Ludwig-Maximilians-Universität

Sigillum Universitatis Ludovici Maximiliani

Temperature and Polarization Studies of the Cosmic Microwave Background

Dissertation an der Fakultät für Physik
der Ludwig-Maximilians-Universität München
für den Grad des
Doctor rerum naturalium

vorgelegt von Mona Silja Frommert
aus Böblingen

München, den 29.10.2009

Sigillum Universitatis Ludovici Maximiliani

1. Gutachter: Prof. Dr. Simon D. M. White

2. Gutachter: Prof. Dr. Jochen Weller

Tag der mündlichen Prüfung: 11.02.2010

Dass ich erkenne, was die Welt /
Im Innersten zusammenhält

Johann Wolfgang von Goethe, Faust

Contents

Contents 1

Zusammenfassung 9

Abstract 11

Introduction 13

1 Cosmology 17

1.1 Basics of General Relativity 17
1.2 Basics of cosmology 18
1.3 The cosmological concordance model 21
1.3.1 Constituents 21
1.3.2 Dark energy 22
1.3.3 Inflation 24
1.4 The cosmic microwave background 25
1.4.1 Origin of the CMB 25
1.4.2 Temperature anisotropies 26
1.4.3 Secondary anisotropies 28
1.4.4 CMB polarization 31
1.5 Other observational probes of the Universe 35
1.5.1 The luminosity distance-redshift relation 35
1.5.2 Angular distances 36
1.5.3 Weak lensing 37
1.5.4 Primordial nucleosynthesis 37

2 Statistical inference **39**

2.1 Notation 39

2.2 Bayesian inference 40

2.3 Information field theory 40

2.3.1 Free theory and the Wiener filter 42

2.3.2 Interacting information fields 44

2.4 Parameter sampling 45

2.4.1 Markov Chains and the Metropolis Algorithm 46

3 Optimal methods for detecting the integrated Sachs-Wolfe effect **49**

3.1 Introduction 49

3.2 Stochastic processes 51

3.2.1 Realisation of the matter distribution 51

3.2.2 CMB detector noise 52

3.2.3 Shot noise 53

3.3 Standard cross-correlation method 55

3.3.1 Description 55

3.3.2 Analysis of error-contributions 58

3.4 Optimal temperature method 60

3.4.1 Derivation of the posterior distribution 61

3.4.2 Estimation of the ISW amplitude 63

3.4.3 Comparison of signal-to-noise ratios and biasing 64

3.5 Optimal polarization method 67

3.5.1 Uncorrelated temperature maps 68

3.5.2 Reduction of variance 70

3.5.3 Application to the ISW effect 73

3.6 Improvement for currently available data 76

3.7 Likelihood for cosmological parameters 80

3.8 Conclusions 81

4 The axis of evil - a polarization perspective **85**

4.1 Introduction 85
4.2 Splitting of the temperature map 87
4.3 Splitting of the polarization map 94
4.4 The axis of evil 95
4.5 Conclusions 102

5 Efficient cosmological parameter sampling using sparse grids 105

5.1 Introduction 105
5.2 Basics of sparse grids 107
5.2.1 General idea of interpolating a function on a full grid 107
5.2.2 Hierarchical basis functions in one dimension 108
5.2.3 Higher-dimensional interpolation on a full grid 110
5.2.4 Sparse grids 110
5.2.5 Extension to functions that are non-zero on the boundary 111
5.3 Interpolation of the WMAP likelihood surface 114
5.3.1 Choice of basis functions 114
5.3.2 Normal parameters 115
5.3.3 Generation of test set and choice of interpolation range 117
5.3.4 Results 118
5.3.5 Improvements with adaptive sparse grids 124
5.4 Conclusions 127

6 Conclusions and outlook 129

A 133

A.1 Derivation of the Wiener filter 133
A.1.1 Lemma 1 134
A.1.2 Lemma 2 135
A.1.3 Lemma 3 135

B 137

B.1 Proof of the equivalence of the number of bins 137

B.2 Proof of the factorization of the likelihood 138

C **141**

C.1 Proof of vanishing correlation between $T^{\mathrm{rec}}_{\mathrm{corr}}$ and $T^{\mathrm{rec}}_{\mathrm{uncorr}}$ 141

C.2 Proof of vanishing correlation between P_{corr} and $P^{\mathrm{rec}}_{\mathrm{uncorr}}$ 142

C.3 Probability for chance alignment in an isotropic universe 143

D **145**

D.1 Inversion of the parameter transformation 145

E Acknowledgements **147**

List of Figures

List of Figures **5**

1.1 Temperature power spectrum of a ΛCDM universe 27
1.2 CMB and ISW power spectrum of a ΛCDM universe 30
1.3 Cartoon explaining the creation of linear polarization by Thomson scattering . . 32
1.4 TE cross power spectrum of a ΛCDM universe 34
1.5 TT and EE power spectrum of a ΛCDM universe 34

3.1 Comparison of signal-to-noise ratios vs l 60
3.2 Probability distribution of the signal-to-noise ratio 65
3.3 Comparison of signal-to-noise ratios vs $z_{\rm max}$ 66
3.4 Reduction of the variance when including polarization data 72
3.5 Cross-correlation between $T_{\rm isw}$ and polarization 73
3.6 Realisation of the maps 74
3.7 Reduction of the variance when including temperature data 75
3.8 Comparison of signal-to-noise ratios vs l 76
3.9 Comparison of signal-to-noise ratios vs $z_{\rm max}$ 77
3.10 Variance reduction with current and future data 78
3.11 Improvement of detection significance for current and future data 79

4.1 Observed polarization maps 88
4.2 Splitting of Stokes Q map 90
4.3 Splitting of Stokes U map 91
4.4 Splitting of temperature map 93
4.5 The *Cold Spot* 94

4.6 Wiener realisation of T_{corr} 97
4.7 Preferred axes for P_{corr} 98
4.8 Preferred axes for $P^{\text{rec}}_{\text{uncorr}}$ 99
4.9 Preferred axes for $T^{\text{rec}}_{\text{corr}}$ 100
4.10 Preferred axes for $T^{\text{rec}}_{\text{uncorr}}$ 101

5.1 One-dimensional piecewise linear interpolation 108
5.2 One-dimensional basis functions 109
5.3 Two-dimensional basis functions 111
5.4 Two-dimensional sub-grids of a sparse grid 112
5.5 Sparse grids in two and three dimensions 112
5.6 Modified one-dimensional basis functions 113
5.7 Scatterplots, 6 dimensions, normal parameters 118
5.8 Scatterplots, 7 dimensions, normal parameters 118
5.9 Histograms, 6 dimensions, normal parameters 120
5.10 Histograms, 7 dimensions, normal parameters 120
5.11 Scatterplots, 6 dimensions, standard parameters 121
5.12 Scatterplots, 7 dimensions, standard parameters 122
5.13 Histograms, 6 dimensions, standard parameters 122
5.14 Histograms, 7 dimensions, standard parameters 123
5.15 Scatterplot, 7 dimensions, standard parameters 126

D.1 The two branches of h_3 versus w_b for $h_2 = 0.45$. 145

List of Tables

List of Tables **7**

1.1 Table of cosmological parameters 23

4.1 Preferred axes and their uncertainties 86

5.1 Comparison of errors for adaptive refinement 127

Zusammenfassung

Der kosmische Mikrowellenhintergrund (cosmic microwave background, CMB) enthält eine Fülle von Informationen über die Eigenschaften unseres Universums. In dieser Dissertation entwickeln wir neue Techniken, mit welchen wir anhand des CMB fundamentale Fragen der Kosmologie erörtern.

Dunkle Energie – falls sie existiert – hinterlässt charakteristische Spuren im CMB, den so genannten *integrierten Sachs-Wolfe* (ISW) Effekt. Man kann diesen schwachen Effekt über seine Kreuzkorrelation mit der großräumigen Materieverteilung (large-scale structure, LSS) detektieren. Wir entwickeln eine optimale Methode zur Detektion des ISW-Effekts, basierend auf Temperatur- und Polarisationsdaten des CMB, die sich von der bisher verwendeten Methode in zwei fundamentalen Punkten unterscheidet: Wir halten die LSS und einen Teil der primordialen Temperaturfluktuationen fest, anstatt über verschiedene Realisierungen zu mitteln, wie es für die Standardmethode der Fall ist. Für ein ideales Szenario ist die ISW-Detektion mit unserer Methode um 23% signifikanter als mit der Standardmethode. Für Polarisationsdaten der *Planck Surveyor*-Mission wird diese Erhöhung mindestens 10% betragen, wobei der limitierende Faktor die Kontamination durch Vordergrundemission unserer Galaxie ist.

Der beobachtete CMB ist fast perfekt isotrop, was als Evidenz für die Isotropie unseres Universums gilt. Allerdings wurden in der Temperaturkarte des *Wilkinson Microwave Anisotropy Probe* (WMAP) Satelliten Anomalien gefunden, die die statistische Isotropie der Temperaturfluktuationen in Frage stellen. Wir versuchen zu verstehen, ob diese Anomalien zufällige Fluktuationen sind oder ihren Ursprung in einer bevorzugten Richtung haben, die der Geometrie des primordialen Universums intrinsisch ist. Dafür berechnen wir den Teil der WMAP-Polarisationskarte, welcher nicht mit der Temperaturkarte korreliert ist, und verwenden diese Karte als statistisch unabhängigen Test für die so genannte *axis of evil*. Letztere stellt eine ungewöhnliche Übereinstimmung der bevorzugten Richtungen des Quadrupols und des Oktopols der Temperaturkarte dar. In der unkorrelierten Polarisationskarte stimmt die Achse des Quadrupols mit der *axis of evil* überein, die Achse des Oktopols jedoch nicht. Allerdings haben wir auf Grund des hohen Rauschanteils in der WMAP-Polarisationskarte eine Unsicherheit von ca. 45° in den Achsen, für welche die Wahrscheinlichkeit, dass wir unser Ergebnis in einem isotropen Universum durch Zufall erhalten, ca. 50% beträgt. Wir erhalten daher keinen Hinweis für oder gegen eine bevorzugte Richtung im primordialen Universum. Für *Planck* erwarten wir eine Verbesserung der Unsicherheit in den Achsen auf $10^\circ - 20^\circ$, je nachdem, wie gut Vordergrundemissionen von der Polarisationskarte entfernt werden können. Unsere Technik angewandt auf *Planck*-Daten wird uns daher als mächtiges Instrument dienen, um den Ursprung der CMB-Anomalien zu verstehen.

Anstatt wie oben beschrieben bestimmte Merkmale des CMB zu analysieren, können wir den CMB auch dazu verwenden, mehrere kosmologische Parameter gleichzeitig einzugrenzen. Die Werte der kosmologischen Parameter, die mit WMAP bestimmt wurden, haben die Ära der Präzisionskosmologie eingeleitet und können als der größte Erfolg der Mission betrachtet werden. In solchen Studien zur Parameterbestimmung ist der geschwindigkeitsbestimmende Schritt normalerweise die Auswertung der Likelihood-Funktion. Um dieses Problem zu umgehen, haben wir daher eine auf dünnen Gittern basierende Interpolation der WMAP-Likelihood-Funktion entwickelt und implementiert, die um Größenordnungen schneller auszuwerten ist als die ursprüngliche Likelihood-Funktion. Unsere Methode ist eine konkurrenzfähige Alternative zu anderen Ansätzen für die Steigerung der Effizienz von Parameterbestimmung.

Abstract

The cosmic microwave background (CMB) provides us with a wealth of information about the properties of our Universe. In this PhD work, we develop and apply new techniques for studying fundamental problems of cosmology using the CMB.

Dark energy, if it exists, leaves a characteristic imprint in the CMB temperature fluctuations, the so-called *integrated Sachs-Wolfe* (ISW) effect. This small effect can be detected via its cross-correlation with the large-scale structure (LSS). We derive an optimal method for ISW detection using temperature and polarization data of the CMB which differs from that usually used in two fundamental ways: we keep the LSS distribution and a part of the primordial temperature fluctuations fixed, rather than averaging over different realisations as done in the standard method. For an ideal scenario, we obtain an overall enhancement of the detection significance of 23 per cent. For polarization data from the *Planck Surveyor* mission, this enhancement will be at least 10 per cent, where the limiting factor will be the contamination by Galactic foregrounds.

The CMB is observed to be almost perfectly isotropic, which is considered strong evidence for the isotropy of the Universe. However, some anomalies have been found in the temperature map of the *Wilkinson Microwave Anisotropy Probe* (WMAP), which seem to question the statistical isotropy of the temperature fluctuations. In order to understand whether these are due to chance fluctuations or to a preferred direction intrinsic to the geometry of the primordial Universe, we compute the part of the WMAP polarization map which is uncorrelated with the temperature map, and use it as a statistically independent probe of the so-called *axis of evil*. The latter is an unusual alignment between the preferred directions of the quadrupole and the octopole in the temperature map. We find that the axis of the quadrupole of the uncorrelated polarization map aligns with the axis of evil, whereas the axis of the octopole does not. However, due to the high noise-level in the WMAP polarization map, we have an uncertainty of about 45° in our axes. With this uncertainty, the probability of at least one axis aligning by chance in an isotropic Universe is around 50 per cent. We therefore do not obtain evidence for or against a preferred direction intrinsic to the primordial Universe. For *Planck*, we expect the uncertainty in the axes to go down to $10^\circ - 20^\circ$, again depending on how well the foregrounds can be removed from the map. Our technique applied to *Planck* data will thus serve as a powerful means to understand the origin of the CMB anomalies.

Instead of studying particular features in the CMB maps as described above, we can also use the CMB to constrain several cosmological parameters simultaneously by sampling the parameter space. The parameter constraints obtained by WMAP marked the beginning of precision cosmology and were the biggest success of the mission. In such parameter sampling studies, the main bottleneck is usually the evaluation of the likelihood. We have thus implemented a sparse-grids based interpolation of the WMAP likelihood surface as a shortcut for the likelihood evaluation. This is orders of magnitude faster to compute than the original likelihood. Our method is a competitive alternative to other approaches for speeding up parameter sampling.

Introduction

For hundreds of years, cosmology had been more of a metaphysical field than a scientific discipline. However, in the course of the last century, great progress has been made in understanding the origin and nature of our Universe. The basics for all modern-day cosmology have been laid by Einstein in the early twentieth century, when he formulated his General Theory of Relativity (Einstein 1916). Then, in 1929, Edwin Hubble discovered the expansion of the Universe (Hubble 1929), which should revolutionise the field of cosmology, given that most cosmologists at that time, including Einstein, had favoured a static Universe. Hubble realised that many of the faint nebulae in the night sky are galaxies outside our Milky Way, and that most of these galaxies are moving away from us with a recession velocity proportional to their distance from the Earth. This was interpreted as an overall expansion of space. After Hubble's discovery, cosmology was dominated by two different schools, one of which supported the idea that the Universe has originated in a Big Bang (Lemaître 1931), whereas the other school favoured the so-called Steady State model (Bondi & Gold 1948), according to which the Universe did not have a beginning, but has always existed and expanded as it does now. The end of the Steady State model was marked by the discovery of the cosmic microwave background (CMB) by Arno Penzias and Robert Wilson in 1965 (Penzias & Wilson 1965; Dicke et al. 1965). The CMB had been predicted to exist in the Big Bang model by Gamow, Alpher and Herman in 1948 (Gamow 1948; Alpher & Herman 1948) but lacked a natural explanation in the Steady State model. In the late nineties, Riess et al. (1998) and Perlmutter et al. (1999) measured the redshift-distance relation of supernovae of type Ia and found that the expansion of the Universe is accelerating rather than decelerating as previously thought. This led cosmologists to postulate some unknown medium named *dark energy*, which pervades the Universe and drives the accelerated expansion of space.

The first full-sky map of the small temperature fluctuations of the CMB were obtained by the COBE satellite in the early nineties (Wright et al. 1992). These fluctuations are created by physical processes in the early (and late) Universe, which depend on the parameters of the cosmological model and can thus be used to infer information about the latter. With the precise analysis of the CMB by the *Wilkinson Microwave Anisotropy Probe* (WMAP) satellite and various balloon-based experiments, together with observations of other cosmological probes, we are now able to constrain the cosmological parameters with an accuracy on the level of a few per cent (Komatsu et al. 2009). This has lead to a widely accepted cosmological model, the so-called concordance model or ΛCDM model. We expect to obtain even stronger constraints on the cosmological parameters from the *Planck Surveyor* mission, which has been launched in May 2009 and will measure the CMB with unprecedented accuracy within the next year.

However, even though we seem to be converging towards a consistent picture of our Universe, there remain many questions yet to be answered. We lack a theoretical understanding of three essential ingredients of the ΛCDM model: dark matter, dark energy, and inflation. Therefore, great effort is put into developing theoretical models and finding observational evidence for them. Even the most fundamental assumption of cosmology, the homogeneity and isotropy of the Universe, is the subject of some doubts triggered by observations of the CMB (de Oliveira-Costa et al. 2004; Eriksen et al. 2007). We can hope to find answers to these unresolved questions in the information contained in cosmological signals such as the CMB. However, the signals encoded in data from cosmological observations are subject to various uncertainties, ranging from detector noise to the complex and imperfect signal transmission by the cosmological processes themselves. A crucial step in understanding our Universe is thus the harvesting of the information content of cosmological data.

In this PhD work, we study specific aspects of the CMB related to both the question of dark energy and the isotropy of the Universe. Both of these studies rely on similar techniques of combining CMB temperature and polarization data using their cross-correlation. This permits us to infer more information about the respective problem than previous studies, provided data of high quality are available. The signature of dark energy in the CMB causes a small coupling between CMB and large-scale structure data. We derive a technique to consistently treat this coupling in cosmological parameter estimation studies. In order to make such parameter estimations more efficient in general, we finally develop and implement a novel likelihood reconstruction method based on the technique of sparse grids.

