncing universe. Hence they cannot explain cosmological data like quasars and galaxies at a redshift of 6 or even the CMB. Solutions below the dashed region emerge from a big bang but recollapse eventually, when either the negative cosmological constant or the positive curvature term render $G(a_{\max}) = 0$.

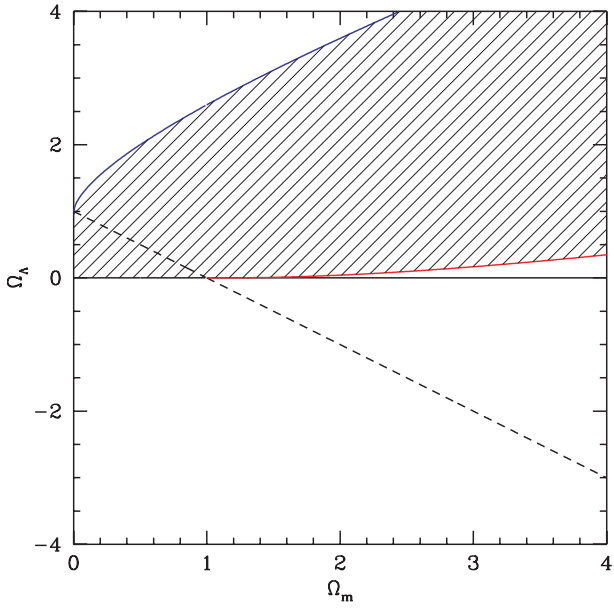


Fig. A10.1. The kinematics of a universe with matter density parameter Ω_m and cosmological constant parameter Ω_Λ. The universes with parameters above the dashed line are positively curved, those below negatively. The universes with values $(\Omega_m, \Omega_\Lambda)$ in the dashed region recollapse into a big crunch. Those below emerge from a big bang, those above emerge from a collapsing universe, they have no big bang in the past.

A10.2 Chapter 2

Exercise 2.1

We want to show that

$$L_X g = a^2 \left[-2 \left(\frac{\dot{a}}{a} T + \dot{T} \right) dt^2 + 2(\dot{L}_i - T_{,i}) dt\, dx^i + \left(2 \frac{\dot{a}}{a} T \gamma_{ij} + L_{i|j} + L_{j|i} \right) dx^i\, dx^j \right], \tag{A10.4}$$

for $X = T\partial_t + L^i \partial_i$ and $g = a^2(t)[-dt^2 + \gamma_{ij}\, dx^i\, dx^j] = a^2(t) S_{\mu\nu}\, dx^\mu\, dx^\nu$.

We use $L_X a^2 = 2\dot{a} a T$ and $L_X(a^2 S) = L_X(a^2) S + a^2 L_X S$. Furthermore, we show below, that for an arbitrary metric S, we have

$$(L_X S)_{\mu\nu} = X_{\mu;\nu} + X_{\nu;\mu}, \tag{A10.5}$$

where here ; denotes the covariant derivative w.r.t. the metric S. For our metric S all Christoffel symbols involving a '0' vanish, so that $X_{\nu;0} = X_{\nu,0}$ and $X_{0;\nu} = X_{0,\nu}$. Furthermore $X_{i;j} = X_{i|j}$, where $|$ denotes the covariant derivative w.r.t. the

three-dimensional metric γ. With this we obtain

$$L_X g = 2\frac{\dot{a}}{a} T a^2 S + a^2 \left(-2\dot{T}\, dt^2 - 2(T, i - \dot{L}_i)\, dt\, dx^i + (L_{i|j} + L_{j|i})\, dx^i\, dx^j\right) \tag{A10.6}$$

which agrees with Eq. (A10.4). It remains to show Eq. (A10.5). For this we use the general expression (A2.20). For a doubly covariant tensor field this gives

$$\begin{aligned}(L_X S)_{\alpha\beta} &= X^\mu S_{\alpha\beta},_\mu + X^\mu,_\alpha\, S_{\mu\beta} + X^\mu,_\beta\, S_{\mu\alpha} \\ &= X_\nu \left(S^{\mu\nu} S_{\alpha\beta},_\mu + S^{\mu\nu},_\alpha\, S_{\mu\beta} + S^{\mu\nu},_\beta\, S_{\mu\alpha}\right) + X_{\alpha,\beta} + X_{\beta,\alpha}.\end{aligned}$$

For the last equals sign we simply inserted $X^\mu = X_\nu S^{\nu\mu}$. We now take the derivative of the identity $S^{\nu\mu} S_{\mu\beta} = \delta^\nu_\beta$ w.r.t. α. This yields $S^{\mu\nu},_\alpha\, S_{\mu\beta} = -S^{\mu\nu} S_{\mu\beta,\alpha}$. Correspondingly $S^{\mu\nu},_\beta\, S_{\mu\alpha} = -S^{\mu\nu} S_{\mu\alpha,\beta}$. Inserting this above and using the definition

$$S_{\alpha;\beta} = S_{\alpha,\beta} - \Gamma^\mu_{\alpha\beta} S_\mu \text{ with } \quad \Gamma^\beta_{\mu\nu} = \frac{1}{2} S^{\beta\alpha}\left(S_{\mu\alpha,\nu} + S_{\nu\alpha,\mu} - S_{\mu\nu,\alpha}\right)\ ,$$

we obtain (A10.5).