The integrated Sachs-Wolfe effect

The first part of this work is devoted to developing new techniques to detect the *integrated Sachs-Wolfe* (ISW) effect in the CMB temperature fluctuations. The ISW effect is a probe of the existence of dark energy (or modified gravity), and can in principle provide us with information about the dark-energy related cosmological parameters. It can be detected via its cross-correlation with the LSS (Ho et al. 2008; Giannantonio et al. 2008), but the measurement of this cross-correlation signal is made difficult by confusion with chance correlations of the primordial CMB fluctuations with the LSS.

The optimal methods for ISW detection developed in this work are designed to infer the maximum amount of information about the ISW effect that the data have to offer, thus reaching a higher detection significance than previously existing methods. As a first step, we keep the realisation of the LSS distribution fixed when trying to detect the ISW, rather than averaging over it as it is done in existing methods. This allows for an enhancement of the detection significance for surveys going to relatively high redshifts. As a subsequent step, we extend this method to include CMB polarization data, which are used to fix a part of the primordial temperature fluctuations in the analysis. Roughly speaking, we remove from the temperature map those structures that are also encoded in the polarization data, and search for the ISW effect in the remaining temperature map, which is now uncorrelated with polarization. Given that the ISW effect is not significantly imprinted in the polarization, we thereby remove variance from the data, which is only a nuisance to the ISW detection. This also considerably enhances the detection significance at small redshifts.

For an ideal scenario, we expect an enhancement of the detection significance by 16 per cent for low redshift surveys such as the SDSS galaxy sample, and by about 23 per cent for surveys ranging to higher redshifts of about 2. For currently available CMB and LSS data, we estimate the improvement of the detection significance of our method as compared to the standard one to be at most 5 per cent. The main reason for that is the low signal-to-noise ratio of the polarization data from WMAP. However, soon the *Planck Surveyor* mission will provide us with polarization measurements of much higher quality than the WMAP data. A very crude estimate yields an improvement of the detection significance for *Planck* of at least 10 per cent, where the limiting factor is how well foregrounds can be removed from the polarization data.

In order to include the information contained in the ISW effect in cosmological parameter estimation, we derive the joint likelihood for cosmological parameter estimation for CMB and LSS data, which consistently includes the coupling between the two data-sets introduced by the ISW effect. Using this likelihood instead of assuming that the likelihoods for CMB and LSS data are independent will result in small changes in the constraints on the dark-energy related parameters.

Anomalies in the CMB

In the second part of this work, we study the so-called *axis of evil*, a strong alignment between the preferred directions of the quadrupole and the octopole in the CMB temperature map. This axis has been found in recent years as one of several direction-dependent phenomena in the CMB that seem to question the isotropy of the Universe (de Oliveira-Costa et al. 2004; Abramo et al. 2006; Land & Magueijo 2007). It is strongly under debate whether these anomalies are simply due to chance fluctuations in the CMB temperature map, if they can be blamed on local structures or on systematics in the measurement, or whether they are actually due to a preferred direction intrinsic to the geometry of the primordial Universe. We can shed more light on this question by probing the anomalies with a statistically independent data set.

Since both the temperature and polarization fluctuations have their physical origin in the primordial gravitational potential, we expect the polarization data to show similar peculiarities as the temperature, provided they are due to some preferred direction intrinsic to the geometry of the primordial Universe. However, since the polarization is not statistically independent of the temperature, anomalies due to chance fluctuations in the temperature can also manifest themselves in the polarization map. We suggest to split the polarization map into a part which is correlated with the temperature map, and an uncorrelated part, analogously to what we have done with the temperature map in the optimal method for ISW detection. The uncorrelated part of the polarization serves as a statistically independent probe of the anomalies described above.

We compute this map for the WMAP data, and use it to search for the *axis of evil*. We find that, within our measurement precision, the axis of the quadrupole of this map aligns with the axis of evil, whereas the axis of the octopole does not. However, due to the high contamination of the WMAP polarization data with detector noise and Galactic foregrounds, the uncertainty in our axes is of the order of 45°. With such an uncertainty, the probability of at least one axis being aligned with the axis of evil within its error bar just by chance amounts to about 50 per cent in an isotropic universe. For *Planck* data, a crude estimate yields an uncertainty in the axes of only 10° - 20°. With *Planck*, we will thus have a powerful test to probe the axis of evil and other CMB anomalies in polarization.

Parameter sampling

As we have indicated before, there is a variety of physical processes that have imprinted their signatures in the CMB and other cosmological data-sets such as the LSS and the redshift-distance relation of supernovae of type Ia. The ISW effect we study in the first part of this work is one example of such a process. We can therefore use cosmological data in order to simultaneously constrain the cosmological parameters that have determined the appearance of one or more of these imprints. To this end, we have to define a likelihood, which measures how well the data fit to the theoretical prediction for given values of the parameters, and multiply it by an adequate prior distribution in order to obtain the probability distribution of the parameters conditional on the data, the so-called *posterior distribution*.

Analysing the posterior distribution is not always easy, since changes in the observables typically do not reflect variations of one particular parameter, but rather depend on a combination of certain parameters of interest. Therefore, a set of parameters has to be determined simultaneously and preferentially with several data-sets in combination. Usually, we are dealing with cosmological models consisting of at least 6 parameters, for which we infer the mean and variance by sampling the parameter space using Markov Chain Monte Carlo (MCMC) simulations. For yielding reliable parameter estimates, MCMCs have to evaluate the posterior distribution (and thus the likelihood) for about 50,000 - 500,000 points in parameter space for a 6-9 dimensional cosmological model. However, the evaluation of the likelihood of the cosmological parameters is very expensive and is thus the main bottleneck in the parameter sampling process. Therefore, fast methods to evaluate the likelihood are becoming of increasing importance, especially in the light of the *Planck Surveyor* mission, which will soon provide us with a huge amount of data that we have to handle.

In the third part of this work, we therefore develop and implement a method to speed up the evaluation of the likelihood by interpolating the likelihood surface using a technique based on sparse grids. We show that projecting our interpolation with MCMCs reproduces the one-dimensional posterior distributions for the cosmological parameters almost perfectly, running in only a fraction of the time it takes to run them on the full likelihood. Using our interpolation, the main bottleneck in parameter sampling studies is now the MCMC algorithm itself rather than the likelihood evaluation. In speed and accuracy, our interpolation method is comparable to approaches of fitting the likelihood surface with polynomials or neural networks, while overcoming some of the drawbacks of the latter. These are, for example, the danger of creating unphysical wiggles if the polynomial degree is chosen too high with respect to the number of available training points, or the comparably long training time required for neural networks. Thus, our approach is a competitive alternative to existing approaches to accelerate parameter estimation (Fendt & Wandelt 2007; Auld et al. 2008).

This thesis is organised as follows. We start by introducing the basic concepts of modern cosmology in chapter 1 and the necessary basics of statistical inference in chapter 2. Chapter 3 is devoted to our work on the ISW effect, whereas chapter 4 describes the project on the axis of evil. In chapter 5, we present the work on the acceleration of cosmological parameter estimation using sparse grids. Concluding remarks and a short outlook are given in chapter 6.

Chapter 1

Cosmology

This chapter is devoted to introducing the fundamental concepts of modern cosmology as a necessary background for this PhD work. We first explain the basics of Einstein's General Relativity in section 1.1, then derive the basic equations for modern cosmology from the Einstein equations in section 1.2. In section 1.3, we briefly describe the cosmological concordance model. We explain in detail the cosmic microwave background radiation in section 1.4, and briefly outline other observational probes of the Universe in section 1.5.

1.1 Basics of General Relativity

In this section, we give a brief introduction to the General Theory of Relativity (GR). For a good and detailed discussion of General Relativity, the reader is referred to Misner et al. (1973).

Einstein's General Theory of Relativity is currently the best description of the laws of gravity that we have. It is based on the idea that spacetime is not just a static background, a framework in which the laws of physics can be described, but spacetime is actually influenced by the matter that lives in it. Matter curves spacetime, that is, it changes its geometry. The geometry of spacetime, in turn, determines how the matter moves through it. We call the quantity that encodes the geometry of spacetime the metric tensor, $g_{\mu\nu}$. It defines the infinitesimal spacetime-interval ds between two neighbouring points in spacetime with coordinate distances dx^μ via the relation $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$. Throughout this work, Greek indices are used as spacetime-indices running from 0 to 3, over one time-dimension and 3 spatial dimensions, whereas Latin indices are used as spatial indices that run from 1 to 3. We also use Einstein's sum convention, according to which indices that occur as upper and as lower index in a term are summed over.

From the first and second derivatives of the metric we can construct the Riemann tensor,

$$R^\mu_{\nu\lambda\rho} \equiv \frac{\partial\Gamma^\mu_{\nu\rho}}{\partial x^\lambda} - \frac{\partial\Gamma^\mu_{\nu\lambda}}{\partial x^\rho} + \Gamma^\mu_{\lambda\alpha}\Gamma^\alpha_{\nu\rho} - \Gamma^\mu_{\rho\alpha}\Gamma^\alpha_{\nu\lambda}\,, \tag{1.1}$$

which describes the curvature of spacetime. Here, we have defined the Christoffel symbols

$$\Gamma^\mu_{\nu\lambda} \equiv \frac{1}{2}g^{\mu\alpha}\left[\frac{\partial g_{\alpha\nu}}{\partial x^\lambda} + \frac{\partial g_{\alpha\lambda}}{\partial x^\nu} - \frac{\partial g_{\nu\lambda}}{\partial x^\alpha}\right]. \tag{1.2}$$

Note that upper/lower indices are called contravariant/covariant indices, and can be converted into one another by applying the metric tensor: $v_\mu = g_{\mu\nu}v^\nu$. Contracting the Riemann tensor over the first and third index yields the Ricci tensor:

$$R_{\mu\nu} \equiv R^\lambda_{\mu\lambda\nu}\,, \tag{1.3}$$

the contraction of which with the metric represents a scalar curvature, which is called the Ricci scalar:

$$R \equiv g^{\mu\nu}R_{\mu\nu}\,. \tag{1.4}$$

Given the above definitions, we can now define the Einstein tensor,

$$G^\mu_\nu \equiv R^\mu_\nu - \frac{1}{2}\delta^\mu_\nu R\,, \tag{1.5}$$

where δ^μ_ν is the Kronecker symbol. The Einstein tensor plays a central role in General Relativity.

We have now defined the necessary quantities describing the geometry of spacetime. The matter distribution in spacetime is described by the energy momentum tensor T^μ_ν, which we require to be divergence-free with respect to the covariant derivative, in order to impose generalised energy and momentum conservation:

$$T^\mu_{\nu;\mu} = 0\,, \tag{1.6}$$

where the covariant derivative is defined as

$$T^\mu_{\nu;\lambda} \equiv \frac{\partial T^\mu_\nu}{\partial x^\lambda} + \Gamma^\mu_{\alpha\lambda}T^\alpha_\nu - \Gamma^\alpha_{\nu\lambda}T^\mu_\alpha\,. \tag{1.7}$$

The fundamental equations relating the Einstein Tensor G^μ_ν and the energy-momentum tensor T^μ_ν are the Einstein equations:

$$G^\mu_\nu = 8\pi G\, T^\mu_\nu\,, \tag{1.8}$$

where we have set the speed of light to $c = 1$, a convention which we will use throughout this work.

1.2 Basics of cosmology

The evolution of the Universe as a whole is described by the General Theory of Relativity, which we have introduced in the last section. Let us now derive the basic equations of cosmology from the Einstein equations. For a more thorough introduction to modern cosmology, see, for example, Coles & Lucchin (2002), Dodelson (2003), or Peacock (1999).

The basic assumption in cosmology is the so-called cosmological principle, which states that the Universe is homogeneous and isotropic on large scales. This considerably simplifies the structure of the metric. The most general metric in a homogeneous and isotropic universe is the Robertson-Walker metric, for which an infinitesimal spacetime interval ds is given by

$$ds^2 = -dt^2 + a(t)^2\left(\frac{dr^2}{1-Kr^2} + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2\right), \tag{1.9}$$

where, r, θ, and ϕ are spherical coordinates on a constant-time hypersurface. The variable t denotes the universal time, $a(t)$ is the overall expansion parameter of space, the so-called *scale factor*, and K is called curvature parameter. The curvature parameter can have the values 0, 1 and -1, for which the metric describes a spatially flat, closed or open universe, respectively. The spatial coordinates r, θ and ϕ are called comoving coordinates, because the time-dependent part is factored out. We can substitute the time coordinate in the Robertson-Walker metric by the *conformal time* η using the definition

$$dt = a(\eta)\, d\eta\,, \tag{1.10}$$

so that the metric becomes

$$ds^2 = a(\eta)^2\Big(-d\eta^2 + \frac{dr^2}{1-Kr^2} + r^2 d\theta^2 + r^2\sin^2\theta d\phi^2\Big). \tag{1.11}$$

The various matter components α of the Universe can be described by the energy momentum tensor of a perfect fluid with the coordinates of the Robertson-Walker metric being fixed to the fluid elements. For baryonic matter, the rest frame of those fluid elements corresponds to the one of the galaxies, if one averages out the proper motion of the individual galaxies. The energy momentum tensor of a perfect fluid is

$$T^{\ \mu}_{(\alpha)\nu} = (\rho_\alpha + p_\alpha)u^\mu u_\nu + p_\alpha \delta^\mu_\nu\,, \tag{1.12}$$

where ρ_α and p_α denote the energy density and pressure of the component α, and $u^\mu \equiv \frac{dx^\mu}{d\lambda}$ is the 4-velocity of the fluid, which is defined as the derivative of the spacetime-coordinate x^μ with respect to the proper time λ of the fluid (in the rest frame of the fluid, $ds^2 = -d\lambda^2$). Since the fluid is at rest in our coordinate system, the spatial components u^i of the 4-velocity vanish. For the zero-components of the 4-velocity, we obtain $u^0 u_0 = u^\mu u_\mu = \frac{dx^\mu}{d\lambda}\frac{dx_\mu}{d\lambda} = \frac{ds^2}{d\lambda^2} = -1$, so that the energy momentum tensor of the fluid becomes

$$T^{\ \mu}_{(\alpha)\nu} = \mathrm{diag}(-\rho_\alpha,\ p_\alpha,\ p_\alpha,\ p_\alpha)\,. \tag{1.13}$$

The total energy momentum tensor is then just the sum over the components,

$$T^\mu_\nu = \sum_\alpha T^{\ \mu}_{(\alpha)\nu} = \mathrm{diag}(-\rho,\ p,\ p,\ p)\,, \tag{1.14}$$

where $\rho \equiv \sum_\alpha \rho_\alpha$ and $p \equiv \sum_\alpha p_\alpha$ are the total energy density and pressure of the matter in the Universe.

Inserting the Robertson-Walker metric, eq. (1.9), and the total energy momentum tensor, eq. (1.14), into the Einstein equations yields the first and second Friedmann equation for the scale factor $a(t)$:

$$\dot a^2 = \frac{8\pi G}{3}\rho\, a^2 - K\,, \tag{1.15}$$

$$\ddot a = -\frac{4\pi G}{3}(\rho + 3p)\, a\,. \tag{1.16}$$

In this work, we use the dot to denote a derivative with respect to the universal time t, whereas the prime denotes the derivative with respect to the conformal time η. From the first Friedmann equation, eq. (1.15), it is easy to show that the curvature parameter K vanishes if the energy density is equal to the critical density

$$\rho_{\mathrm{crit}} \equiv \frac{3}{8\pi G}\frac{\dot a^2}{a^2}\,. \tag{1.17}$$

We define the density parameters for the constituents of the Universe as the present ratio of their energy density to the critical density,

$$\Omega_\alpha \equiv \frac{\rho_{\alpha,0}}{\rho_{\mathrm{crit},0}}\,, \tag{1.18}$$

where the subscript 0 indicates that we refer to the variable at the present time t_0. We will use this convention throughout the work. A flat universe (i.e. a universe with vanishing curvature) is thus equivalent to $\Omega \equiv \sum_\alpha \Omega_\alpha = 1$.

In the course of the last years, there has been increasing evidence for the Universe being remarkably flat. This evidence comes mainly from measurements of the the cosmic microwave background by WMAP Komatsu et al. (2009), combined with observations of supernovae of type Ia (Riess et al. 1998; Perlmutter et al. 1999). There is also a strong theoretical motivation for a flat universe from the theory of inflation, which we will briefly discuss in section 1.3.3.

Instead of working with the two Friedmann equations, we can replace the second Friedmann equation, eq. (1.16), by the continuity equation

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0\,, \tag{1.19}$$

which can easily be derived from $T^\mu_{0;\mu} = 0$ with the aid of the Christoffel symbols given in Appendix A of Kodama & Sasaki (1984). The continuity equation also holds for the various matter components of the Universe separately, if they are not explicitely coupled to other components.

It is convenient to express the pressure p_α as a function of ρ_α in the equation of state (EoS),

$$p_\alpha = w_\alpha\,\rho_\alpha\,, \tag{1.20}$$

where w_α is called the equation of state parameter, and has the value 1/3 for radiation (including all relativistic particle species), and 0 for cold dark matter and baryons. Using the equation of state, we can write the continuity equation for the components as follows:

$$\dot{\rho}_\alpha + 3\frac{\dot{a}}{a}(1 + w_\alpha)\rho_\alpha = 0\,. \tag{1.21}$$

Integrating this equation, we find the following scaling behaviour for a constant w_α:

$$\rho_\alpha = \rho_{\alpha,0}\,a^{-3\,(1+w_\alpha)}\,, \tag{1.22}$$

where we have adopted the convention $a_0 \equiv 1$. We will use this convention throughout the work.

In order to complete this section about cosmology, let us now introduce some important cosmological quantities which will be used in this work. Consider two astronomical objects, one of which is sitting in the origin of our coordinate system. The *proper distance* d_p between these two objects is defined as the integral over the spacetime interval ds with $dt = d\theta = d\phi = 0$ at time t:

$$d_p \equiv a(t)\int_0^{r_{\mathrm{obj}}} \frac{dr}{\sqrt{1 - Kr^2}}\,. \tag{1.23}$$

The recession velocity of these two objects from one another is

$$v_p \equiv \frac{d}{dt}d_p = \frac{\dot{a}}{a}d_p\,. \tag{1.24}$$

Eq. (1.24) is called the *Hubble law*, and the quantity

$$H(t) \equiv \frac{\dot{a}}{a} \tag{1.25}$$

is is the so-called *Hubble parameter*. It is convenient to define the dimensionless quantity h by

$$H_0 \equiv h \times 100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}\,, \tag{1.26}$$

where H_0 is the present value of the Hubble parameter. From recent experiments, astronomers have determined $h = 0.72 \pm 0.05$ (Komatsu et al. 2009).

The inverse of the Hubble parameter, $\frac{1}{H(t)}$, is called the *Hubble radius* or *Hubble horizon*, and represents the upper limit of the size of regions being in causal contact at time t. This quantity is important in the theory of cosmological structure growth, since it defines the largest scale on which cosmological perturbations can still grow. On scales outside the Hubble horizon, perturbations are frozen in time. Perturbations on scales much smaller than the Hubble horizon are called *subhorizon modes*.

The *redshift z*, which is defined by

$$1 + z \equiv \frac{1}{a(t)}\,, \tag{1.27}$$

is a measure of how much the wavelength of radiation, emitted by a source at time t, has been stretched by the expansion of the Universe since the light has been emitted. It is closely connected to the distance between the emitting source and the Earth. However, the exact relation between redshift and distance depends on the expansion history of the Universe. One of the main challenges of modern cosmology is to obtain information about the expansion history by observing this relation.

1.3 The cosmological concordance model

1.3.1 Constituents

During the last ten years, theory and observations have converged to yield a consistent model of our Universe, the so-called *concordance model*. We believe that our Universe is spatially flat and that it consists of dark energy (~70%), cold dark matter (CDM, ~25%), baryons (~ 5%), and radiation and neutrinos (~10^{-3}%).

It is noteworthy that the baryonic matter, which is the matter we can actually observe in form of galaxies and gas, only contributes about 5% to the total energy content of the Universe. 95% of the constituents of the Universe are not directly observable, and their existence can only be inferred from their gravitational impact. This is somewhat suspicious, and it could actually be a hint that the laws of physics as we know them, in particular the General Theory of Relativity, have to modified on cosmological scales.

Cold dark matter is thought to consist of heavy (and thus non-relativistic) particles, which do not or hardly interact with baryons and photons, and thus do not emit light from which we could observe it. The concept of cold dark matter explains the flat rotation curves of galaxies, the mass of galaxy-clusters obtained from dynamical estimates, the formation of the LSS from small initial fluctuations at the time of last scattering, and gravitational lensing observations. Candidates for dark matter particles are so-called weakly interacting massive particles (WIMPs), the most promising candidate for which is the lightest supersymmetric particle, the neutralino. Attempts to explain the observed behaviours listed above by modifying the laws of gravity rather than by introducing dark matter include *modified Newtonian gravity* (MOND) (Milgrom 1983), and *Tensor-vector-scalar gravity* (TeVeS) (Bekenstein 2004). However, these theories suffer from some difficulties as shown, for instance, by Feix et al. (2008), Klypin & Prada (2009), and Mavromatos et al. (2009).

Dark energy has been postulated in order to explain the observed accelerated expansion of the Universe discovered by Riess et al. (1998) and Perlmutter et al. (1999). We will explain dark energy and its observational evidence in detail in section 1.3.2.

1.3.2 Dark energy

An accelerated expansion of the Universe as observed by Riess et al. (1998) and Perlmutter et al. (1999) cannot be obtained with normal matter (i.e. baryons and dark matter) in a Friedmann-Robertson-Walker cosmological model, as can be easily shown. During an accelerated expansion, the second time derivative of the scale factor a is by definition positive, $\ddot{a} > 0$. Considering the second Friedmann equation, eq. (1.16),

$$\ddot{a} = -\frac{4\pi G}{3}(\rho + 3p)\, a\,,$$

it becomes clear that $\ddot{a} > 0$ only if the effective equation of state (EoS) parameter $w \equiv \frac{p}{\rho}$ of the total energy content of the universe is smaller than $-\frac{1}{3}$. But all known forms of matter including dark matter have $w_\alpha \geq 0$ and can thus only decelerate the expansion. Therefore, if the accelerated expansion is to be explained by some exotic form of energy, the latter is required to have a negative pressure such that the total EoS parameter $w < -\frac{1}{3}$. Such an exotic form of energy was postulated after the observations of Riess et al. (1998) and Perlmutter et al. (1999), and it was named *dark energy*.

In addition to explaining the accelerated expansion, dark energy can account for the missing mass which is necessary to reach the critical density. Measurements of the CMB combined with observations of supernovae of type Ia provide strong evidence of our Universe being spatially flat. As we have seen in section 1.2, the total energy density in a flat universe is necessarily equal to the critical density. But the matter we actually observe by emitted light (baryonic matter) and by its gravitational impact (dark matter) only accounts for about 30% of the critical density. This makes it necessary to postulate some form of energy which accounts for the missing 70% of the critical density. Dark energy naturally accounts for this missing mass.

Another evidence for the existence of dark energy, which is independent of the observations described above, is the integrated Sachs-Wolfe (ISW) effect in the CMB, which we will explain in detail in section 1.4.3. Chapter 3 of this work is devoted to developing optimal methods for the detection of the ISW effect.

There are several different theoretical models for dark energy, or rather, for explaining the observations described above. Some models postulate the existence of an exotic form of energy while leaving General Relativity unchanged, whereas others modify the laws of gravity. In the following, we briefly describe the most common models and indicate how they could be distinguished by observations.

parameter	WMAP 5 year ML	WMAP 5 year +BAO+SN Mean	explanation
Ω_Λ	0.751	0.721 ± 0.015	density parameter of dark energy
Ω_m	$1 - \Omega_\Lambda$	$1 - \Omega_\Lambda$	density parameter of matter (dark + baryonic)
$\Omega_b h^2$	0.02268	0.02265 ± 0.00059	density parameter of baryonic matter $\times h^2$
h	0.724	0.701 ± 0.013	Hubble constant
τ	0.089	0.084 ± 0.016	optical depth to last scattering
n_s	0.961	0.960 ± 0.014	spectral index of the primordial power spectrum
σ_8	0.787	0.817 ± 0.026	Fluctuation amplitude at 8/h Mpc

Table 1.1: Table of the main cosmological parameters of the flat ΛCDM model as given by Komatsu et al. (2009), table 1. The values of the parameters in the second column are the maximum likelihood values for the 5 year WMAP data, which are used in the analysis in chapter 4. The values in the third column are the mean values from combining the 5 year WMAP data with measurements of the baryon acoustic oscillations and supernovae of type Ia. These values are used in chapter 3.

The cosmological constant

One possible candidate for dark energy is the *cosmological constant* Λ, which was originally introduced by Einstein in order to obtain a static universe (Einstein 1917). The Einstein equations with the cosmological constant read

$$G^\mu_\nu + \Lambda\delta^\mu_\nu = 8\pi G\, T^\mu_\nu\,. \tag{1.28}$$

After Hubble discovered the distance-redshift relation of galaxies, which was interpreted as an overall expansion of the Universe, Einstein called the idea of the cosmological constant “the biggest blunder in my life” . However, now the discussion about Λ has been revived by the apparent accelerated expansion of our Universe. Instead of adding the Λ-term to the left hand side (the geometrical part) of the Einstein equations, one can equivalently include it in the energy momentum tensor. In this picture, the cosmological constant corresponds to the vacuum energy density of the Universe and has the constant EoS parameter $w_\Lambda = -1$.

The cosmological model that contains about 70% vacuum energy density, 25% cold dark matter, 5% baryons and 10^{-3}% radiation and neutrinos is referred to as ΛCDM model. The ΛCDM model fits surprisingly well to observations of the CMB, the matter distribution in the Universe, and the distance-redshift relation of supernovae of type Ia. Being the simplest cosmological model comprising dark energy, it is the model which is most widely used by cosmologists. In this work we will use the ΛCDM model with parameter values given by Komatsu et al. (2009), table 1, which are listed in Table 1.1.

Dynamical scalar fields

Instead of considering dark energy to be vacuum energy density, it is also possible to obtain the behaviour of dark energy from dynamical scalar fields. Dynamical scalar fields are fields which are not just sitting in the ground state of their potential, as it is the case for vacuum energy, but the field has started out in some non-equilibrium state and then evolves according to its equations of motion. The main potentially observable difference to the cosmological constant is that the EoS parameter of the dynamical models evolves with time. In principle, it is possible to observe such an evolution, for example by measuring the redshift dependence of the so-called *baryon acoustic oscillations* in the matter power spectrum, which will be introduced in section 1.5.2. There is a wide range of models for dark energy using dynamical scalar fields, including quintessence (Wetterich 1988; Peebles & Ratra 1988), k-essence (Armendariz-Picon et al. 2001), and phantom energy (Caldwell 2002), just to name a few of them.

Other

In what we have described above, we have implicitely assumed that Einstein's General Theory of Relativity is the correct theory to describe our Universe on cosmological scales. This is not necessarily true, and in fact General Relativity has only been confirmed on scales up to solar system scales. There are attempts to reproduce the observations described above by changing the laws of gravity on large scales rather than by introducing dark energy. One example is represented by scalar tensor theories (Hwang 1990a,b). These can, however, be reformulated in terms of GR theories in which a scalar field, possibly representing dark energy, interacts universally with all matter fields.

Another, highly controversial, approach to circumvent dark energy is to use backreactions of inhomogeneities in the Universe on the background expansion (Buchert 2008; Räsänen 2004; Martineau & Brandenberger 2005; Kolb et al. 2008; Kasai et al. 2006). An inhomogeneous Universe may on average evolve differently from a homogeneous solution of Einstein's laws of gravity. This could result in an apparent acceleration of the expansion of space.

Yet another idea of explaining the apparent acceleration is the hypothesis that our observed Universe can be described by a Lemaître-Tolman-Bondi model, which is a spherically symmetric but inhomogeneous dust Universe. Such a model can in principle mimic an accelerated expansion (Enqvist 2008), but in order to preserve the observed isotropy of the CMB, we would presumably need to be placed quite close to the centre of such a structure, violating the Copernican principle (Caldwell & Stebbins 2008). One particular scenario which has been investigated is that we live in a giant void, with a larger Hubble rate inside than out (Tomita 2001). However, it was shown recently that a giant void is not necessary in order to circumvent dark energy in a Lemaître-Tolman-Bondi model, but a large hump would do the job as well (Célérier et al. 2009).

1.3.3 Inflation

There is yet another poorly understood but essential component of the cosmological concordance model, the so-called *inflationary phase* of the very early Universe, about $10^{-35}s$ after the Big Bang. This inflationary period was supposedly a phase of rapid and accelerated expansion, in which the Universe was blown up by a factor of 60 e-folds within a fraction of a second. This exponential expansion is thought to be driven by one or more scalar fields (sometimes called the *inflaton field*), as suggested by Alan Guth and Andrej Linde (Guth 1981; Guth & Weinberg 1983; Linde 1982). Without inflation, we lack a natural explanation for why our Universe should be spatially flat to such a high degree, which is commonly referred to as the *flatness problem*. In addition, in a Friedmann-Robertson-Walker Universe without inflation, there is no reason why the CMB should be isotropic, since the different regions from which we observe the CMB cannot have been in causal contact by the time when they emitted the radiation. This problem is called the *horizon problem*. Inflation solves both the flatness problem and the horizon problem with the same mechanism: The Universe we observe is just a tiny patch of the pre-inflationary Universe, which has been blown up by a factor of about e^{60} by inflation. This implies that the observed Universe is spatially flat, as can be easily seen from dividing eq. (1.15) by a^2:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2}\,. \tag{1.29}$$

The inflaton field has an EoS parameter close to -1, and thus $\rho \approx$ const, see eq. (1.19). Therefore, since the scale factor a is growing with time, the curvature term in the differential equation, K/a^2, eventually becomes negligible compared to the constant density term. The horizon problem is solved by the fact that the different regions we observe in the CMB radiation were all part of the same pre-inflationary patch and thus in thermodynamic equilibrium before inflation.

Furthermore, the theory of inflation naturally explains the presence of small inhomogeneities in the early Universe: During inflation, quantum fluctuations of the inflaton field get blown up to cosmological scales and act as seeds for the formation of cosmological structures. The power spectrum of these initial fluctuations is predicted to be nearly scale-invariant. This has been confirmed by observations of the CMB and can thus be considered the first observed ‘prediction’ of inflation. The seeds for structure formation are furthermore predicted to be close to Gaussian, but to exhibit small non-Gaussianities that depend on the specific model of inflation. Observations of the CMB roughly confirm the Gaussianity of the fluctuations, apart from certain non-Gaussian features that still lack an explanation (Vielva et al. 2006; Ayaita et al. 2009). The detection of small primordial non-Gaussianities of the type predicted by inflation could be a handle to distinguish between the different inflationary models. Furthermore, there is a chance of detecting relics of gravitational waves created during the inflationary phase in the so-called B-mode of the CMB polarization fluctuations, which will be introduced in section 1.4.4. Being the first direct evidence for an inflationary period, such a detection would be one of the greatest successes of modern cosmology.

1.4 The cosmic microwave background

In this section, we introduce the cosmic microwave background radiation (CMB), which is one of the richest sources of information about our Universe that we have. The analysis of different aspects of the CMB will be the main focus of this work. The integrated Sachs-Wolfe (ISW) effect, for which we will develop new detection methods in chapter 3, is explained in detail in section 1.4.3. Other secondary effects on the CMB, which we could in principle apply our methods to, are briefly described.

1.4.1 Origin of the CMB

After the Big Bang, the matter in the Universe was extremely hot and dense, and it successively became cooler and less dense as the Universe expanded. Due to these high temperatures, the hydrogen in the early Universe was ionised, and baryons and photons were tightly coupled via Thomson scattering of the CMB photons by the free electrons. Baryons and photons thus formed the so-called *baryon-photon fluid*. Approximately 400,000 years after the Big Bang, the Universe had become cool enough for the free protons and electrons to combine and form neutral hydrogen, a process which is referred to as *recombination*. After recombination, the photons could no longer get scattered by free electrons, and thus baryons and radiation decoupled. Since then, the photons have been free streaming through the Universe, and can nowadays be observed as the CMB. We measure the CMB radiation in the microwave band, because the wavelength of the photons has been stretched by the expansion of the Universe. We denote the time when the photons have been scattered for the last time the time of *last scattering* (ls). The corresponding surface on our backwards lightcone is referred to as the surface of last scattering. It has a redshift of $z_{\rm ls} \approx 1100$. The CMB features an almost perfect blackbody-spectrum, which tells us that the baryons and photons have been in thermal equilibrium at last scattering.

1.4.2 Temperature anisotropies

The CMB is almost isotropic over the whole sky. There are, however, small temperature-anisotropies of the blackbody-spectrum of $\frac{\Delta T}{T_0} \sim 10^{-5}$, most of which originate in density fluctuations in the baryon-photon-fluid before last scattering. Others, the so-called *secondary effects*, are imprinted in the CMB after last scattering. For notational simplicity, we redefine $T(\hat{\boldsymbol{n}}) \equiv \frac{T(\hat{\boldsymbol{n}})-T_0}{T_0}$, where T_0 denotes the average CMB temperature of 2.725 K, and $\hat{\boldsymbol{n}}$ is the direction on the sky.