Exercise 2.3

We consider a perturbed FL universe containing two non-interacting fluids with energy densities ρ_α and pressure P_α. The total energy density and pressure are $\rho = \rho_1 + \rho_2$ and $P = P_1 + P_2$. We first note that for both components the intrinsic entropy perturbation is given by

$$\Gamma_\alpha = \pi^{(\alpha)}_L - \frac{c^2_\alpha}{w_\alpha}\delta_\alpha = \frac{\delta P_\alpha}{P_\alpha} - c^2_\alpha \frac{\delta\rho_\alpha}{P_\alpha} \tag{A10.7}$$

and the total sound speed is

$$c^2_s = \frac{\dot{P}_1 + \dot{P}_2}{\dot{\rho}} = \frac{c^2_1\dot{\rho}_1 + c^2_2\dot{\rho}_2}{\dot{\rho}} = \frac{(1+w_1)c^2_1\rho_1 + (1+w_2)c^2_2\rho_2}{(1+w)\rho}\ . \tag{A10.8}$$

For the second equality sign we have used that both components are separately conserved. Defining now $R = \rho_2/\rho$, so that $\rho_1/\rho = 1 - R$ we can also write

$$(1+w)c^2_s = (1+w_1)c^2_1(1-R) + c^2_2(1+w_2)R\ . \tag{A10.9}$$

Let us first assume $\Gamma_\alpha = 0$, so that $\delta P_\alpha = c^2_\alpha \delta\rho_\alpha$. The total entropy perturbation is then given by $\Gamma = \Gamma_{\rm rel}$ with

$$P\Gamma_{\rm rel} = c^2_1\delta\rho_1 + c^2_2\delta\rho_2 - c^2_s(\delta\rho_1 + \delta\rho_2) = \left(c^2_1 - c^2_s\right)\delta\rho_1 + \left(c^2_2 - c^2_s\right)\delta\rho_2\ . \tag{A10.10}$$

To express $\Gamma_{\rm rel}$ in terms of gauge-invariant variables we now use

$$\delta\rho_\alpha = [D^{(\alpha)}_g + (1+w_\alpha)(3H_L + H_T)]\rho_\alpha\ .$$

Inserting this in Eq. (A10.10) yields

$$\begin{aligned}w\Gamma_{\rm rel} = &\left(c^2_1 - c^2_s\right)(1-R)D^{(1)}_g + \left(c^2_2 - c^2_s\right)R D^{(2)}_g + (3H_L + H_T) \\ &\times\left[\left(c^2_1 - c^2_s\right)(1-R)(1+w_1) + \left(c^2_2 - c^2_s\right)R(1+w_2)\right]\ .\end{aligned} \tag{A10.11}$$

Using Eq. (A10.9) and

$$1+w=\frac{\rho+P}{\rho}=\frac{\rho_1+P_1+\rho_2+P_2}{\rho}=(1+w_1)(1-R)+(1+w_2)R\,,$$

we find that the square bracket above vanishes and $\Gamma_{\rm rel}$ is gauge invariant, as it should be. In fact, with the relation (A10.9)

$$[\quad]=c_1^2(1-R)(1+w_1)+c_2^2R(1+w_2)-c_s^2(1+w)=0\,.$$

Multiplying Eq. (A10.11) with $1+w$ and using Eq. (A10.9) to replace c_s^2 finally leads to

$$w(1+w)\Gamma_{\rm rel}=R(1-R)\big(c_1^2-c_2^2\big)\left[(1+w_2)D_g^{(1)}-(1+w_1)D_g^{(2)}\right]\,. \tag{A10.12}$$

From this equation we already conclude that $\Gamma_{\rm rel}$ vanishes if both sound speeds are equal, $c_1^2=c_2^2$ or if one of the two components is largely subdominant, $R\simeq 0$ or $R\simeq 1$. If none of these conditions is fulfilled, perturbations are adiabatic if

$$(1+w_2)D_g^{(1)}=(1+w_1)D_g^{(2)}\quad\text{(adiabaticity)}\,. \tag{A10.13}$$

To determine Γ when $\Gamma_\alpha\neq 0$ we simply note that in this case $\delta P_\alpha=P_\alpha\Gamma_\alpha+c_\alpha\delta\rho_\alpha$ so that

$$P\Gamma=P_1\Gamma_1+P_2\Gamma_2+P\Gamma_{\rm rel}\,.$$

Inserting our result for $\Gamma_{\rm rel}$ we find

$$\Gamma=\frac{w_1}{w}(1-R)\Gamma_1+\frac{w_2}{w}R\Gamma_2+\Gamma_{\rm rel}\,. \tag{A10.14}$$

We now want to derive an evolution equation for $\Gamma_{\rm rel}$ in the case where $\Gamma_\alpha=0$ and $w_\alpha=c_\alpha^2=$ constant for both components. We use the conservation equation (2.114) which in this case reduces to

$$\dot{D}_g^{(\alpha)}=-k(1+w_\alpha)V_\alpha\,. \tag{A10.15}$$

Defining

$$f=\frac{R(1-R)}{w(1+w)}\big(c_1^2-c_2^2\big)\,,$$

the derivative of $\Gamma_{\rm rel}$ can be written as

$$\dot{\Gamma}_{\rm rel}=\frac{\dot{f}}{f}\Gamma_{\rm rel}+kf(1+w_1)(1+w_2)[V_2-V_1]\,. \tag{A10.16}$$

This shows that even if perturbations of a two-component fluid are initially adiabatic, they develop a relative entropy perturbation if $V_1\neq V_2$. This is already clear from the adiabaticity condition (A10.13), which cannot be maintained if $V_1\neq V_2$ due to the time evolution of $D_g^{(\alpha)}$ given in Eq. (A10.15). Especially, on sub-Hubble scales, where V_1 and V_2 evolve differently (we consider the non-trivial case $c_1\neq c_2$), adiabaticity between different components cannot be maintained. When talking about adiabatic perturbations, we therefore always refer to super-Hubble scales.

A10.3 Chapter 3

Exercise 3.1

We want to show that only exponential potentials allow for power law inflation, $a \propto t^q$ with some constant q and we want to express q in terms of the parameters of the potential. We assume a spatially flat FL universe, $K = 0$.