The temperature anisotropies $T(\hat{\boldsymbol{n}})$ can be expanded in spherical harmonics Y_{lm}:

$$T(\hat{\boldsymbol{n}}) = \sum_{l,m} a^T_{lm} Y_{lm}(\hat{\boldsymbol{n}})\,. \tag{1.30}$$

Here, the a^T_{lm} are the expansion coefficients, which are defined by

$$a^T_{lm} \equiv \int_S d\Omega\, T(\hat{\boldsymbol{n}})\, Y^*_{lm}(\hat{\boldsymbol{n}})\,, \tag{1.31}$$

where the integral is taken over the whole sphere S and the star denotes complex conjugation. We can define the autocorrelation function of the temperature distribution as

$$\begin{aligned} \langle T(\hat{\boldsymbol{n}})T(\hat{\boldsymbol{n}}')\rangle_{\mathcal{P}(T|p)} &= \sum_{l,l',m,m'} \langle a^T_{lm} a^{T*}_{l'm'}\rangle Y_{lm}(\hat{\boldsymbol{n}}) Y^*_{l'm'}(\hat{\boldsymbol{n}}') \\ &= \sum_l \frac{2l+1}{4\pi} C^T_l\, P_l(\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{n}}')\,, \end{aligned} \tag{1.32}$$

where the average is to be taken over an ensemble of different realisations of T given the cosmological parameters p, i.e. over the probability distribution $\mathcal{P}(T|p)$. The power spectrum C^T_l of the temperature anisotropies is defined by

$$\langle a^T_{lm} a^{T*}_{l'm'}\rangle = \delta_{ll'}\delta_{mm'} C^T_l\,, \tag{1.33}$$

where we have assumed statistical isotropy of the temperature fluctuations. $P_l(x)$ are the Legendre polynomials, which we have introduced into the equation by using the addition theorem for spherical harmonics,

$$\sum_{m=-l}^{l} Y_{lm}(\hat{\boldsymbol{n}}) Y^*_{lm}(\hat{\boldsymbol{n}}') = \frac{2l+1}{4\pi} P_l(\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{n}}')\,. \tag{1.34}$$

Since we only observe one CMB, we will never be able to obtain the average over different CMB realisations in order to measure the power spectrum. We are therefore forced to assume the ‘ergodic hypothesis’ that the average over different directions in the sky gives the same result as an ensemble

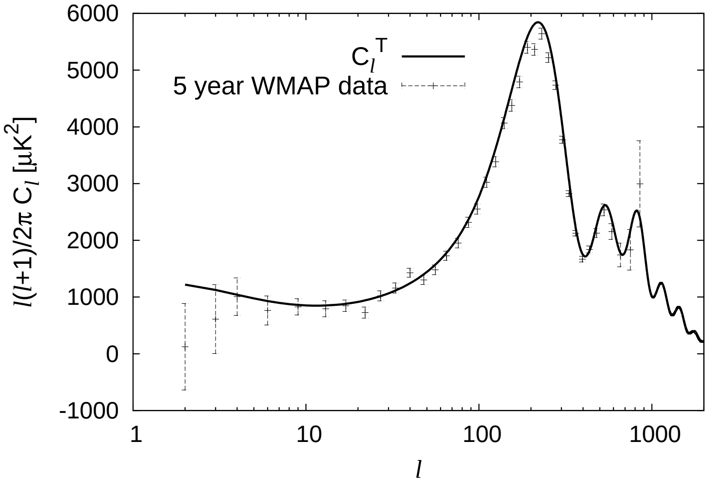


Figure 1.1: Temperature power spectrum of a ΛCDM universe with the current best-fit values of the cosmological parameters as given by Dunkley et al. (2009), together with the WMAP 5 year measurements.

average. This leads to the problem of *cosmic variance*: it is not possible to obtain a good spatial average over large-scale fluctuations, since we simply don't have enough of these patterns to average over. As a straightforward calculation shows, the minimal error with which we can measure the CMB power spectrum from one sky is given by

$$\sqrt{\left\langle \left(\hat{C}_l^T - C_l^T \right)^2 \right\rangle} = \sqrt{\frac{2}{2l+1}}\, C_l^T \,, \tag{1.35}$$

where $\hat{C}_l^T \equiv 1/(2l+1)\sum_m |a_{lm}^T|^2$ is the power spectrum estimated from our CMB realisation, and the average has to be taken over an ensemble of CMB realisations. This cosmic variance is a fundamental limit on the estimation of the power spectrum, which becomes large for low l.

For a given cosmological model, we can calculate the theoretical CMB power spectrum by solving the coupled system of perturbation equations for baryons, dark matter, photons and neutrinos. These perturbation equations comprise the linearised Einstein equations, generalised energy-momentum conservation for baryons and dark matter, and the Boltzmann equation for photons as well as the collisionless Boltzmann equation for neutrinos. A detailed derivation of these equations can be found in Durrer (2001). The temperature power spectrum has been measured with high precision by different experiments (Nolta et al. 2009; Masi et al. 2007; Reichardt et al. 2009; Sievers et al. 2009). In Fig. 1.1, we plot the temperature power spectrum of a ΛCDM universe with the current best-fit values of the cosmological parameters, as given by Dunkley et al. (2009), together with the WMAP 5 year measurements, which are almost cosmic variance limited for very low l.

The most prominent feature of the CMB spectrum are the acoustic peaks, which originate at the time before last scattering when baryons and photons are still tightly coupled and form the baryon-photon fluid. Dark matter, which is not coupled to the photons, had already started to clump and form potential wells well before last scattering. The radiation pressure from the photons resists the gravitational compression of the fluid into these potential wells and sets up acoustic oscillations in the fluid. The frequency of a mode with wavelength λ is given by $\nu = \frac{c_s}{\lambda}$, c_s being the sound speed. Therefore, the modes with shorter wavelengths have time to oscillate more often before last scattering than the longer wavelength modes. Consider the 'zero-mode', λ_0, for which there has been just enough time to go through half an oscillation before last scattering. It is the longest wavelength mode that is at the maximum of its oscillation at last scattering. Therefore, it leads to the first acoustic peak in the CMB-spectrum. Half of the wavelength of this mode is equal to the distance d_s which sound can travel before last scattering, as one can easily see: $d_s = \frac{1}{2\nu_0}c_s = \frac{1}{2}\lambda_0$. This distance is commonly referred to as the *sound horizon* at last scattering. The mode with half the wavelength of the zero-mode oscillates twice as fast and therefore undergoes one full oscillation before last scattering. At last scattering, it will also be at the maximum of its oscillation and corresponds to the second peak of the spectrum. The other peaks are due to higher harmonics of the zero-mode.

On large scales, i.e. for small l's, the power spectrum is dominated by the Sachs-Wolfe (SW) and the integrated Sachs-Wolfe (ISW) effect. Both are related to the gravitational redshift a photon experiences when climbing out of a gravitational potential well. At last scattering, the CMB photons are sitting in potential wells and on potential hills created by the density fluctuations of dark matter. When the photons are set free at last scattering, they have to climb out of the wells, or fall off the hills, and thus get red- or blueshifted, respectively. This is referred to as the Sachs-Wolfe effect. The integrated Sachs-Wolfe effect will be explained in detail in section 1.4.3.

1.4.3 Secondary anisotropies

The temperature anisotropies we have explained in the last section are primordial anisotropies originating at the surface of last scattering. However, the matter inhomogeneities the photons pass through on their way towards us, leave imprints on the CMB fluctuations in various ways. These imprints are called secondary effects. The most important secondary effects are the ISW effect, the Rees-Sciama (RS) effect, the thermal and kinetic Sunyaev-Zel'dovich (SZ) effects, and gravitational lensing. We describe these effects in the following.

The integrated Sachs-Wolfe effect

In section 1.4.2, we have explained the SW effect as the gravitational redshift of photons that have to climb out of potential wells at the surface of last scattering. The same happens when the CMB photons pass through gravitational wells or hills on their way to us. If the gravitational potential of a cosmic matter structure is static, the net frequency shift for a photon travelling through it is zero. However, if the depth of the potential changes in time the red- and blueshift no longer cancel, and the photon experiences a net frequency shift. On smaller scales, the frequency shifts from changing potentials average out, because the photon passes through many small overdense and underdense regions on its way to us. On the largest scales, however, the photon will only pass through few potential wells and hills, and there will remain a net frequency shift, the ISW effect, which can be observed on the large angular scales of the CMB. On those scales, cosmological structure growth can still be described by linear perturbation theory.

In an Einstein-de Sitter universe, i.e. a flat matter dominated universe, the gravitational potentials are constant on linear scales, and thus no ISW effect can be observed. However, in a universe which contains dark energy, gravitational potentials decay with time and give rise to an ISW effect. The same holds if the apparent acceleration of the Universe is not caused by dark energy, but is a consequence of modified gravity, which we have introduced in section 1.3. Curvature of space also causes gravitational potentials to decay, but as we said before there is strong evidence against such a curvature, so that the ISW effect is considered to be a probe of dark energy or modified gravity.

On non-linear scales, on which the matter fluctuations are decoupled from the background expansion, gravitational potentials grow due to the matter infall into potential wells. This non-linear ISW effect is called Rees-Sciama (RS) effect. It is quite small since the matter flow velocities causing the changing potential are well below 1 per cent.

The temperature anisotropies coming from the integrated Sachs-Wolfe effect are given by

$$T_{\mathrm{isw}}(\hat{\boldsymbol{n}}) = 2 \int_{\eta_{\mathrm{ls}}}^{\eta_0} \Psi' \left(\eta, (\eta_0 - \eta)\, \hat{\boldsymbol{n}}\right)\, d\eta\,. \tag{1.36}$$

Recall that η denotes the conformal time and the prime stands for the derivative with respect to the latter. η_{ls} and η_0 denote the conformal time at last scattering and the present epoch, respectively, and $\hat{\boldsymbol{n}}$ is the direction on the sky. Ψ is the gauge invariant Bardeen potential (Bardeen 1980), which coincides with the Newtonian gravitational potential in the Newtonian gauge[1] used in this work. Note that the integral in the above equation has to be taken along the backwards light cone.

In Newtonian gauge, T_{isw} can be obtained by applying a suitably constructed linear operator $\mathcal{Q}$ to the present matter density contrast $\delta_m(\eta_0)$:

$$T_{\mathrm{isw}} = \mathcal{Q}\, \delta_m(\eta_0)\,. \tag{1.37}$$

The matter density contrast is defined as $\delta_m(\boldsymbol{x}) \equiv \left[\rho_m(\boldsymbol{x}) - \bar{\rho}_m\right] / \bar{\rho}_m$, where $\rho_m(\boldsymbol{x})$ denotes the density of matter in the Universe at position $\boldsymbol{x}$, and $\bar{\rho}_m$ is the background matter density. Eq. (1.37) can be verified by using the perturbation equations derived by, e.g., Kodama & Sasaki (1984) or Durrer (2001): In order to obtain the expression for the operator $\mathcal{Q}$ in the subhorizon-limit, let us look at the Poisson equation

$$\Delta\Psi = \frac{3H_0^2}{2}\,(1+z)\,\Omega_m\,\delta_m\,, \tag{1.38}$$

where Δ denotes the Laplace operator in comoving coordinates. From the Poisson equation, we obtain

$$\Psi'(\boldsymbol{k},\eta) = \frac{3H_0^2\Omega_m}{2k^2}\, H(\eta)\, \left(1 - f(\eta)\right)\, \mathcal{D}(\eta)\, \delta_m(\boldsymbol{k},\eta_0)\,, \tag{1.39}$$

where $f \equiv d\ln\delta_m / d\ln a$ is the growth function, $\mathcal{D}(\eta) \equiv \delta_m(\boldsymbol{k},\eta)/\delta_m(\boldsymbol{k},\eta_0)$ denotes the linear growth factor, k stands for the absolute value of $\boldsymbol{k}$, and we define Fourier transformed quantities as

$$\delta_m(\boldsymbol{k},\eta) = \int d^3x\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\, \delta_m(\boldsymbol{x},\eta)\,, \tag{1.40}$$

with the inverse transformation

$$\delta_m(\boldsymbol{x},\eta) = \frac{1}{(2\pi)^3} \int d^3k\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\, \delta_m(\boldsymbol{k},\eta)\,. \tag{1.41}$$

The expression for the operator $\mathcal{Q}$ can then be obtained by Fourier transforming eq. (1.39) and

---

[1] For a detailed explanation of the gauge freedom in linear perturbation theory, and a definition of Newtonian gauge, see Kodama & Sasaki (1984).

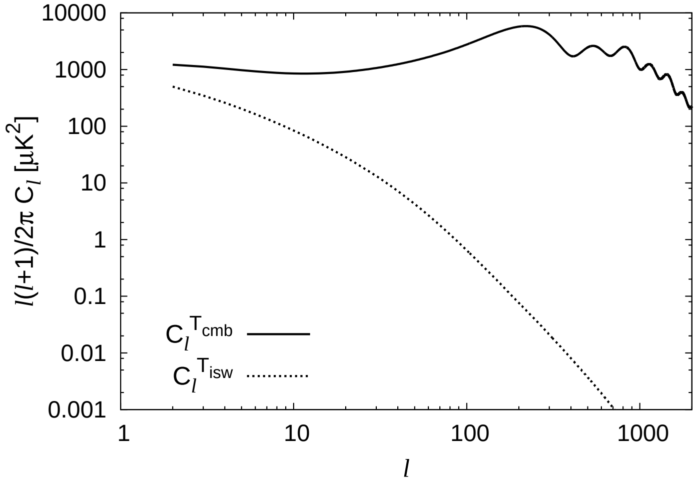


Figure 1.2: CMB and ISW power spectrum of a ΛCDM universe with the current best-fit values of the cosmological parameters as given by Dunkley et al. (2009).

inserting it into eq. (1.36). Note, though, that we have not used the subhorizon-limit in this work, as eq. (1.37) is valid on superhorizon-scales as well.[2] In Fig. 1.2, we compare the power spectrum of the ISW effect with the one of the CMB. Note that the ISW effect only significantly contributes to the total CMB power spectrum at the lowest multipoles.

The Sunyaev-Zel'dovich effect

The Sunyaev-Zel'dovich (SZ) effect (Sunyaev & Zeldovich 1972) is the interaction of CMB photons with ionised gas in galaxy clusters and filaments. We distinguish between the thermal and the kinetic SZ effect. In both cases, the relatively cool CMB photons undergo inverse Compton scattering by free electrons of the X-ray gas in the cluster, i.e. the photons get kicked to higher or lower energies while the electrons lose or gain energy.

In the case of the thermal SZ effect, the energy kick of the CMB photon is taken from the thermal energy of electrons in clusters. Since the temperature in clusters is much higher than the temperature of the CMB photons, the photons are on average up-scattered in energy. This causes a deviation of the photon energy distribution of the CMB from the blackbody spectrum. Therefore, the thermal SZ effect can be separated from the primordial CMB by using its spectral signature. The thermal SZ effect leaves an imprint on the CMB spectrum at very small scales and might even dominate it at $l > 2000$.

In the case of the kinetic SZ effect, the energy change of the scattered photon comes from the bulk motion of the ionised gas. The kinetic SZ effect does not destroy the blackbody spectrum

---

[2] The correct formula for Q in Newtonian gauge, which also holds on superhorizon-scales, can be obtained by differentiating and Fourier transforming the expression

$$\Psi(\boldsymbol{k},\eta) = \exp\left(-\int_0^{\eta} p\,(\boldsymbol{k},\eta')\,d\eta'\right)\int_0^{\eta}\frac{H_0^2\,\Omega_m}{2Ha^2}\mathcal{D}(\boldsymbol{k},\eta')\,\delta_m(\boldsymbol{k},\eta_0)\exp\left(\int_0^{\eta'} p(\boldsymbol{k},\eta'')\,d\eta''\right)d\eta'\,,$$

and inserting it into eq. (1.36), instead of the expression for $\Psi(\boldsymbol{k},\eta)$ in the subhorizon-limit, eq. (1.39). Here, we have defined $p(\boldsymbol{k},\eta) \equiv \frac{k^2+3a^2H^2}{3aH}$ and the linear growth factor $\mathcal{D}(\boldsymbol{k},\eta) \equiv \frac{\delta(\boldsymbol{k},\eta)}{\delta(\boldsymbol{k},\eta_0)}$, which in general depends on the Fourier mode $\boldsymbol{k}$.

shape of the CMB but only changes its temperature. Therefore, it cannot be distinguished from the primordial CMB signal without additional information about the distribution of the clusters which create the SZ effect. The kinetic SZ signal of clusters is typically several times smaller than their thermal SZ effect.

Gravitational lensing

According to Einstein's General Theory of Relativity, the curvature of spacetime determines how massive particles and light rays propagate through space. The resulting deflection of light by gravitational potentials is called *gravitational lensing*. Einstein's prediction of the deflection angle of starlight which gets deflected by the sun has been confirmed by Arthur Eddington's measurements during a solar eclipse in 1919 (Dyson et al. 1920), and was one of the great successes of general relativity.

The CMB photons are subject to gravitational lensing when travelling through the inhomogeneous gravitational field of the large-scale structure (LSS) on their way from the last scattering surface to us. This causes a change in size and shape of the warm and cold patches in the CMB temperature fluctuations, and also affects the polarization of the CMB, which we will introduce in the next section. The effect of gravitational lensing manifests itself in the CMB power spectra by smoothing out sharp features in the latter. These effects contribute about 1 per cent to the CMB power spectra at $l \sim 400$, and become larger than the primordial fluctuations at $l \gtrsim 3000$.

1.4.4 CMB polarization

Thomson scattering, which is the dominant process that couples photons to free electrons before last scattering, creates linear polarization of the CMB photons. In the following, we give an intuitive explanation of this process. For a thorough discussion of CMB polarization, the reader is referred to Durrer (2001) or Zaldarriaga & Seljak (1997).

We consider a linearly polarized incoming wave with polarization direction $\hat{\boldsymbol{\epsilon}}'$, which is scattered into an outgoing wave with polarization direction $\hat{\boldsymbol{\epsilon}}$ (the polarization direction is defined as the direction in which the electric field oscillates). The differential Thomson scattering cross section for this process is given by Jackson (1975)

$$\frac{d\sigma}{d\Omega} = \frac{3\sigma_T}{8\pi} |\hat{\boldsymbol{\epsilon}}' \cdot \hat{\boldsymbol{\epsilon}}| \,, \tag{1.42}$$

where σ_T is the total Thomson cross section. For an unpolarized incoming beam, this leads to a differential cross section which is independent of the scattering angle θ if $\hat{\boldsymbol{\epsilon}}$ is perpendicular to the scattering plane, and proportional to $\cos^2(\theta)$ if $\hat{\boldsymbol{\epsilon}}$ lies in the scattering plane. If the incident photons come in isotropically from all directions, there is no net polarization of the scattered photons. However, if the intensity of the incoming radiation exhibits a quadrupole moment in the rest frame of the scattering electron, Thomson scattering leads to some linear polarization of the outgoing beam. This can be intuitively understood by considering two incident light beams coming in from directions separated by 90° (cf. Fig. 1.3). From the mathematical derivation (Durrer 2001; Zaldarriaga & Seljak 1997), it becomes clear that among all incident multipoles the quadrupole is the only source of polarization of the outgoing beam.

Due to the process described above, the CMB photons leaving the surface of last scattering are linearly polarized to a certain degree. After recombination, nearly all of the free electrons and

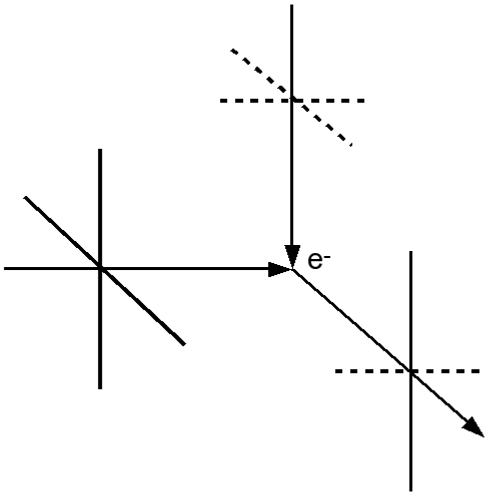


Figure 1.3: Two unpolarized light beams are coming in from directions separated by 90°. The light beam coming in from the left has a higher intensity, as indicated by the longer polarization vectors. The outgoing light beam is linearly polarized to a certain degree.

protons have combined to form Hydrogen, and thus the probability of the photons getting scattered is very low. Only during the epoch of reionization at $z \sim 10-20$, there are again free electrons created by ionising photons from the first supernovae, and about 10 per cent of the CMB photons get scattered again. The free-streaming of the photons coming from the last scattering surface has enhanced the quadrupole incident on the scattering electrons at reionization (Zaldarriaga 1997). Therefore, the re-scattering of CMB photons during reionization creates additional polarization on large scales.

Let us now introduce the mathematical framework used to describe linear polarization on a sphere, which has been developed by Zaldarriaga & Seljak (1997). Consider a coordinate system with the z-direction pointing towards us along the line of sight. The electric field of a monochromatic electromagnetic wave travelling in z-direction is of the form $\boldsymbol{E} = \boldsymbol{E_0}\, e^{i(kz-\omega t)}$, where k and ω are wave-vector and frequency, respectively. $\boldsymbol{E_0} \equiv (E_0^x\, \hat{\boldsymbol{\epsilon}}_{(x)} + E_0^y\, \hat{\boldsymbol{\epsilon}}_{(y)})$ is the complex amplitude of the electromagnetic wave, with $\hat{\boldsymbol{\epsilon}}_{(x)}$ and $\hat{\boldsymbol{\epsilon}}_{(y)}$ being the unit vectors in x and y direction, respectively. We can characterise a linearly polarized monochromatic wave by the three Stokes parameters I, Q, and U, which are defined as

$$I \equiv |E_x|^2 + |E_y|^2 \,, \tag{1.43}$$

$$Q \equiv |E_x|^2 - |E_y|^2 \,, \tag{1.44}$$

$$U \equiv E_x^* E_y + E_x E_y^* \,, \tag{1.45}$$

where I is proportional to the total intensity of the wave, Q encodes the polarized intensity in the directions of $\hat{\boldsymbol{\epsilon}}_{(x)}$ and $\hat{\boldsymbol{\epsilon}}_{(y)}$, and U describes the polarized intensity in the directions inclined by 45° to that. I is related to the temperature of the CMB introduced in section 1.4.2 by[3]

$$\frac{\delta T}{T} = \frac{1}{4}\frac{\delta I}{I} \,. \tag{1.46}$$

---

[3] Note that we had redefined $T(\hat{\boldsymbol{n}}) \equiv \frac{T(\hat{\boldsymbol{n}})-T_0}{T_0}$ in section 1.4.2, which we have ignored in eq. (1.46) for clarity.

Since the Stokes parameters depend on the chosen coordinate system, they are not well suited for describing polarization on the full sphere (Zaldarriaga & Seljak 1997). However, the combinations

$$Q \pm iU\,, \tag{1.47}$$

where i denotes the imaginary unit, transform as spin-2 quantities under rotation around the z-axis. They can therefore be expanded in spin-weighted spherical harmonics, ${}_sY_{lm}(\hat{\boldsymbol{n}})$, with the index $s = \pm 2$:

$$\begin{aligned}(Q + iU)(\hat{\boldsymbol{n}}) &= \sum_{l,m} a_{2,lm}\,{}_2Y_{lm}(\hat{\boldsymbol{n}})\,, \\ (Q - iU)(\hat{\boldsymbol{n}}) &= \sum_{l,m} a_{-2,lm}\,{}_{-2}Y_{lm}(\hat{\boldsymbol{n}})\,, \end{aligned} \tag{1.48}$$

where the expansion coefficients can be obtained from the following equations:

$$\begin{aligned} a_{2,lm} &\equiv \int d\Omega\,{}_2Y^*_{lm}(\hat{\boldsymbol{n}})(Q + iU)(\hat{\boldsymbol{n}})\,, \\ a_{-2,lm} &\equiv \int d\Omega\,{}_{-2}Y^*_{lm}(\hat{\boldsymbol{n}})(Q - iU)(\hat{\boldsymbol{n}})\,. \end{aligned} \tag{1.49}$$

Instead of working with the coefficients $a_{\pm 2,lm}$, it is convenient to introduce their linear combinations, the so-called E and B modes:

$$a^E_{lm} \equiv -\frac{1}{2}(a_{2,lm} + a_{-2,lm})\,, \tag{1.50}$$

$$a^B_{lm} \equiv \frac{i}{2}(a_{2,lm} - a_{-2,lm})\,. \tag{1.51}$$

Under parity transformations, the E mode remains unchanged whereas the B mode changes sign, in analogy to electric and magnetic fields.

With the above definitions, we can fully characterise the statistics of CMB fluctuations by the following four power spectra:

$$\langle a^T_{lm} a^{T*}_{l'm'} \rangle = C^T_l\,\delta_{ll'}\delta_{mm'}\,, \tag{1.52}$$

$$\langle a^T_{lm} a^{E*}_{l'm'} \rangle = C^{TE}_l\,\delta_{ll'}\delta_{mm'}\,, \tag{1.53}$$

$$\langle a^E_{lm} a^{E*}_{l'm'} \rangle = C^E_l\,\delta_{ll'}\delta_{mm'}\,, \tag{1.54}$$

$$\langle a^B_{lm} a^{B*}_{l'm'} \rangle = C^B_l\,\delta_{ll'}\delta_{mm'}\,. \tag{1.55}$$

Note that the TB and EB cross power spectra vanish in a universe symmetric under parity transformation, because B has the opposite parity of T and E.

In Fig. 1.4, we plot the theoretical TE cross power spectrum for the WMAP 5 year best fit cosmological model, together with the values measured by WMAP. In Fig. 1.5, we plot C^T_l and C^E_l for comparison. Note that the power contained in CMB polarization fluctuations is about a factor of 100 less than the power contained in the temperature fluctuations. Nevertheless, the polarization of the CMB contains valuable information about our Universe. The E-mode can be used to obtain information about the epoch of reionization, and the B mode, which contains only a fraction of

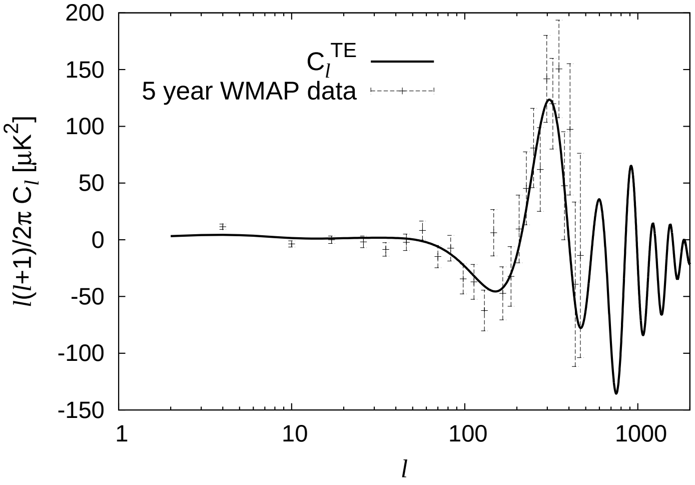


Figure 1.4: TE cross power spectrum of a ΛCDM universe with the current best-fit values of the cosmological parameters as given by Dunkley et al. (2009), together with the values observed by WMAP.

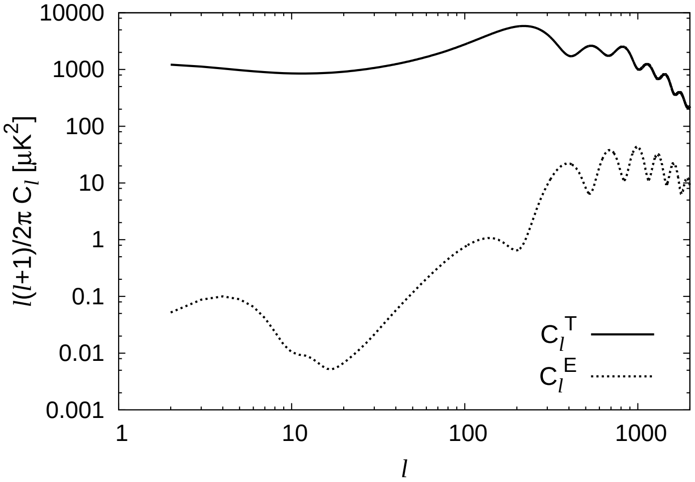


Figure 1.5: TT and EE power spectrum of a ΛCDM universe with the current best-fit values of the cosmological parameters as given by Dunkley et al. (2009).

the power of the E-mode, is considered to be a smoking gun for detecting primordial gravitational waves created during inflation.

The currently available full-sky polarization data from WMAP are highly contaminated by detector noise and Galactic foregrounds. However, the *Planck* satellite, which has been launched in May 2009, will soon provide us with full-sky polarization data of unprecedented accuracy. Whereas *Planck* will presumably only detect the E-mode fluctuations, the next generation of CMB experiments, such as PolarBeaR[4] or CMBPol[5] are designed to measure the B-mode as well. Given that inflation is at present merely a hypothesis without any observational evidence, measuring the primordial gravitational waves would be a tremendous success for the standard model of cosmology, in which inflation plays a crucial role.

1.5 Other observational probes of the Universe

In addition to the CMB explained in the last section, there are a number of other observational probes that are used to infer information about our Universe. In the following, we briefly introduce the most important probes and discuss the constraints that can be obtained from their analysis.

1.5.1 The luminosity distance-redshift relation

The relation of the distance of astronomical objects to their redshift encodes information about the geometry and expansion history of the Universe. The proper distance to an object as defined in section 1.2 corresponds to the distance measured by a chain of rulers at a fixed time t. It is not possible to determine the proper distance by observation, because we observe astronomical objects through their light, which takes a certain time to reach us. Therefore, we have to rely on observationally motivated distance measures to characterise the distance to astronomical objects.

An important distance measure is the so-called luminosity distance d_L, which is defined such as to preserve the Euclidean inverse-square law for the diminution of light:

$$d_L \equiv \sqrt{\frac{L}{4\pi f}}\,. \tag{1.56}$$

Here, L is the absolute luminosity emitted by the source at time t and comoving coordinate distance r, and f is the flux we observe at time t_0.

The area of a sphere centred on the source and passing through the position of the observer at time t_0 is just $4\pi r^2$, as can be easily seen from the Robertson-Walker metric, eq. (1.9). In an expanding Universe, the diminution of light is proportional to an additional factor of a^2, due to the redshift of the single photons and to the dilution of the number of photons:

$$f = \frac{L}{4\pi r^2} a^2\,, \tag{1.57}$$

from which we readily obtain

$$d_L = \frac{r}{a}\,. \tag{1.58}$$

---

[4] `http://bolo.berkeley.edu/polarbear/index.html`

[5] Baumann et al. (2008), `http://cmbpol.uchicago.edu`

Expanding the scale factor a in a Taylor series around t_0 and integrating the Robertson-Walker metric for a light ray (for which $ds^2 = 0$), it is quite straightforward to obtain the relation between the distance coordinate r and the redshift z (Coles & Lucchin 2002):

$$r = \frac{1}{H_0}\left[z - \frac{1}{2}(1+q_0)z^2 + O(z^3)\right], \tag{1.59}$$

where q_0 is called the deceleration parameter, which is defined as

$$q_0 \equiv -\frac{\ddot{a}(t_0)}{\dot{a}(t_0)^2}. \tag{1.60}$$

Inserting this relation into eq. (1.58), we obtain the luminosity distance-redshift relation

$$d_L = \frac{1}{H_0}\left[z + \frac{1}{2}(1-q_0)z^2 + O(z^3)\right]. \tag{1.61}$$

If the absolute luminosity L of an object and thus d_L is known, the distance-redshift relation given in eq. (1.61) can be used to infer information about the Hubble parameter H_0 and the deceleration parameter q_0. Examples for astronomical objects of known absolute luminosity, the so-called *standard candles*, are certain variable stars (Cepheids), and supernovae of type Ia. In case of the supernovae of type Ia, an empirical relation has been found between the shape of the light-curve emitted by the supernova and its absolute luminosity. Up to present times, however, we still lack a theoretical understanding of this relation. Nevertheless, supernovae of type Ia are considered reliable standard candles, and have been used to determine the deceleration parameter of the Universe by Riess et al. (1998) and Perlmutter et al. (1999), with the surprising result that the expansion is actually accelerating rather than decelerating (cf. section 1.3.2).

1.5.2 Angular distances

In analogy to the luminosity distance introduced in the last section, we can define the so-called angular diameter distance d_A. It is defined in such a way as to preserve the variation of the angular size of an object in Euclidean space. Let D_s be the proper (physical) diameter of a source placed at coordinate r at time t, and $\Delta\theta$ the angle on the sky subtended by the source. We then define the angular diameter distance to be

$$d_A \equiv \frac{D_s}{\Delta\theta}. \tag{1.62}$$

From the Robertson-Walker metric, eq. (1.9), we obtain the relation

$$D_s = ar\Delta\theta, \tag{1.63}$$

which we insert into eq. (1.62) in order to obtain

$$d_A = ar. \tag{1.64}$$

Just as for the luminosity distance, we can again expand the expression for the angular diameter distance, eq. (1.64), in a Taylor series in z. For astronomical objects of known physical diameter, the so-called *standard rulers*, we can then use the angular diameter distance to infer the Hubble parameter and the deceleration parameter.

The potentially most powerful standard ruler is the baryon acoustic oscillations (BAOs) in the power spectra of the CMB and the matter distribution. The BAOs originate in the oscillations of the baryon-photon plasma at the time before last scattering. We have already described these oscillations and the imprint they leave in the CMB spectrum in section 1.4. We have also explained that the physical scale of the first peak in the CMB spectrum is given by the sound horizon at last scattering. The BAOs are visible not only in the spectrum of the CMB, but also in the matter power spectrum, which can in principle be measured up to a redshift of about 7 or higher from the galaxy and quasar distributions and from 21cm observations. By determining the BAO scale at different redshifts, it is thus in principle possible to actually measure the expansion history of the Universe. Therefore, BAOs are considered to be among the most promising future sources of information about dark energy.

1.5.3 Weak lensing

In section 1.4.3, we have already introduced gravitational lensing as a secondary effect on the CMB. Just like the CMB photons, also the light emitted from galaxies gets deflected by gravitational potentials it passes through, resulting in a distortion and magnification of the observed galaxy images. Since gravitational lensing is sensitive to anything that creates a gravitational potential, it can be used to obtain maps of the projected dark matter distribution and to determine the dark matter power spectrum. This is done by analysing gravitational lensing in the weak lensing regime, in which the distortions of the galaxy images can be described by a complex field called *shear*.

Obtaining the shear from the elliptical distortion of galaxy images involves taking the mean over a large sample of galaxies, in order to average out the intrinsic shapes of the galaxies. At present, weak lensing measurements therefore still suffer from a high noise-level. Nevertheless, once there are enough galaxies to average over, weak lensing measurements provide a potentially very powerful tool to map our Universe.

1.5.4 Primordial nucleosynthesis

Within the first three minutes after the Big Bang, the light elements deuterium, ^{3}He, ^{4}He and ^{7}Li have formed from Hydrogen during primordial nucleosynthesis (Big Bang nucleosynthesis, BBN). The relative abundance of these elements after BBN depends on the baryon content of the Universe. This allows us to obtain bounds on the density parameter of baryons, Ω_b, by comparing the observed abundances of the light elements with predictions of BBN computations.

Chapter 2

Statistical inference

Note: Section 2.3 of this chapter is taken from Enßlin et al. (2009), to which I have contributed as a second author, but which was primarily the work of Torsten Enßlin.

Modern-day cosmology heavily relies on the use of probability theory. Cosmological models make predictions of the statistical properties of cosmological signals such as the CMB, which we compare to the statistical properties of the observed signal. In this chapter, we give a brief introduction to the basic concepts of Bayesian inference used in this PhD work. For a more thorough treatment, please refer to Bolstad (2004), Robert (2001), or Gelman et al. (2004).

2.1 Notation

Most of the random variables we will be dealing with in this work, including certain cosmological signals and detector noise, can be approximately described by a Gaussian distribution. For simplifying the notation, we therefore define

$$\mathcal{G}(\chi, C) \equiv \frac{1}{\sqrt{|2\pi C|}} \exp\left(-\frac{1}{2}\chi^{\dagger}\, C^{-1}\chi\right) \tag{2.1}$$

to denote the probability density function of a Gaussian distributed vector χ with zero mean. By 'vector' we generally mean a vector in function space (i.e. a field), for example the value of the CMB temperature fluctuations as a function of position on the sphere. The covariance matrix, C, is defined as

$$C \equiv \langle \chi\chi^{\dagger}\rangle_{\mathcal{G}(\chi,C)} \equiv \int \mathcal{D}\chi\,(\chi\chi^{\dagger})\,\mathcal{G}(\chi, C)\,, \tag{2.2}$$

where the integral has to be taken over all possible field configurations of χ. We usually work with pixelised quantities, in which case this translates to an integral over all pixels i,

$$\mathcal{D}\chi \equiv \Pi_{i=1}^{N_{\rm pix}} d\chi^{i}\,. \tag{2.3}$$

Note that in general the covariance matrix depends on the cosmological parameters p, which is not explicitly stated in our notation. A daggered vector or matrix denotes its transposed and complex conjugated version, as usual. Hence, given two vectors a and b, $a\,b^{\dagger}$ must be read as the tensor product, whereas $a^\dagger b$ denotes the scalar product.