For a spatially flat FL universe, the Friedmann equation and energy–momentum conservation imply

$$\dot{\mathcal{H}} = -\frac{1+3w}{2}\mathcal{H}^2 \,.$$

Now if $a \propto t^q$ we have $\mathcal{H} = q/t$ and $\dot{\mathcal{H}} = -q/t^2$. Inserting this above gives

$$q = \frac{2}{1+3w} \quad \text{hence} \quad w = \frac{2-q}{3q} = \text{constant} \,.$$

Furthermore, integrating $d\tau = a\,dt \propto t^q\,dt$ yields $\tau \propto t^{q+1}$, hence

$$a \propto \tau^p \quad \text{with} \quad p = \frac{q}{q+1} = \frac{2}{3+3w} \,.$$

Since $w = P/\rho$ has to be constant if q is constant, we have

$$\frac{(\dot{\rho} - \dot{P})}{\rho - P} = \frac{\dot{\rho}}{\rho} = -3(1+w)\mathcal{H} \,.$$

But $\rho - P = 2W$ so that

$$\frac{(\dot{\rho} - \dot{P})}{\rho - P} = \frac{W_{,\varphi}\dot{\varphi}}{W} = -3(1+w)\mathcal{H} \quad \text{hence} \quad \frac{W_{,\varphi}}{W} = -3(1+w)\frac{\mathcal{H}}{\dot{\varphi}} \,.$$

Here, $W_{,\varphi} = \frac{dW}{d\varphi}$. The same procedure for $\rho + P = (\dot{\varphi})^2/a^2$ yields

$$\ddot{\varphi} = -\frac{1+3w}{2}\mathcal{H}\dot{\varphi} \,.$$

Using these results to replace $\ddot{\varphi}$ and $W_{,\varphi}$ in the equation of motion for φ,

$$\ddot{\varphi} + 2\mathcal{H}\dot{\varphi} + a^2 W_{,\varphi} = 0 \quad \text{gives} \quad \frac{1}{2}(1-w)\dot{\varphi}^2 = (1+w)a^2 W \,.$$

Inserting this in the Friedmann equation, we obtain

$$\mathcal{H}^2 = \frac{4\pi}{3m_p^2}\left(1 + \frac{1-w}{1+w}\right)\dot{\varphi}^2 \,; \quad \mathcal{H} = A\frac{\dot{\varphi}}{m_P} \quad \text{with} \quad A = \sqrt{\frac{8\pi}{3(1+w)}}$$

and $W_{,\varphi}/W = -3(1+w)A/m_P = \text{constant}$. Hence we obtain an exponential potential

$$W = W_0 \exp\left(-\alpha\frac{\varphi}{m_P}\right) \quad \text{with} \quad \alpha = \sqrt{24\pi(1+w)} = 4\sqrt{\pi\,\frac{1+q}{q}} = 4\sqrt{\frac{\pi}{p}} \,.$$

To obtain inflationary expansion we need $p > 1$, i.e. $q < -1$ or, equivalently $w < -\frac{1}{3}$. This is equivalent to $\alpha < 4\sqrt{\pi}$. Exponential inflation is obtained in the limit $q \to -1$ which is equivalent to $p \to \infty$ and $\alpha \to 0$, $W = W_0 = \text{constant}$.

A10.4 Chapter 4

Exercise 4.3

We start with Eq. (4.135) which yields

$$\frac{1}{4\pi}\sum_\ell (2\ell+1)C_\ell^{(V)}P_\ell(\mathbf{n}\cdot\mathbf{n}') = \sum_\ell \frac{(2\ell+1)^2}{(2\pi)^3}\int d^3k\, M_\ell^{(V)}(k)P_\ell(\mu)P_\ell(\mu')(\mathbf{n}\cdot\mathbf{n}'-\mu\mu')\,.$$

For the last factor we made use of Eq. (4.136). Before we continue we now show Eq. (4.191). The addition theorem of spherical harmonics yields

$$\int d\Omega_{\hat{\mathbf{k}}}\, P_\ell(\mu)P_{\ell'}(\mu') = \frac{(4\pi)^2}{(2\ell+1)(2\ell'+1)}\sum_{mm'}\int d\Omega_{\hat{\mathbf{k}}}\, Y_{\ell m}(\hat{\mathbf{k}})\, Y^*_{\ell m}(\mathbf{n})Y^*_{\ell' m'}(\hat{\mathbf{k}})Y_{\ell' m'}(\mathbf{n}').$$

Using the orthogonality of spherical harmonics this implies

$$\int d\Omega_{\hat{\mathbf{k}}}\, P_\ell(\mu)P_{\ell'}(\mu') = \delta_{\ell\ell'}\frac{(4\pi)^2}{(2\ell+1)^2}\sum_m Y^*_{\ell m}(\mathbf{n})Y_{\ell m}(\mathbf{n}') = \frac{4\pi}{2\ell+1}P_\ell(\mathbf{n}\cdot\mathbf{n}')\,.$$

For the last equals sign we have again applied the addition theorem. With the help of the recursion relation

$$\mu P_\ell(\mu) = \frac{\ell+1}{2\ell+1}P_{\ell+1}(\mu) + \frac{\ell}{2\ell+1}P_\ell(\mu)\,,$$

we can now perform the angular integration,

$$\begin{aligned}
&\int d^3k\, M_\ell^{(V)}(k)P_\ell(\mu)P_\ell(\mu')(\mathbf{n}\cdot\mathbf{n}'-\mu\mu') \\
&= 4\pi\int dk\, k^2 M_\ell^{(V)}(k)\left[\frac{1}{2\ell+1}(\mathbf{n}\cdot\mathbf{n}')P_\ell(\mathbf{n}\cdot\mathbf{n}')\right. \\
&\left. - \frac{(\ell+1)^2}{(2\ell+1)^2(2\ell+3)}P_{\ell+1}(\mathbf{n}\cdot\mathbf{n}') - \frac{\ell^2}{(2\ell+1)^2(2\ell-1)}P_{\ell-1}(\mathbf{n}\cdot\mathbf{n}')\right] \\
&= \frac{4\pi}{(2\ell+1)^2}\int dk\, k^2 M_\ell^{(V)}(k)\left[\frac{(\ell+1)(\ell+2)}{2\ell+3}P_{\ell+1}(\mathbf{n}\cdot\mathbf{n}') + \frac{\ell(\ell-1)}{2\ell-1}P_{\ell-1}(\mathbf{n}\cdot\mathbf{n}')\right].
\end{aligned}$$