2.2 Bayesian inference

The Bayesian approach to probability theory (Bayes 1763) is to interpret missing knowledge as probabilistic uncertainty. It is therefore better suited for cosmological problems than the frequentist interpretation of probability as relative frequency of occurrence, since we have no possibility to do a cosmological 'experiment' many times in a row. Bayesian statistics can furthermore be used to assign probabilities to the values of the parameters of our statistical models, and even to the underlying models themselves. Thus, Bayesian inference is the method best adapted to constrain cosmological parameters, and it is also widely used in other contexts in cosmology. In this work, we sometimes talk about a *sample of universes*, in order to visualise the uncertainty of our signal in question. We understand this as a *hypothetical* sample of universes, though, which does not necessarily have to exist.

Let us consider a signal s which we are interested in and which represents some specific aspect of the physical state of our Universe, for example the fluctuations of the CMB radiation or the matter distribution in the Universe. Since the signal does not contain the full physical state of the Universe, any degree of freedom which is not present in the signal but influences the data will be received as probabilistic uncertainty, or shortly noise.

In order to infer information about the signal from the data, we need to specify a statistical model describing our state of knowledge about the signal *before* the data are observed. This knowledge can be taken from theory, or from previous measurements of data containing information about the signal. The corresponding probability distribution, $\mathcal{P}(s)$, which we assign to the signal is called the *prior distribution* or simply the *prior*.

Our state of knowledge about the signal *after* the data have been measured is described by the *posterior distribution* or *posterior*, $\mathcal{P}(s\,|\,d)$. This is the probability distribution of the signal conditional on the measured data. Usually, we can not easily write down a model for the posterior. It is much more straightforward to define the so-called *likelihood* for the signal from theoretical modelling, i.e. the probability distribution of the data given the signal, $\mathcal{L}(s) \equiv \mathcal{P}(d\,|\,s)$. From the prior and the likelihood we then obtain the posterior via Bayes' Theorem:

$$\mathcal{P}(s\,|\,d) = \frac{\mathcal{P}(d\,|\,s)\,\mathcal{P}(s)}{\mathcal{P}(d)}\,, \tag{2.4}$$

where the normalisation,

$$\mathcal{P}(d) \equiv \int \mathcal{D}s\,\mathcal{P}(d\,|\,s)\,\mathcal{P}(s)\,, \tag{2.5}$$

is called the *evidence*. The posterior is the key quantity from which we deduce information about our signal s. The evidence also plays a central role in Bayesian inference, since it is used to assign probabilities to the statistical models in Bayesian model selection.

2.3 Information field theory

Throughout this work, we make the assumption that all the processes we are dealing with are well described by Gaussian probability distributions, which holds very well for everything done in this thesis. However, the work on the ISW detection presented in chapter 3 can be extended to other secondary effects in the CMB, for which the assumption of Gaussianity is rather poor. This can be done using the mathematical framework of *information field theory* (IFT) presented in Enßlin et al. (2009). In this section, we give a brief introduction to the general framework of IFT.

Let us assume we want to infer information about a signal s from its posterior distribution $\mathcal{P}(s\,|\,d)$ via Bayesian inference. This is very straightforward as long as the posterior is Gaussian, but it quickly becomes very difficult for non-Gaussian posteriors. IFT provides us with approximate solutions of all moments of a non-Gaussian distributed signal, provided that the prior of the signal as well as the likelihood are known or can at least be Taylor-Fréchet-expanded around some reference field configuration t. Then Bayes' Theorem permits us to express the posterior as

$$\mathcal{P}(s\,|\,d) = \frac{\mathcal{P}(d\,|\,s)\,\mathcal{P}(s)}{\mathcal{P}(d)} \equiv \frac{1}{Z}\,e^{-H[s]}\,, \tag{2.6}$$

where we have introduced the Hamiltonian

$$H[s] \equiv H_d[s] \equiv -\log\left[\mathcal{P}(d,s)\right] = -\log\left[\mathcal{P}(d\,|\,s)\,\mathcal{P}(s)\right]\,. \tag{2.7}$$

Let us also define the partition function $Z \equiv Z_d$ as

$$Z \equiv \mathcal{P}(d) \equiv \int \mathcal{D}s\;\mathcal{P}(d\,|\,s)\,\mathcal{P}(s) = \int \mathcal{D}s\;e^{-H[s]}\,. \tag{2.8}$$

It is extremely convenient to include a moment generating function into the definition of the partition function

$$Z_d[J] \equiv \int \mathcal{D}s\;e^{-H[s]+J^\dagger s}\,. \tag{2.9}$$

This means $\mathcal{P}(d) = Z = Z[0]$, but also permits us to calculate any moment of the signal field via Fréchet-differentiation of eq. (2.9),

$$\langle s(\boldsymbol{x}_1)\cdots s(\boldsymbol{x}_n)\rangle_{\mathcal{P}(s|d)} = \frac{1}{Z}\left.\frac{\delta^n Z_d[J]}{\delta J(\boldsymbol{x}_1)\cdots\delta J(\boldsymbol{x}_n)}\right|_{J=0}\,, \tag{2.10}$$

where the average is taken over the posterior distribution of the signal.

Of special importance are the so-called connected correlation functions or cumulants

$$\langle s(\boldsymbol{x}_1)\cdots s(\boldsymbol{x}_n)\rangle^{\mathrm{c}}_{\mathcal{P}(s|d)} \equiv \left.\frac{\delta^n \log Z_d[J]}{\delta J(\boldsymbol{x}_1)\cdots\delta J(\boldsymbol{x}_n)}\right|_{J=0}\,, \tag{2.11}$$

which are corrected for the contribution of lower moments to a correlator of order n. For example, the connected mean and dispersion are expressed in terms of their unconnected counterparts as:

$$\begin{aligned}
\langle s(\boldsymbol{x})\rangle^{\mathrm{c}}_{\mathcal{P}(s|d)} &= \langle s(\boldsymbol{x})\rangle_{\mathcal{P}(s|d)}\,,\\
\langle s(\boldsymbol{x})\,s(\boldsymbol{y})\rangle^{\mathrm{c}}_{\mathcal{P}(s|d)} &= \langle s(\boldsymbol{x})\,s(\boldsymbol{y})\rangle_{\mathcal{P}(s|d)} - \langle s(\boldsymbol{x})\rangle_{\mathcal{P}(s|d)}\,\langle s(\boldsymbol{y})\rangle_{\mathcal{P}(s|d)}\,,
\end{aligned} \tag{2.12}$$

where the last term represents such a correction. For Gaussian random fields all higher order connected correlators vanish:

$$\langle s(\boldsymbol{x}_1)\cdots s(\boldsymbol{x}_n)\rangle^{\mathrm{c}}_{\mathcal{P}(s|d)} = 0 \tag{2.13}$$

for $n > 2$, but are in general non-zero for non-Gaussian random fields.

The assumption that the Hamiltonian can be Taylor-Fréchet expanded in the signal field permits us to write

$$H[s] = \frac{1}{2}\, s^\dagger D^{-1}\, s - j^\dagger s + H_0 + \sum_{n=3}^{\infty} \frac{1}{n!}\, \Lambda^{(n)}_{\boldsymbol{x}_1 \ldots \boldsymbol{x}_n}\, s_{\boldsymbol{x}_1} \cdots s_{\boldsymbol{x}_n}\,, \tag{2.14}$$

where repeated coordinates are thought to be integrated over. The first three Taylor coefficients have special roles. The constant H_0 (not to be confused with the Hubble parameter introduced in chapter 1) is fixed by the normalisation condition of the joint probability density of signal and data. If $H'_d[s]$ denotes some unnormalised Hamiltonian, its normalisation constant is given by

$$H_0 = \log \int \mathcal{D}s \int \mathcal{D}d\; e^{-H'_d[s]}\,. \tag{2.15}$$

Often H_0 is irrelevant unless different models or hyperparameters are to be compared.

We call the linear coefficient j information source. This term is usually directly and linearly related to the data. The quadratic coefficient, D^{-1}, defines the information propagator $D(\boldsymbol{x}, \boldsymbol{y})$, which propagates information about the signal at $\boldsymbol{y}$ to location $\boldsymbol{x}$, and thereby permits us, e.g., to partially reconstruct the signal at locations where no data was taken. Finally, the anharmonic tensors $\Lambda^{(n)}$ create interactions between the modes of the free, harmonic theory. Since this free theory will be the basis for the full interaction theory, we first investigate the case $\Lambda^{(n)} = 0$.

2.3.1 Free theory and the Wiener filter

A very simple and widely used data model specifying the relation between signal and data is the model

$$d = Rs + n\,, \tag{2.16}$$

where the data are given by a linear *response matrix* R applied to the signal plus an additive noise term n. For precise definitions of R and n, the reader is referred to Enßlin et al. (2009). For the free theory, we assume both the signal prior and the noise distribution to be Gaussian, i.e. $\mathcal{P}(s) = \mathcal{G}(s, S)$ and $\mathcal{P}(n) = \mathcal{G}(n, N)$, with the signal and noise covariances $S \equiv \langle s s^\dagger \rangle_{\mathcal{P}(s)}$ and $N \equiv \langle n n^\dagger \rangle_{\mathcal{P}(n)}$.

Since the noise is just the difference of the data to the signal-response, $n = d - R\,s$, the likelihood is given by

$$\mathcal{P}(d\,|\,s) = \mathcal{P}(n = d - R\,s\,|\,s) = \mathcal{G}(d - R\,s, N)\,, \tag{2.17}$$

and thus the joint distribution of signal and data for the Gaussian theory is

$$\begin{aligned} \mathcal{P}(s, d) &= [\mathcal{P}(d\,|\,s)\mathcal{P}(s)] \\ &= [\mathcal{G}(d - R\,s, N)\,\mathcal{G}(s, S)]\,. \end{aligned} \tag{2.18}$$

This leads to the Hamiltonian

$$\begin{aligned} H_{\mathcal{G}}[s] &= -\log[\mathcal{P}(d\,|\,s)\mathcal{P}(s)] \\ &= \frac{1}{2} s^\dagger D^{-1} s - j^\dagger s + H_0^{\mathcal{G}}\,, \end{aligned} \tag{2.19}$$

with

$$D = \left[S^{-1} + R^\dagger N^{-1} R\right]^{-1} \tag{2.20}$$

being the propagator of the free theory and

$$j = R^\dagger N^{-1} d \tag{2.21}$$

the information source. Finally,

$$H_0^{\mathcal{G}} = \frac{1}{2}\, d^\dagger\, N^{-1}\, d + \frac{1}{2}\, \log\left(|2\,\pi\, S\,|\, |2\,\pi\, N|\right) \tag{2.22}$$

has absorbed all s-independent normalisation constants. We prove eq. (2.19) in Appendix A.1.

The partition function of the free field theory,

$$\begin{aligned} Z_{\mathcal{G}}[J] &= \int \mathcal{D}s\, e^{-H_{\mathcal{G}}[s]+J^\dagger s} \\ &= \int \mathcal{D}s\, \exp\left\{-\frac{1}{2} s^\dagger D^{-1} s + (J+j)^\dagger s \,-H_0^{\mathcal{G}}\right\}\,, \end{aligned} \tag{2.23}$$

is a Gaussian path integral, which can be calculated exactly, yielding

$$Z_{\mathcal{G}}[J] = \sqrt{|2\pi\, D|}\, \exp\left\{+\frac{1}{2}(J+j)^\dagger D (J+j) - H_0^{\mathcal{G}}\right\}\,. \tag{2.24}$$

The explicit partition function permits us to calculate via eq. (2.11) the posterior mean of the signal, i.e. the expectation of the signal given the data:

$$\begin{aligned} s_{\mathrm{rec}} &= \langle s\rangle_{\mathcal{P}(s|d)} = \left.\frac{\delta \log Z_G}{\delta J}\right|_{J=0} = D\, j \\ &= \underbrace{\left[S^{-1} + R^\dagger N^{-1} R\right]^{-1} R^\dagger N^{-1}}_{F_{\mathrm{WF}}}\, d\,. \end{aligned} \tag{2.25}$$

The last expression shows that the posterior mean is given by the data after applying a generalised Wiener filter, $s_{\mathrm{rec}} = F_{\mathrm{WF}}\, d$, which has first been derived by Wiener (1950). We call s_{rec} the *Wiener reconstruction* of the signal, hence the index 'rec'. The propagator $D(\boldsymbol{x}, \boldsymbol{y})$ describes how the information on the density field contained in the data at location $\boldsymbol{x}$ propagates to position $\boldsymbol{y}$: $s_{\mathrm{rec}}(\boldsymbol{y}) = \int d^d x\, D(\boldsymbol{y}, \boldsymbol{x})\, j(\boldsymbol{x})$.

In Appendix A.1, we explicitely derive the signal posterior, which is

$$\mathcal{P}(s\,|\,d) = \mathcal{G}\,(s - s_{\mathrm{rec}}, D)\ . \tag{2.26}$$

From eq. (2.26), we readily see that the real signal s virtually fluctuates around the reconstruction s_{rec} with the covariance D, due to remaining uncertainties. The propagator is thus also called the *Wiener variance*. The Wiener reconstruction is often used to reconstruct the matter distribution in the Universe from galaxy catalogues, and to obtain maps of the CMB fluctuations from time-ordered data.

2.3.2 Interacting information fields

In the previous section, we have introduced the free theory, corresponding to a Gaussian posterior distribution. A non-Gaussian signal or noise, a non-linear response, or a signal dependent noise create anharmonic terms in the Hamiltonian. These describe interactions between the eigenmodes of the free Hamiltonian.

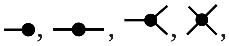
We assume the Hamiltonian can be Taylor expanded in the signal fields, which permits us to write

$$H[s] = \underbrace{\frac{1}{2}\, s^\dagger D^{-1}\, s - j^\dagger s + H_0^{\mathcal{G}}}_{H_{\mathcal{G}}[s]} + \underbrace{\sum_{n=0}^{\infty} \frac{1}{n!}\, \Lambda^{(n)}_{\mathbf{x}_1\ldots\mathbf{x}_n}\, s_{\mathbf{x}_1} \cdots s_{\mathbf{x}_n}}_{H_{\mathrm{int}}[s]} \,. \tag{2.27}$$

Repeated coordinates are thought to be integrated over. In contrast to eq. (2.14), we have now included perturbations which are constant, linear and quadratic in the signal field, because we are summing from $n = 0$. This permits us to treat certain non-ideal effects perturbatively. For example, if a mostly position-independent propagator gets a small position dependent contamination, it might be more convenient to treat the latter perturbatively and not to include it into the propagator used in the calculation. Note further, that all coefficients can be assumed to be symmetric with respect to their coordinate-indices.[1]

Since all the information about any correlation function of the fields is contained in the partition sum and can be extracted from it, only the latter needs to be calculated:

$$\begin{aligned}
Z[J] &= \int \mathcal{D}s\, e^{-H[s]+J^\dagger s} \\
&= \int \mathcal{D}s\, \exp\left[-\sum_{n=0}^{\infty} \frac{1}{n!}\, \Lambda^{(n)}_{\mathbf{x}_1\ldots\mathbf{x}_n}\, s_{\mathbf{x}_1} \cdots s_{\mathbf{x}_n}\right] e^{-H_{\mathcal{G}}[s]+J^\dagger s} \\
&= \exp\left[-\sum_{n=0}^{\infty} \frac{1}{n!}\, \Lambda^{(n)}_{\mathbf{x}_1\ldots\mathbf{x}_n}\, \frac{\delta}{\delta J_{\mathbf{x}_1}} \cdots \frac{\delta}{\delta J_{\mathbf{x}_n}}\right] \int \mathcal{D}s\, e^{-H_{\mathcal{G}}[s]+J^\dagger s} \\
&= \exp\left[-H_{\mathrm{int}}[\frac{\delta}{\delta J}]\right] Z_{\mathcal{G}}[J] \,.
\end{aligned} \tag{2.28}$$

There exist well-known diagrammatic expansion techniques for such expressions (e.g. Binney et al. 1992). The expansion terms of the logarithm of the partition sum, from which any connected moments can be calculated, are represented by all possible connected diagrams build out of lines (——), vertices (with a number of legs connecting to lines, like , ...) and without any external line-ends (any line ends in a vertex). These diagrams are interpreted according to the following Feynman rules:

1. Open ends of lines in diagrams correspond to external coordinates and are labelled by such. Since the partition sum in particular does not depend on any external coordinate, it

---

[1] This means $D_{\mathbf{x}\,\mathbf{y}} = D_{\mathbf{y}\,\mathbf{x}}$ and $\Lambda^{(n)}_{\mathbf{x}_{\pi(1)}\ldots\mathbf{x}_{\pi(n)}} = \Lambda^{(n)}_{\mathbf{x}_1\ldots\mathbf{x}_n}$ with π any permutation of $\{1, \ldots, n\}$, since even non-symmetric coefficients would automatically be symmetrized by the integration over all repeated coordinates. Therefore, we assume in the following that such a symmetrization operation has been already done, or we impose it by hand before we continue with any perturbative calculation by applying

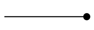
$$\Lambda^{(n)}_{\mathbf{x}_1\ldots\mathbf{x}_n} \longmapsto \frac{1}{n!} \sum_{\pi\in\mathcal{P}_n} \Lambda^{(n)}_{\mathbf{x}_{\pi(1)}\ldots\mathbf{x}_{\pi(n)}} \,.$$

This clearly leaves any symmetric tensor invariant if $\mathcal{P}_n$ is the space of all permutations of $\{1, \ldots, n\}$.

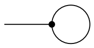
is calculated only from summing up closed diagrams. However, the field expectation value $s_{\mathrm{rec}}(\boldsymbol{x}) \equiv \langle s(\boldsymbol{x})\rangle_{\mathcal{P}(s|d)} = d\log Z[J]/dJ(\boldsymbol{x})|_{J=0}$ and higher order correlation functions depend on coordinates and therefore are calculated from diagrams with one or more open ends, respectively.

2. A line with coordinates $\boldsymbol{x}'$ and $\boldsymbol{y}'$ at its end represents the propagator $D_{\boldsymbol{x}'\boldsymbol{y}'}$ connecting these locations.
3. Vertices with one leg get an individual internal, integrated coordinate $\boldsymbol{x}'$ and represent the term $j_{\boldsymbol{x}'} + J_{\boldsymbol{x}'} - \Lambda^{(1)}_{\boldsymbol{x}'}$.
4. Vertices with n legs represent the term $-\Lambda^{(n)}_{\boldsymbol{x}'_1 \dots \boldsymbol{x}'_n}$, where each individual leg is labelled by one of the internal coordinates $\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_n$. This more complex vertex-structure, as compared to QFT, is a consequence of non-locality in IFT.
5. All internal (and therefore repeatedly occurring) coordinates are integrated over, whereas external coordinates are not.
6. Every diagram is divided by its symmetry factor, the number of permutations of vertex legs leaving the topology invariant, as described in any book on field theory (e.g. Binney et al. 1992).

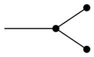
The n-th moment of s is generated by taking the n-th derivative of $\log Z[J]$ with respect to J, and then setting $J = 0$. This correspond to removing n end-vertices from all diagrams. For example, the first four diagrams contributing to a signal reconstruction ($s_{\mathrm{rec}} \equiv \langle s\rangle_{\mathcal{P}(s|d)}$) are



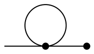
$$= D_{\boldsymbol{xy}}\, j_{\boldsymbol{y}}$$



$$= -\frac{1}{2} D_{\boldsymbol{xy}}\, \Lambda^{(3)}_{\boldsymbol{yzu}}\, D_{\boldsymbol{zu}}$$



$$= -\frac{1}{2} D_{\boldsymbol{xy}}\, \Lambda^{(3)}_{\boldsymbol{yuz}}\, D_{\boldsymbol{zz}'}\, j_{\boldsymbol{z}'}\, D_{\boldsymbol{uu}'}\, j_{\boldsymbol{u}'}$$



$$= -\frac{1}{2} D_{\boldsymbol{xy}}\, \Lambda^{(4)}_{\boldsymbol{yzuv}}\, D_{\boldsymbol{zu}}\, D_{\boldsymbol{vv}'}\, j_{\boldsymbol{v}'}\,. \tag{2.29}$$

Here, we have assumed that any first and second order perturbation was absorbed into the data source and the propagator, thus $\Lambda^{(1)} = \Lambda^{(2)} = 0$. Repeated indices are assumed to be integrated (or summed) over. The Feynman diagrams are to be interpreted as computational algorithms and can be implemented using existing map-making codes or linear algebra packages for the information propagator and vertices.

In Enßlin et al. (2009), we have shown the utility of IFT on two examples: The derivation of a Bayesian estimator for the non-Gaussianity in the CMB, and the non-linear reconstruction of a Gaussian signal with Poissonian noise in one dimension. The latter serves as a one-dimensional toy model for the LSS distribution measured from galaxy counts.

2.4 Parameter sampling

In section 2.3, we have assumed that we have an analytic form of the posterior distribution $\mathcal{P}(s\,|\,d)$, or at least an analytical approximation to the latter given by the Taylor expansion of the

Hamiltonian. However, this is not always the case, since sometimes we can only compute the likelihood numerically. We cannot use IFT to infer information about our signal in these cases. In this section, we explain cosmological parameter estimation as an example of such a problem, and introduce the concept of *parameter sampling*, which we can use to address it.

The cosmological data we observe are created by complicated physical processes, which leave characteristic imprints on the data. The corresponding physical processes can be theoretically described by an adequate model, which ususally depends on several cosmological parameters p. By comparing the predictions of such a model with observations, we can thus infer information about the cosmological parameters from the data. However, usually we have to constrain several parameters simultaneously, since their effects on the data are often degenerate. The considered cosmological parameter spaces typically have between 6 and 12 dimensions.

In order to obtain constraints on our cosmological parameters, we use Bayes' Theorem, eq. (2.4), with the cosmological parameters p being the signal:

$$\mathcal{P}(p\,|\,d) = \frac{\mathcal{P}(d\,|\,p)\,\mathcal{P}(p)}{\mathcal{P}(d)}\,. \tag{2.30}$$

As already mentioned, the likelihood $\mathcal{P}(d\,|\,p)$ often does not have an analytic form, but needs to be evaluated numerically. In the case of the CMB, for instance, evaluating the likelihood implies running a Boltzmann code such as CMBFAST (Seljak & Zaldarriaga 1996), CAMB (Lewis et al. 2000), or CMBEASY (Doran 2005) to calculate the power spectrum, which is then fed into a likelihood code. Since we lack an analytic form of the likelihood and thus of the posterior, we need a representation of the latter in parameter space, from which we can compute the posterior mean values and variances of the cosmological parameters. The usual way of doing so is to run Markov Chain Monte Carlo simulations, which we will describe briefly in section 2.4.1.

2.4.1 Markov Chains and the Metropolis Algorithm

Markov Chain Monte Carlo (MCMC) simulations (Chib & Greenberg 1995; Gamerman 1997; Neal 1993) are used to draw samples from a probability distribution. The statistical properties of the distribution, such as its mean and variance can then be estimated from the sample. Usually, one wishes to sample the posterior distribution of the (cosmological) parameters, $\mathcal{P}(p\,|\,d)$, but the technique can equally well be used to sample any other probability distribution. MCMCs are especially well-suited for high-dimensional problems, because the computational effort increases only linearly with the number of parameters.

The samples are drawn by running a Markov Chain, which is defined as a sequence of random variables (in our case points in parameter space) chosen by a random process such that a given element of the sequence, p_i, depends solely on the previous element, p_{i-1}. The aim is to choose the next point in the chain based on the previous point such that the distribution of the points becomes stationary, with $\mathcal{P}(p\,|\,d)$ being the stationary distribution, in the limit of the number of points going to infinity.

One possibility of implementing such a process is the Metropolis-Hastings (M-H) algorithm (Metropolis et al. 1953), which we briefly introduce in the following. For a given point p_i in the chain, the M-H algorithm draws a point $\tilde{p}$ from a proposal distribution $q(\tilde{p}\,|\,p_i)$. The proposed point is accepted, i.e. $p_{i+1} \equiv \tilde{p}$, with the transition probability

$$a(p_i, \tilde{p}) = \min\left\{\frac{q(p_i \,|\, \tilde{p})\, \mathcal{P}(\tilde{p} \,|\, d)}{q(\tilde{p} \,|\, p_i)\, \mathcal{P}(p_i \,|\, d)}, 1\right\} . \tag{2.31}$$

In practise, this is implemented by drawing a uniformly distributed random variable u from [0,1], accepting the proposed point if

$$\frac{q(p_i \,|\, \tilde{p})\, \mathcal{P}(\tilde{p} \,|\, d)}{q(\tilde{p} \,|\, p_i)\, \mathcal{P}(p_i \,|\, d)} > u \,, \tag{2.32}$$

and rejecting it otherwise. If $\tilde{p}$ is rejected, we retain the old point and set $p_{i+1} \equiv p_i$. If the proposal distribution is symmetric, $q(\tilde{p} \,|\, p_i) = q(p_i \,|\, \tilde{p})$, the algorithm is called the Metropolis algorithm (Metropolis et al. 1953). The Metropolis algorithm is used in the MCMC driver of CMBEASY (Doran & Müller 2004).

Chapter 3

Optimal methods for detecting the integrated Sachs-Wolfe effect

Note: Sections 3.2-3.4 and section 3.7 of this chapter, as well as appendix B.1 and A.1, have been published in Frommert et al. (2008). The bulk of section 3.5 and appendix B.2 have been published in Frommert & Enßlin (2009a). Section 3.6 has been added.

3.1 Introduction

As we have seen in section 1.4.3, the integrated Sachs-Wolfe effect (Sachs & Wolfe 1967) is an important probe of the existence and nature of dark energy (see also Crittenden & Turok 1996) and the nature of gravity (see Lue et al. 2004; Zhang 2006b). However, the detection of the ISW signal is a challenging task, for it is much smaller than the primordial temperature fluctuations in the CMB, which originate at the time of last scattering. We can try to detect the ISW effect via its cross-correlation with the large-scale structure (LSS). Such a correlation exists, since the ISW effect is created by the interaction of CMB photons with the gravitational potential of the LSS. The primordial temperature fluctuations of the CMB, on the other hand, should be uncorrelated with the LSS distribution. In recent years, substantial effort has been made to detect the ISW effect via cross-correlation of the CMB temperature fluctuations with LSS surveys, such as optical galaxy and quasar surveys[1], radio surveys[2], and X-ray surveys[3].[4]

The *standard method* for detecting the cross-correlation between the LSS and the CMB, which has been used by the studies mentioned above, involves comparing the observed cross-correlation function with its theoretical prediction for a given fiducial cosmological model. The theoretical prediction is by construction an ensemble average over all possible realisations of the universe given the fiducial parameters, i.e. over all possible realisations of the primordial CMB, which originates at the surface of last scattering, and all realisations of the local matter distribution. Assuming ergodicity, this second ensemble average can also be thought of as an average over all possible positions of the observer in the Universe ('cosmic mean'). The specific realisations of both the LSS and the primordial temperature fluctuations of the CMB in our Universe thus contribute to the error budget of the detection. We estimate the contribution of these two sources of uncertainty to the total variance in the detected signal under the simplifying assumption that there is no shot noise in the galaxy distribution. The contribution of the LSS to the total uncertainty, which we refer to as *local variance*, amounts to about 11 per cent in the case of an ideal LSS survey going out to about redshift 2 and covering enough volume to include the large scales relevant for the ISW. We will show that this local variance leads to a biased detection significance in the standard method for ISW detection.

---

[1] Sloan Digital Sky Survey, Adelman-McCarthy et al. (2008), Two-Micron All-Sky Survey, Jarrett et al. (2000)

[2] NRAO VLA Sky Survey, Condon et al. (1998)

[3] High Energy Astrophysics Observatory, Boldt (1987)

[4] Such cross-correlation studies have, for example, been done by Boughn et al. (1998), Boughn & Crittenden (2004), Boughn & Crittenden (2005), Afshordi et al. (2004), Rassat et al. (2007), Raccanelli et al. (2008), McEwen et al. (2007), Pietrobon et al. (2006), Fosalba et al. (2003), Fosalba & Gaztañaga (2004), Vielva et al. (2006), Liu & Zhang (2006), Ho et al. (2008) and Giannantonio et al. (2008), just to name a few of them.

In this chapter, we present new methods for the detection of the ISW effect, which reduce both sources of uncertainty mentioned above by working conditional on the LSS distribution and on the measured CMB polarization. The method which only operates conditional on the LSS distribution, without using polarization data, will be referred to as the *optimal temperature method*. The conditionality on the LSS implies that the signal-to-noise ratio or detection significance in the optimal temperature method depends on the specific realisation of the LSS in our Universe. Note that we use the two expressions *signal-to-noise ratio* and *detection significance* as synonyms. On average, the detection significance is about 7 per cent higher than for the standard method, due to the reduction of local variance. Here, we have assumed a perfect galaxy survey covering all of the relevant volume. In addition to reducing local variance, we can reduce the variance coming from the primordial temperature fluctuations of the CMB by inferring information about the latter from CMB polarization data. The resulting method is called the *optimal polarization method*. Note that, of course, the optimal polarization method uses not only polarization data but also temperature data. The latter reaches a detection significance of up to 8.5, which is about 16 per cent higher than the standard one for shallow LSS surveys such as the SDSS main galaxy sample, and about 23 per cent for a full-sky survey reaching out to a redshift of 2. Again, these estimates hold for ideal (noiseless) data. Unfortunately, for currently available CMB and LSS surveys, the detection significance of our optimal polarization method is not notably above the standard one, which is mainly due to the high contamination of the WMAP polarization data by detector noise and Galactic foregrounds. A very crude estimate for data from the *Planck Surveyor* mission promises an enhancement of detection significance of at least 10 per cent for the optimal polarization method as compared to the standard method.

Many of the cross-correlation studies mentioned above have attempted to constrain cosmological parameters using a likelihood function for the cosmological parameters p given the observed cross-correlation function between CMB temperature fluctuations and LSS data. Just like the detection significance, these parameter estimates suffer from biasing due to local variance. Furthermore, to our knowledge, there is no straightforward way of combining the likelihood function for the cross-correlation with the likelihoods for CMB and LSS data so far. In this chapter, we derive the correct joint likelihood function $\mathcal{P}(T, P, \delta_g \,|\, p)$ for cosmological parameters, given the CMB temperature and polarization maps T and P and the LSS data δ_g, from first principles for the linear LSS formation regime. This joint likelihood consistently includes the coupling between the two data-sets introduced by the ISW effect, which so far has been neglected in analyses deriving cosmological parameter constraints by combining CMB and LSS data (Tegmark et al. 2004; Spergel & et al 2007). For parameter sampling studies using our likelihood, we expect small changes of the dark-energy related parameters with respect to studies neglecting the coupling between the data-sets.

This chapter is organised as follows. We start by explaining the different stochastic processes that are relevant for the measurement of the ISW effect in section 3.2. In section 3.3, we review the standard method for detecting the ISW effect via cross-correlation and estimate the contribution of the local variance to the total variance of the detected signal. Section 3.4 is devoted to presenting the optimal temperature method of ISW detection we developed, and to comparing it to the standard method, whereas in section 3.5, we present the optimal polarization method and compare it to the other two methods. In section 3.6, we estimate the improvement we obtain from the optimal polarization method for currently available data. We discuss the role of the biasing effect due to local variance in parameter constraints and derive the joint likelihood function $\mathcal{P}(T, P, \delta_g \,|\, p)$ in section 3.7. Concluding remarks are given in section 3.8.

3.2 Stochastic processes

In order to understand the methods for ISW detection introduced in this chapter, it is necessary to be familiar with the different stochastic processes that need to be considered. We introduce those processes in the following.

3.2.1 Realisation of the matter distribution

During inflation, the matter density perturbations have been created from quantum fluctuations of the inflaton field. This stochastic process is believed to have been close to Gaussian (Mukhanov 2005), permitting to write down the probability distribution for the matter density contrast given the cosmological parameters p as

$$\mathcal{P}(\delta_m \,|\, p) = \mathcal{G}(\delta_m, S_m)\,, \tag{3.1}$$

where the covariance matrix $S_m \equiv \langle \delta_m \delta_m^\dagger \rangle_{\mathcal{P}(\delta_m \,|\, p)}$, depends on the cosmological parameters p. The average $\langle .. \rangle_{\mathcal{P}(\delta_m \,|\, p)}$ is defined as ensemble average over the different realisations of δ_m, the index $\mathcal{P}(\delta_m \,|\, p)$ explicitly states which probability distribution the average has to be taken over. Given homogeneity and isotropy, we note that the Fourier transformation of S is diagonal:

$$\langle \delta_m(\boldsymbol{k}) \delta_m(\boldsymbol{k}')^* \rangle_{\mathcal{P}(\delta_m \,|\, p)} = (2\pi)^3 \delta(\boldsymbol{k} - \boldsymbol{k}') P(k)\,, \tag{3.2}$$

where $P(k)$ is the power spectrum, $\delta(..)$ denotes the Dirac delta function, and the star is used for denoting complex conjugation.