Identifying the coefficient of P_ℓ finally results in

$$C_\ell = \frac{2\ell(\ell+1)}{\pi(2\ell+1)^2}\int dk\, k^2\left[M_{\ell+1}^{(V)}(k) + M_{\ell-1}^{(V)}(k)\right]\,. \tag{A10.17}$$

A10.5 Chapter 6

Exercise 6.2

We parametrize the initial conditions by

$$C_{ij} = \langle X_i(\mathbf{k}) X_j^*(\mathbf{k}') \rangle = A_{ij} (k/H_0)^{n_{ij}} \delta(\mathbf{k} - \mathbf{k}') \,.$$

Clearly, for C_{ij} to be positive semi-definite for all values of k, the matrix $A_{ij} = C_{ij}(k = H_0)$ has to be positive semi-definite. Let us now consider $i \neq j$ with $A_{ij} \neq 0$. If neither $n_{ii} \leq n_{ij} \leq n_{jj}$ nor $n_{jj} \leq n_{ij} \leq n_{ii}$ is true, n_{ij} is either the largest or the smallest of these three spectral indices. Let us first assume it to be the smallest. In order to show that C_{ij} is not positive semi-definite, we have to find a vector V so that $C_{mn} V^m V^n < 0$. If $A_{ij} > 0$, we choose $V^i = -V^j = 1$, and if $A_{ij} < 0$, we choose $V^i = V^j = 1$, so that $A_{ij} V^i V^j = -|A_{ij}|$ (no sum!). Since n_{ij} is smaller than n_{ii} and n_{jj} we can choose k to be sufficiently small so that $|A_{ij}|(k/H_0)^{n_{ij}} \gg |A_{ii}|(k/H_0)^{n_{ii}}$ and $|A_{ij}|(k/H_0)^{n_{ij}} \gg |A_{jj}|(k/H_0)^{n_{jj}}$. Setting all other components of V to zero we obtain for such values of k

$$\sum_{mn} V^m V^n C_{mn}(k) = -|A_{ij}|(k/H_0)^{n_{ij}} + A_{ii}(k/H_0)^{n_{ii}} + A_{jj}(k/H_0)^{n_{jj}} < 0 \,.$$

If n_{ij} is larger than n_{ii} and n_{jj} we just have to choose k sufficiently large.

A10.6 Chapter 7

Exercise 7.1

We consider a mass M positioned at $\mathbf{x} = 0$ with gravitational potential $\Psi = GM/r$. To first order in Ψ the corresponding metric is given by

$$ds^2 = -(1 + 2\Psi)\, dt^2 + (1 - 2\Psi)\, d\mathbf{x}^2 \,.$$

We want to determine the deflection of a photon in this metric. Angles are invariant under conformal transformations of the geometry. We may therefore calculate the deflection in the conformally related metric $d\tilde{s}^2 = (1 + 2\Psi)\, ds^2$. To first order in Ψ we have

$$d\tilde{s}^2 = -(1 + 4\Psi)\, dt^2 + d\mathbf{x}^2 \,.$$

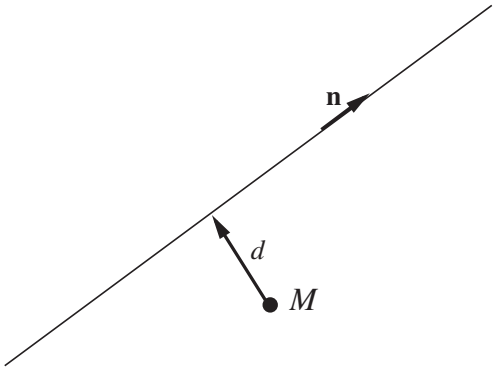


Fig. A10.2. A photon passing the mass M in direction $\mathbf{n}$ with impact parameter d.

We consider a photon along the unperturbed path $\mathbf{x}(s) = d\mathbf{e} + s\mathbf{n}$. The spatial unit vector $\mathbf{n}$ is the direction of motion of the photon and $\mathbf{e}$ is a spatial unit vector normal to $\mathbf{n}$. Hence d is the impact parameter, i.e. the closest distance of the photon from the mass M at $\mathbf{x} = 0$, see Fig. A10.2. The unperturbed photon velocity is given by $(n^\mu) = (1, \mathbf{n})$. Since Ψ is spherically symmetric, angular momentum is conserved and also the perturbed motion will be in the plane $(\mathbf{e}, \mathbf{n})$. We define the perturbed velocity by

$$(n^\mu + \delta n^\mu) = (1 + \delta n^0, \mathbf{n} + \delta\mathbf{n}) \,.$$

As it lies in the plane $(\mathbf{e}, \mathbf{n})$, the spatial part of δn^μ is of the form $\delta\mathbf{n} = \varphi\mathbf{e} + \alpha\mathbf{n}$, where φ is the deflection angle and α is related to the gravitational redshift. The Christoffel symbols are of first order in Ψ, so that the first-order equation of motion for the photon trajectory gives

$$\delta\dot{n}^\mu + \tilde{\Gamma}^\mu_{00} + 2\tilde{\Gamma}^\mu_{0j} n^j + \tilde{\Gamma}^\mu_{ij} n^i n^j = 0 \,.$$

For the metric $d\tilde{s}^2$ the only non-vanishing Christoffel symbols are

$$\tilde{\Gamma}^0_{0i} = \tilde{\Gamma}^0_{i0} = \tilde{\Gamma}^i_{00} = 2\partial_i \Psi \,.$$

For the deflection angle we therefore obtain

$$\dot{\varphi} = (\delta\dot{\mathbf{n}} \cdot \mathbf{e}) = -2\mathbf{e} \cdot \nabla\Psi = 2MG \frac{d}{(d^2 + s^2)^{3/2}} \,.$$

Integrating this from $s = -\infty$ to $s = \infty$ yields

$$\varphi = \frac{4MG}{d} \,. \tag{A10.18}$$

References

Abbott, L. F. and Schaefer, R. K. (1986). A general analysis of the cosmic microwave anisotropy, *Astrophys. J.* **308**, 546.

Abraham, R. and Marsden, J. (1982). *Foundations of Mechanics* (New York, Addison and Wesley), Ch. 3.

Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions*, 9th edn (New York, Dover Publications).

Adams, J. A., Ross, G. G. and Sarkar, S. (1997). Multiple inflation, *Nucl. Phys.* **B503**, 405.

Adams, J. F. (1981). Spin (8), triality, F_4 and all that. In *Superspace and Supergravity*, ed. S. W. Hawking and M. Röcek (Cambridge, Cambridge University Press).

Albrecht, A., Coulson, D. Ferreira, P. and Magueijo, J. (1996). Causality, randomness, and the microwave background, *Phys. Rev. Lett.* **76**, 1413.

Arfken, G. B. and Weber, H. J. (2001) *Mathematical Methods for Physicists*, 5th edn (Harcourt, Academic Press).

Arnol'd, V. I. (1978). *Mathematical Methods of Classical Mechanics* (Berlin, Springer-Verlag).

Astier, P. *et al.* (2006). The supernova legacy survey: measurement of Ω_M, Ω_Λ and w from the first year data set, *Astron. Astrophys.* **447**, 31.

Bahcall, N. and Cen, R. (1992). Galaxy clusters and cold dark matter – A low-density unbiased universe?, *Astrophys. J.* **398**, L81.

Bardeen, J. (1980). Gauge-invariant cosmological perturbations, *Phys. Rev.* **D22**, 1882.

Bashinsky, S. (2006). Gravity of cosmological perturbations in the CMB, *Phys. Rev.* **D 74**, 043007.

Bennett, C. L. *et al.* (2003). First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results, *Astrophys. J. Suppl.* **148**, 1.

Bernstein, J., Brown, L. and Feinberg, G. (1989). Cosmological helium production simplified, *Rev. Mod. Phys.* **61**, 25.

Bevis, N., Kunz, M. and Hindmarsh, M. (2004). WMAP constraints on inflationary models with global defects, *Phys. Rev.* **D70**, 043508.

Bevis, N., Hindmarsh, M., Kunz, M. and Urrestilla, J. (2007). CMB polarization power spectra contributions from a network of cosmic strings, *Phys. Rev.* **D76**, 043005.

Bonvin, C., Durrer, R. and Kunz, M. (2006). The dipole of the luminosity distance: a direct measure of $H(z)$, *Phys. Rev. Lett.* **96**, 191302.

Bruni M., Dunsby, P. and Ellis, G. (1992). Cosmological perturbations and the physical meaning of gauge-invariant variables, *Astrophys. J.* **395**, 34.

Bucher, M., Moodley, K. and Turok, N. (2000). General primordial cosmic perturbations *Phys. Rev.* **D62**, 083508. astro-ph/9904231.
Bucher, M., Moodley, K. and Turok, N. (2001). Constraining iso-curvature perturbations with CMB polarization, *Phys. Rev. Lett.* **87**, 191301.
Burles, S., Nollett, K. and Turner, M. (2001). Big-bang nucleosynthesis predictions for precision cosmology, *Astrophys. J.* **552**, L1.
Caldwell, R. R., Dave, R. and Steinhardt, P. J. (1998). Cosmological imprint of an energy component with general equation of state, *Phys. Rev. Lett.* **80**, 1582.
Caldwell, R. R., Kamionkowski, M. and Weinberg, N. N. (2003). Phantom energy and cosmic doomsday, *Phys. Rev. Lett.* **91**, 071301.
Caprini, C., Durrer, R. and Kahniashvili, T. (2004). The cosmic microwave background and helical magnetic fields: the tensor mode, *Phys. Rev.* **D69**, 063006.
Challinor, A. and Lewis, A. (2005). Lensed CMB power spectra from all-sky correlation functions, *Phys. Rev.* **D71**, 103010.
Chandrasekhar, S. (1939). *An Introduction to the Study of Stellar Structure* (Chicago, Chicago University Press).
Colafrancesco, S. (2007). Beyond the standard lore of the SZ effect, *New Astron. Rev.* **51**, 394.
Cole, S. *et al.* (2005). The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final dataset and cosmological implications, *Mon. Not. Roy. Astron. Soc.* **362**, 505.
Conklin, E. K. (1969). Velocity of the Earth with respect to the cosmic background radiation, *Nature* **222**, 971.
De Petris, M. *et al.* (2002). MITO measurements of the Sunyaev–Zel'dovich effect in the Coma cluster of galaxies, *Astrophys. J.* **574**, L119.
Dodelson, S. (2003). *Modern Cosmology* (New York, Academic Press).
Doran, M. (2005). CMBeasy: an object oriented code for the cosmic microwave background, *JCAP* **0510**, 011.
Durrer, R. (1990). Gauge-invariant cosmological perturbation theory with seeds, *Phys. Rev.* **D42**, 2533.
Durrer, R. (1994). Gauge invariant cosmological perturbation theory, *Fund. Cosmic Phys.* **15**, 209.
Durrer, R. and Straumann, N. (1988). Some applications of the $3 + 1$ formalism of general relativity, *Helv. Phys. Acta* **61**, 1027.
Durrer, R. *et al.* (1999). Seeds of large-scale anisotropy in string cosmology, *Phys. Rev.* **D59**, 043511.
Durrer, R., Kunz, M. and Melchiorri, A. (2002). Cosmic structure formation with topological defects, *Phys. Rep.* **364**, 1.
Durrer, R., Gabrielli, A., Joyce, M. and Sylos-Labini, F. (2003). Bias and the power spectrum beyond the turn-over, *Astrophys. J. Lett.* **585**, L1.
Ehlers, J. (1971). General relativity and kinetic theory. In *Proceedings of the Varenna School 1969, Course XLVII*, ed. R. K. Sachs (New York, Academic Press).
Eisenstein, D. J. *et al.* (2005). Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies, *Astrophys. J.* **633**, 560.
Ellis, G. F. R. and Bruni, M. (1989). Covariant and gauge-invariant approach to cosmological density fluctuations, *Phys. Rev.* **D40**, 1804.
ESA (2005). PLANCK, the scientific programme, ESA-SCI, 1.
Fields, B. D. and Sarkar, S. (2006). Big-bang nucleosynthesis, *J. Phys.* **G33**, 1.
Fixsen, D. J. *et al.* (1996). The cosmic microwave background spectrum from the full COBE FIRAS data set, *Astrophys. J.* **473**, 567.

Freedman, W. L. *et al.* (2001). Final results from the Hubble Space Telescope key project to measure the Hubble constant, *Astrophys. J.* **553**, 47.
Friedmann, A. (1922). Über die Krümming des Raumes, *Z. Phys.* **10**, 377.
Friedmann, A. (1924). Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes, *Z. Phys.* **21**, 326.
Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulations for Bayesian Inference* (London, Chapman and Hall).
Goldberg, J. N. *et al.* (1967). Spin-s Spherical Harmonics and $\eth$, *J. Math. Phys.* **8**, 2155.
Gorski, K. M. (1994). On determining the spectrum of primordial inhomogeneity from the COBE DMR sky maps: I. Method, *Astrophys. J.* **430**, L85.
Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Table of Integrals, Series and Products*, 6th edn (New York, Academic Press).
Gunn, J. E. and Peterson, B. A. (1965). On the density of neutral hydrogen in intergalactic space, *Astrophys. J.* **142**, 1633.
Hajian, A. (2007). Efficient cosmological parameter estimation with Hamiltonian Monte Carlo, *Phys. Rev.* **D75**, 083525.
Harrison E. R. (1970). Fluctuations at the threshold of classical cosmology, *Phys. Rev.* **D1**, 2726–2730.
Hawking, S. and Ellis, G. F. R. (1973). *The Large Scale Structure of the Universe* (Cambridge, Cambridge University Press).
Henry, P. S. (1971). Isotropy of the 3 K background, *Nature* **231**, 516.
Hinshaw, G. *et al.* (2007), Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: temperature analysis, *Astrophys. J. Suppl.* **170**, 288.
Hoffman, M. and Turner, M. (2001). Kinematic constraints to the key inflationary observables, *Phys. Rev.* **D64**, 023506.
Hogg, D. *et al.* (2005). Cosmic homogeneity demonstrated with luminous red galaxies, *Astrophys. J.* **624**, 54–58.
Hu, W. and Silk, J. (1993). Thermalization and spectral distortions of the cosmic microwave background radiation, *Phys. Rev.* **D48**, 485.
Hu, W. and Sugiayma, N. (1995). Anisotropies in the cosmic microwave background: an analytic approach, *Astrophys. J.* **444**, 489.
Hu, W. and White, M. (1996). CMB anisotropies in the weak coupling limit, *Astron. Astrophys.* **315**, 33.
Hu, W. and White, M. (1997a). Tensor anisotropies in an open universe, *Astrophys. J.* **486**, L1.
Hu, W. and White, M. (1997b). CMB anisotropies: total angular momentum method, *Phys. Rev.* **D56**, 596.
Hu, W., Scott, D., Sugiayma, N. and White, M. (1995). The effect of physical assumptions on the calculation of microwave background anisotropies, *Phys. Rev.* **D52**, 5498.
Hu, W., Seljak, U. A., White, M. and Zaldarriaga, M. (1998). Complete treatment of CMB anisotropies in a FRW universe, *Phys. Rev.* **D57**, 3290.
Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae, *Proc. Natl. Acad. Sci.*, USA **15**, 168.
Israel, W. and Stewart, J. (1980) *Einstein Commemorative Volume*, ed. A. Held (New York, Plenum Press).
Jackson, J. D. (1975). *Classical Electrodynamics*, 2nd edn (New York, Wiley & Sons).
Jauch, J. M. and Rorlich, F. (1976). *The Theory of Photons and Electrons* (New York, Springer-Verlag).
Jones, W. C. *et al.* (2006). A measurement of the angular power spectrum of the CMB temperature anisotropy from the 2003 flight of Boomerang, *Astrophys. J.* **647** 823.

Kamionkowski, M., Kosowsky, A. and Stebbins, A. (1997). Statistics of cosmic microwave background polarization, *Phys. Rev. Lett.* **78**, 2058.
Knox, L., Scoccimaro, R. and Dodelson, S. (1998). The impact of inhomogeneous reionization on cosmic microwave background anisotropy, *Phys. Rev. Lett.* **81**, 2004.
Kodama, H. and Sasaki, M. (1984). Cosmological perturbation theory, *Prog. Theor. Phys. Suppl.* **78**, 1.
Kogut, A. *et al.* (2006). ARCADE: absolute radiometer for cosmology, astrophysics, and diffuse emission, *New Astron. Rev.* **50**, 925.
Kolb, E. and Turner, M. (1990). *The Early Universe* (New York, Addison Wesley).
Kunz, M., Trotta, R. and Parkinson, D. (2006). Measuring the effective complexity of cosmological models, *Phys. Rev.* **D74**, 023503.
Kuo, C. L. *et al.* (2006). Improved measurements of the CMB power spectrum with ACBAR, arXiv:astro-ph/0611198.
Lachieze-Rey, M. and Luminet, J. P. (1995). Cosmic topology, *Phys. Rep.* **254**, 135–214.
LaRoque, S. *et al.* (2006). X-ray and Sunyaev–Zel'dovich effect measurements of the gas mass fraction in galaxy clusters, *Astrophys. J.* **652**, 917.
Lemaître, G. (1927). L'univers en expansion, *Ann. Soc. Bruxelles* **47A**, 49.
Lemaître, G. (1931). Expansion of the universe, a homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae, *Mon. Not. R. Ast. Soc.* **91** 483–490; Expansion of the universe, the expanding universe, *Mon. Not. R. Astron. Soc.* **91**, 490–501.
Lewis, A. and Bridle, S. (2002). Cosmological parameters from CMB and other data: a Monte Carlo approach, *Phys. Rev.* **D66**, 103511. (see also http://cosmologist.info/cosmomc/).
Lewis, A. and Challinor, A. (2006). Weak gravitational lensing of the CMB, *Phys. Rep.* **429**, 1.
Lewis, A., Challinor, A. and Lasenby, A. (2000). Efficient computation of CMB anisotropies in closed FRW models, *Astrophys. J.* **538**, 473. (see http://camb.info)
Liddle, A. and Lyth, D. (2000). *Cosmological Inflation and Large Scale Structure* (Cambridge, Cambridge University Press).
Lifshitz, E. (1946). About gravitational stability of expanding worlds, *JETP* **10**, 116.
Lifshitz, E. and Pitajewski, L. (1983). *Lehrbuch der Theoretischen Physik*, vol. X (Berlin, Akademie Verlag).
Linde, A. (1989). *Inflation and Quantum Cosmology* (New York, Academic Press).
MacKay, D. J. C. (2003). *Information Theory*, *Inference*, *and Learning Algorithms* (Cambridge, Cambridge University Press) (see also http:www.inference.phy.cam.ac.uk/mackay/itprnn/book.html).
Maggiore, M. (2005). *A Modern Introduction to Quantum Field Theory* (Oxford, Oxford University Press).
Maggiore, M. (2007). *Gravitational Waves* (Oxford, Oxford University Press).
Maldacena, J. (2003). Non-Gaussian features of primordial fluctuations in single field inflationary models, *J. High Energy Phys.* **0305**, 013.
Maoli, R. *et al.* (2001). Cosmic shear analysis in 50 uncorrelated VLT fields. Implications for Ω_0, σ_8, *Astron. Astrophys.* **368**, 766.
Martin, J. and Schwarz, D. (2003). WKB approximation for inflationary cosmological perturbations, *Phys. Rev.* **D67**, 083512.
McDonald, P. *et al.* (2005). The linear theory power spectrum from the Lyman-alpha forest in the Sloan Digital Sky Survey, *Astrophys. J.* **635**, 761.

Mészáros, P. (1974). The behaviour of point masses in an expanding cosmological substratum, *Astron. Astrophys.* **37**, 225.
Moodley, K. *et al*. (2004). Constraints on iso-curvature models from the WMAP first-year data, *Phys. Rev.* **D70**, 103520.
Mukhanov, V. F. (2005). *Physical Foundations of Cosmology* (Cambridge, Cambridge University Press).
Mukhanov, V. F. and Chibisov, G. (1982). The vacuum energy and large scale structure of the Universe, *JETP* **56**, 258.
Mukhanov, V. F., Feldman, H. A. and Brandenberger, R. H. (1992). Theory of cosmological perturbations, *Phys. Rep.* **215**, 203.
Newman, E. T. and Penrose, R. (1966). Note on the Bondi–Metzner–Sachs group, *J. Math. Phys.* **7**, 863.
Newton, I. (1958). Letters from Sir Isaac Newton to Dr. Bentley, Letter I, 203ff quoted by A. Koyré, *From the Classical World to the Infinite Universe* (New York, Harper and Row).
Olive, K. A., Steigman, G. and Walker, T. P. (2000). Primordial nucleosynthesis: theory and observations, *Phys. Rep.* **333**, 389.
Padmanabhan, N. *et al.* (2005). Correlating the CMB with luminous red galaxies: the integrated Sachs–Wolfe effect, *Phys. Rev.* **D72**, 043525.
Padmanabhan, T. (2000). *Theoretical Astrophysics* vol. I (Cambridge, Cambridge University Press).
Page, L. *et al.* (2007). Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: polarization analysis, *Astrophys. J. Suppl.* **170**, 335.
Particle Data Group (2004). Review of particle physics, *Phys. Lett.* **B592**, 191–227.
Particle Data Group (2006). Review of particle physics, ed. W.-M. Yao *et al.*, *J. Phys.* **G33**, 1.
Peacock, J. A. (1999). *Cosmological Physics* (Cambridge, Cambridge University Press).
Peebles, P. J. E. (1993). *Principles of Physical Cosmology* (Princeton, Princeton University Press).
Pierpaoli, E., Scott, D. and White, M. (2001). Power spectrum normalization from the local abundance of rich clusters of galaxies, *Mon. Not. R. Astron. Soc.* **325**, 77.
Pietrobon, D., Balbi, A. and Marinucci D. (2006). Integrated Sachs–Wolfe effect from the cross correlation of WMAP3 year and the NRAO VLA sky survey data: new results and constraints on dark energy, *Phys.Rev.* **D74**, 043524.
Press, W. and Schechter, P. (1974). Formation of galaxies and clusters of galaxies by self-similar gravitational condensation, *Astrophys. J.* **187**, 425.
Reed, M. and Simon, B. (1980). *Methods of Modern Mathematical Physics*, vol 1. *Functional Analysis* (New York, Academic Press).
Robertson, H. P. (1936). Kinematics and world structure I, II, III, *Astrophys. J.* **82**, 284–301; *Astrophys. J.* **83**, 187–201, 257–271.
Rocher, J. and Sakellariadou, M. (2005). Constraints on supersymmetric grand unified theories from cosmology, *J. Cosmol. Astroport. Phys.* **0503**, 004.
Rubino-Martin, J. A., Chluba, J. and Sunyaev, R. A. (2006). Lines in the cosmic microwave background spectrum from the epoch of cosmological hydrogen recombination, *Mon. Not. R. Astron. Soc.* **371**, 1939.
Rybicki, G. B. and Lightman, A. P. (1979). *Radiative Processes in Astrophysics* (New York, John Wiley and Sons).
Sachs, R. K. and Wolfe, A. M. (1967). Perturbations of a cosmological model and angular variations of the microwave background, *Astrophys. J.* **147**, 73.
Sakurai, J. J. (1993). *Modern Quantum Mechanics* (New York, Addison-Wesley).

Sandage, A. *et al.* (2006). The Hubble constant: a summary of the HST program for the luminosity calibration of type Ia supernovae by means of cepheids, *Astrophys. J.* **653**, 843.
Schneider, P. (2007). Weak gravitational lensing. In *33rd Saas Fee Lectures*, ed. P. Jetzer and P. North (Berlin, Springer-Verlag).
Schwarz, D., Terrero-Escalante, C. and Garcia, A. (2001). Higher order corrections to primordial spectra from cosmological inflation, *Phys. Lett.* **B517**, 243–249.
Seljak, U. (1996a). Rees–Sciama effect in a CDM universe, *Astrophys. J.* **460**, 549.
Seljak, U. (1996b). Measuring polarization in the cosmic microwave background, *Astrophys. J.* **482**, 6.
Seljak, U. and Zaldarriaga, M. (1996). A line of sight integration approach to cosmic microwave background anisotropies, *Astrophys. J.* **469**, 437.
Seljak, U., Slosar, A. and MacDonald, P. (2006). Cosmological parameters from combining the Lyman-alpha forest with CMB, galaxy clustering and SN constraints, *J. Cosmol. Astro. Part. Phys.* **0610**, 014.
Singal, J. *et al.* (2005). Design and performance of sliced-aperture corrugated feed horn antennas, *Rev. Sci. Instrum.* **76**, 124703. (For updates see http://arcade.gsfc.nasa.gov/)
Smoot, G. F. *et al.* (1992). Structure in the COBE differential microwave radiometer first-year maps, *Astrophys. J.* **396**, L1.
Songaila, A. and Cowie, L. L. (1996). Metal enrichment and ionization balance in the Lyman α forest at $z = 3$, *Astron. J.* **112**, 335.
Spergel, D. and Zaldarriaga, M. (1997). CMB polarization as a direct test of inflation, *Phys. Rev. Lett.* **79**, 2180.
Spergel, D. N. *et al.* (2003). First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: determinations of cosmological parameters, *Astrophys. J. Suppl.* **148**, 175.
Spergel, D. N. *et al.* (2007). Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology, *Astrophys. J. Suppl.* **170**, 377.
Stewart, J. M. (1971). *Non-equilibrium relativistic kinetic theory*, *Springer Lecture Notes* (Berlin, Springer-Verlag).
Stewart, J. M. and Walker, M. (1974). Perturbations of space-times in general relativity, *Proc. R. Soc. London* **A341**, 49.
Straumann, N. (1974). Minimal assumptions leading to a Robertson–Walker model of the Universe, *Hel. Phys. Acta* **47**, 379.
Straumann, N. (1984). *General Relativity and Relativistic Astrophysics* (Berlin, Springer-Verlag).
Straumann, N. (2004). *General Relativity With Applications to Astrophysics* (Berlin, Springer-Verlag).
Sylos Labini, F., Montuori, M. and Pietronero, L. (1998). Scale-invariance of galaxy clustering, *Phys. Rep.* **293**, 61–226.
Tegmark, M. *et al.* (2004). The 3D power spectrum of galaxies from the SDSS, *Astrophys. J.* **606**, 702.
Thorne, K. (1980). Multipole expansions of gravitational radiation, *Rev. Mod. Phys.* **52**, 299.
Trotta, R. (2006). The iso-curvature fraction after WMAP 3-year data, *Mon. Not. R. Astron. Soc. Lett.* **375**, L26.
Trotta, R., Riazuelo, A. and Durrer, R. (2001). Reproducing cosmic microwave background anisotropies with mixed iso-curvature perturbations, *Phys. Rev. Lett.* **87**, 231301.

Trotta, R., Riazuelo, A. and Durrer, R. (2003). The cosmological constant and general iso-curvature initial conditions, *Phys. Rev.* **D67**, 063520.
Vilenkin, N. Y. and Smorodinskii, Y. A. (1964). *Sov. Phys. JETP*, **19**, 1209.
Wald, R. M. (1984). *General Relativity* (Chicago, University of Chicago Press).
Walker, A. G. (1936). *Proc. Lond. Math. Soc*. **42**, 90.
Wolf, J. (1974). *Spaces of Constant Curvature* (Boston, Publish or Perish).
Wong, W. Y., Seager, S. and Scott, D. (2006). Spectral distortions to the cosmic microwave background from the recombination of hydrogen and helium, *Mon. Not. R. Astron. Soc*. **367**, 1666.
Wood-Vasey, W. M. *et al.* (2007). Observational constraints on the nature of dark energy: first cosmological results from the ESSENCE supernova survey, astro-ph/0701041.
Zaldarriaga, M. and Seljak, U. (1997). An all-sky analysis of polarization in the microwave background, *Phys. Rev.* **D55**, 1830–1840.
Zel'dovich, Y. B. (1972). A hypothesis, unifying the structure and the entropy of the Universe, *Mon. Not. R. Astron. Soc.* **160**, 1.

Index