The stochastic process due to the inflationary quantum fluctuations created the angular fluctuations in the CMB, that is, the primordial temperature fluctuations originating from the surface of last scattering at redshift $z = 1100$, as well as the integrated Sachs-Wolfe effect imprinted by the more local matter distribution at $z < 2$. Throughout this work we will assume that the primordial fluctuations and the ISW effect are stochastically independent, which is a safe assumption, given that they are associated with matter perturbations of very different wavelengths that are spatially well separated, so that very little intrinsic cross-correlation can be expected (Boughn et al. 1998). In fact, for notational convenience we will use the symbol δ_m to only denote the local matter distribution at $z < 2$. The joint probability distribution for $T_{\mathrm{isw}} = \mathcal{Q}\,\delta_m$ (cf. eq. (1.37)) and the primordial temperature fluctuations T_{prim} then factorises

$$\mathcal{P}(T_{\mathrm{isw}}, T_{\mathrm{prim}} \,|\, p) = \mathcal{P}(T_{\mathrm{isw}} \,|\, p)\, \mathcal{P}(T_{\mathrm{prim}} \,|\, p)\,, \tag{3.3}$$

with

$$\mathcal{P}(T_{\rm isw}\,|\,p) = \mathcal{G}(T_{\rm isw}, C_{\rm isw})\,, \tag{3.4}$$

and

$$\mathcal{P}(T_{\rm prim}\,|\,p) = \mathcal{G}(T_{\rm prim}, C_{\rm prim})\,, \tag{3.5}$$

where we have defined the angular two-point auto-correlation function for the fluctuation T_X (X being 'isw' or 'prim')

$$C_X \equiv \langle T_X T_X^\dagger \rangle_{\mathcal{P}(T_X\,|\,p)}\,. \tag{3.6}$$

Again, given homogeneity and isotropy, C_X is diagonal in spherical harmonics space

$$\langle a^X_{lm} a^{X\,*}_{l'm'} \rangle_{\mathcal{P}(T_X\,|\,p)} = C^X_l \delta_{ll'}\, \delta_{mm'}\,, \tag{3.7}$$

where C^X_l is the angular power spectrum of the quantity X, and we have used the expansion coefficients of T_X into spherical harmonics Y_{lm},

$$a^X_{lm} \equiv \int_S d\Omega\, T_X(\hat{\boldsymbol{n}})\, Y^*_{lm}(\hat{\boldsymbol{n}})\,, \tag{3.8}$$

where the integral is taken over the sphere. Given that the joint distribution $\mathcal{P}(T_{\rm isw}, T_{\rm prim}\,|\,p)$ factorises into two Gaussian distributions, the sum $T = T_{\rm isw} + T_{\rm prim}$, which denotes the temperature fluctuation of the CMB, is again Gaussian distributed

$$\mathcal{P}(T\,|\,p) = \mathcal{G}(T, C_T)\,, \tag{3.9}$$

with

$$C_T = C_{\rm isw} + C_{\rm prim}\,. \tag{3.10}$$

Given the cosmological parameters, the angular power spectra C^T_l, $C^{\rm isw}_l$, and $C^{\rm prim}_l$ can all be calculated using CMBFAST (`http://ascl.net/cmbfast.html`, Seljak & Zaldarriaga (1996)), CAMB (`http://camb.info`, Lewis et al. (2000)), or CMBEASY (`www.cmbeasy.org`, Doran (2005)). In particular, $C_{\rm isw}$ can be obtained from the three-dimensional matter covariance matrix S_m by

$$C_{\rm isw} = Q\, S_m\, Q^\dagger\,, \tag{3.11}$$

where we have used that linear transformations of Gaussian random variables are again Gaussian distributed, with the covariance matrix transformed accordingly (see also Cooray 2002a).

3.2.2 CMB detector noise

From CMB detectors, we do not read off the actual CMB temperature fluctuations T as defined in the last section, but a temperature where the detector noise $T_{\rm det}$ has been added. Again, this can be modelled as a Gaussian random process,

$$\mathcal{P}(T_{\rm det}) = \mathcal{G}(T_{\rm det}, C_{\rm det})\,, \tag{3.12}$$

where $C_{\rm det}$ denotes the detector noise covariance. This process is independent of the process that created the CMB fluctuations T, such that if we redefine $T \equiv T + T_{\rm det}$ to be the temperature we read off our detector, we obtain

$$\mathcal{P}(T \mid p) = \mathcal{G}(T, C_T)\,, \tag{3.13}$$

where $C_T \equiv C_{\rm prim} + C_{\rm isw} + C_{\rm det}$ now includes the covariance of the detector temperature.

However, in most of this work we will neglect the detector noise in the CMB temperature ($C_{\rm det} \approx 0$), since the ISW is only present on the largest angular scales, where the dominant source of noise is cosmic variance (Afshordi 2004). However, if needed, the detector noise can be easily included by just substituting $C_{\rm prim} \rightarrow C_{\rm prim} + C_{\rm det}$. The only part where we explicitely include the temperature detector noise will be in section 3.7, where we derive the joint likelihood for the cosmological parameters, given CMB and LSS data, since in this likelihood we also include smaller angular scales.

3.2.3 Shot noise

Unfortunately, the matter distribution is not directly known, and we have to rely on LSS catalogues from which we can try to reconstruct it. A process to be considered when working with such catalogues is the stochastic distribution of the galaxies, which only on average follows the matter distribution. Since the galaxies are discrete sources from which we want to infer the properties of the underlying matter overdensity field, we have to deal with shot noise in the galaxy distribution. More specifically, we assume the observed number $N_g(\boldsymbol{x_i})$ of galaxies in a volume element $\Delta V(\boldsymbol{x_i})$ at a discrete position $\boldsymbol{x_i}$ to be distributed according to a Poisson distribution

$$\mathcal{P}(N_g(\boldsymbol{x_i}) \mid \lambda(\boldsymbol{x_i})) = \frac{\lambda(\boldsymbol{x_i})^{N_g(\boldsymbol{x_i})} e^{-\lambda(\boldsymbol{x_i})}}{N_g(\boldsymbol{x_i})!}\,. \tag{3.14}$$

Here, $\lambda(\boldsymbol{x})$ denotes the expected mean number of observed galaxies within $\Delta V(\boldsymbol{x})$, given the matter density contrast,

$$\lambda(\boldsymbol{x}) = w(\boldsymbol{x})\,\overline{n_g^r}\,\Delta V\,\left[1 + b\,\delta_m(\boldsymbol{x})\right]\,. \tag{3.15}$$

In the above equation, $\overline{n_g^r} \equiv {N_g}^{r,\,\rm tot}/V$ denotes the cosmic mean galaxy density, with ${N_g}^{r,\,\rm tot}$ being the total number of galaxies in the volume V. Note that we have added an index 'r' to stress that these are the actual (real) number of galaxies present in ΔV, not the observed number of galaxies N_g, which can be smaller due to observational detection limits. The window $w(\boldsymbol{x}) \equiv \Phi(\boldsymbol{x})\,m(\boldsymbol{\hat{n}})$ denotes the combined selection function $\Phi(\boldsymbol{x})$ and sky mask $m(\boldsymbol{\hat{n}})$ of the survey, and b the galaxy bias, which in general depends on redshift, scale, and galaxy type. The variance in the observed number of galaxies $N_g(\boldsymbol{x})$ within $\Delta V(\boldsymbol{x})$ is then $\sigma_g^2(\boldsymbol{x}) \equiv \langle \left(N_g(\boldsymbol{x}) - \lambda(\boldsymbol{x})\right)^2 \rangle_{N_g} = \lambda(\boldsymbol{x})$. Here, we have used the index N_g on the average to indicate the average over the Poisson distribution in eq. (3.14).

If the average number of galaxies $\lambda(\boldsymbol{x})$ is large, the Poisson distribution is well approximated by a Gaussian distribution around $\lambda(\boldsymbol{x})$. For simplicity we will use the Gaussian approximation throughout this work. Furthermore we will ignore the dependence of the noise on $\delta_m(\boldsymbol{x})$ by using $\sigma_g^2(\boldsymbol{x}) = w(\boldsymbol{x})\,\overline{n_g^r}\,\Delta V$ instead of the correct noise term $\sigma_g^2(\boldsymbol{x}) = \lambda(\boldsymbol{x})$, for the latter would require a non-linear and iterative approach. Such an approach is beyond the scope of this paper, but is also irrelevant for the main finding of this work. However, see Enßlin et al. (2009) for a better handling of the Poisson noise and bias variations.

Since the cosmic mean galaxy density $\overline{n_g^r}$ is not known, we have to estimate it from the observed galaxy counts by

$$\overline{n_g^r}\,\Delta V \equiv \frac{N_g^{\mathrm{tot}}}{\sum_{i=0}^{N_{\mathrm{pix}}} w(\boldsymbol{x_i})}\,, \tag{3.16}$$

where N_g^{tot} is the total number of observed galaxies and the sum goes over all the pixels in our volume.

With the above-mentioned simplifications, we can now work with the following linear data model. First we define the observed galaxy density contrast at position $\boldsymbol{x}$ to be

$$\delta_g(\boldsymbol{x}) \equiv \frac{N_g(\boldsymbol{x}) - w(\boldsymbol{x})\,\overline{n_g^r}\,\Delta V}{\overline{n_g^r}\,\Delta V}\,, \tag{3.17}$$

which is the convention used in Kitaura et al. (2009). Note that this definition differs from the one usually used in cross-correlation studies by a factor of $w(\boldsymbol{x})$ (see, e.g., Pogosian et al. 2005). We then write

$$\delta_g = R\,\delta_m + \epsilon\,, \tag{3.18}$$

where $\epsilon(\boldsymbol{x})$ is the additive noise-term that originates in the Poissonian distribution of $N_g(\boldsymbol{x})$, and R is the linear response operator. In the simplest case, $R(\boldsymbol{x_i}, \boldsymbol{x_j}) \equiv b\,w(\boldsymbol{x_i})\,\delta_{ij}$, but in general R maps the continuous space in which δ_m lives onto the discrete pixel space of our data δ_g, and it can also include the mapping from redshift-space onto comoving coordinate space. In the latter case, the matter density contrast δ_m would have to be read as a density contrast in redshift space.

Gravitational lensing introduces a magnification bias in the observed galaxy density contrast, as described by Loverde et al. (2007). In our data model, it is straightforward to take this effect into account by letting

$$\begin{aligned} R\,\delta_m(\hat{\boldsymbol{n}}, z) \;\equiv\; & w(\hat{\boldsymbol{n}}, z)[b\,\delta_m(r(z)\,\hat{\boldsymbol{n}}, z) + 3\,\Omega_m\,H_0^2\,(2.5\,s(z) - 1) \\ & \times \int_0^z dz'\,\frac{1}{H(z')}\,\frac{r(z')(r(z) - r(z'))}{r(z)} \\ & (1 + z')\,\delta_m(r(z)\,\hat{\boldsymbol{n}}, z')]\,, \end{aligned} \tag{3.19}$$

where $r(z)$ is the comoving distance corresponding to redshift z, and the slope s of the number count of the source galaxies is defined as

$$s \equiv \frac{d\log_{10} N(< m)}{dm}\,, \tag{3.20}$$

with m being the limiting magnitude and $N(< m)$ being the count of objects brighter than m. Note that in order to get the correct formula for the magnification bias term in 3 dimensions, we used the Dirac delta function as the normalised selection function used by Loverde et al. (2007), $W(z, z') \equiv \delta(z - z')$.

From the Poisson distribution in eq. (3.14), we see that $\langle\delta_g\rangle_{N_g} = R\,\delta_m$, and thus with the above simplifications the noise ϵ is Gaussian distributed around zero

$$\mathcal{P}(\epsilon\,|\,p) = \mathcal{G}(\epsilon, N_\epsilon)\,, \tag{3.21}$$

with the noise covariance matrix

$$N_\epsilon(\boldsymbol{x_i}, \boldsymbol{x_j}) \equiv \langle \epsilon(\boldsymbol{x_i})\, \epsilon(\boldsymbol{x_j}) \rangle_{N_g} = \frac{w(\boldsymbol{x_i})}{\overline{n_g^r}\, \Delta V}\, \delta_{ij}\,. \tag{3.22}$$

3.3 Standard cross-correlation method

In this section, we briefly review the standard method for detecting the cross-correlation of the CMB with the projected galaxy density contrast, which was first described by Boughn et al. (1998), but see for example also Ho et al. (2008) and Giannantonio et al. (2008)). Note that we use the word galaxy density contrast for convenience, but the method is of course the same when working with other tracers of the LSS.

3.3.1 Description

The theoretical cross-correlation function of two quantities $X(\hat{\boldsymbol{n}})$ and $Y(\hat{\boldsymbol{n}})$ on the sky is defined in spherical harmonics space as

$$C_l^{X,Y} \equiv \langle a_{lm}^X a_{lm}^{Y\,*} \rangle_{\mathrm{all}}\,. \tag{3.23}$$

The average in the above definition is an ensemble average over all possible realisations of the universe with given cosmological parameters, i.e. over $\mathcal{P}(\delta_m, \delta_g, T\,|\,p)$. This is indicated by the index ’all’ on the average. We will denote the abstract cross-correlation function as a vector in Hilbert space by $\xi^{X,Y}$ to simplify the notation. This can be understood as a vector in pixel space or as a vector in a_{lm}-space. Only when evaluating the expressions we derive, we will choose the representation of the abstract vector $\xi^{X,Y}$ in spherical harmonics space, $(\xi^{X,Y})_{lml'm'} = C_l^{X,Y} \delta_{ll'} \delta_{mm'}$. In the following we will work with the cross-correlation function of the projected galaxy density contrast with the CMB temperature fluctuations, $\xi^{g,\,T}$, in order to reproduce the standard approach in the literature.

The observed projected galaxy density contrast δ_g^{proj} for a redshift bin centred around redshift z_i in a given direction $\hat{\boldsymbol{n}}$ on the sky is

$$\begin{aligned} \delta_g^{\mathrm{proj}}(\hat{\boldsymbol{n}}, z_i) &= \int dz\, W(z, z_i)\, \delta_g(\hat{\boldsymbol{n}}, z) \\ &= \int dz W(z, z_i)\, [R\, \delta_m(\hat{\boldsymbol{n}}, z) + \epsilon(\hat{\boldsymbol{n}}, z)]\,, \end{aligned} \tag{3.24}$$

where $W(z, z_i)$ denotes the normalised selection function that defines the ith bin, and δ_g is given by eq. (3.17). Note that in many cross-correlation studies the normalised selection function $\Phi(\boldsymbol{x})$ of the survey is used to define the bin. However, since later on we will consider a perfect galaxy survey covering all the redshift range relevant for the ISW, we need to introduce the additional narrow selection function $W(z, z_i)$ defining the bin.

If the LSS survey and the CMB map cover the full sky, it is convenient to define an estimator for the cross-correlation function of the projected galaxy density contrast with the CMB in spherical harmonics space (Rassat et al. 2007),

$$\widehat{C}_l^{g,T} \equiv \frac{1}{2l+1} \sum_m Re\left(a_{lm}^g a_{lm}^{T\,*}\right)\,, \tag{3.25}$$

where a^g_{lm} and a^T_{lm} are the expansion coefficients of the observed δ_g^{proj} and T into spherical harmonics as defined in eq. (3.8). The hat has been added to discriminate the estimator of the cross-correlation function from its theoretical counterpart $C_l^{g,T}$. In the case that the experiments cover only a part of the sky, one has to take into account the effects of mode-coupling when working in spherical harmonics space. In this case it is therefore more straightforward to define other estimators for the cross-correlation function, such as averages over the sphere in real space (see, e.g., Giannantonio et al. 2008) or quadratic estimators as in Afshordi et al. (2004). However, for the statement we will make in this work the actual definition of the estimator is not relevant, and we find the one defined in spherical harmonics space the most convenient to work with, since a closely related quantity also appears within the framework of the optimal detection method presented later on in section 3.4. Again we use the abstract notation $\widehat{\xi^{g,T}}$ for the estimator of the cross-correlation $\xi^{g,T}$. In order to keep the notation simple, we will from now on understand $\widehat{\xi^{g,T}}$ and $\xi^{g,T}$ as being vectors in spherical harmonics-space as well as in bin-space, containing the cross-correlation functions for all the different bins.

In the literature, the probability distribution of the above-defined estimator $\widehat{\xi^{g,T}}$ around the theoretical cross-correlation function $\xi^{g,T}$ is usually approximated by a Gaussian,

$$P\left(\widehat{\xi^{g,T}}\mid p\right) = \mathcal{G}\left(\widehat{\xi^{g,T}} - \xi^{g,T}, C_{cc}\right) , \tag{3.26}$$

where the covariance matrix of the cross-correlation estimator is defined as

$$C_{cc} \equiv \langle \left(\widehat{\xi^{g,T}} - \langle\widehat{\xi^{g,T}}\rangle_{\text{all}}\right)\left(\widehat{\xi^{g,T}} - \langle\widehat{\xi^{g,T}}\rangle_{\text{all}}\right)^{\dagger}\rangle_{\text{all}} . \tag{3.27}$$

The first question usually addressed in the above-mentioned cross-correlation studies is whether a non-zero cross-correlation function can be detected at all. To this end one assumes a fiducial cosmological model, which is used to predict the theoretical cross-correlation function and covariance matrix C_{cc}. In this chapter, we use the flat ΛCDM model with parameter values given by Komatsu et al. (2009), table 1: $\Omega_b h^2 = 0.02265, \Omega_\Lambda = 0.721, h = 0.701, n_s = 0.96, \tau = 0.084, \sigma_8 = 0.817$. The covariance matrix is usually estimated by Monte Carlo simulations (see Cabré et al. (2007) for an overview), or analytically as in Afshordi et al. (2004). The analytical prediction is possible in the case that the joint probability distribution for the projected galaxy density contrast and CMB given the cosmological parameters, $\mathcal{P}(\delta_{gi}^{\text{proj}}, \delta_{gj}^{\text{proj}}, T\mid p)$, is Gaussian, which is valid in the framework of linear perturbation theory. Here we have used the index gi to denote the projected galaxy density contrast of bin i. Then the covariance matrix in spherical harmonics space can be expressed in terms of two-point correlation functions as

$$C_l^{cc}(i,j) = \frac{1}{(2l+1)f_{\text{sky}}}\left[C_l^{gi,T}C_l^{gj,T} + C_l^{gi,gj}\,C_l^T\right] , \tag{3.28}$$

where we have used the auto-correlation power spectrum for the CMB, as defined in eq. (3.7). $C_l^{gi,gj}$ contains by definition the power coming from the underlying matter distribution plus the shot noise. Note that, in principle, C_l^T in the above formula also includes detector noise, which we neglect here as discussed in section 3.2.2. f_{sky} is the fraction of the sky covered by both, the galaxy survey and the CMB experiment. In the following, we will assume $f_{\text{sky}} = 1$ whenever we go to spherical harmonics space.

Putting an amplitude or fudge factor A_{cc} in front of the theoretical cross-correlation function

$\xi^{g,T}$ by hand, one can now find out whether it is possible to detect a non-zero A_{cc}. The index 'cc' on the amplitude indicates that it is the amplitude of the cross-correlation function. Of course this amplitude should be one in the fiducial model. However, even if the data are taken from a universe in which the underlying cosmology is the fiducial model we will in general not estimate the amplitude to be one. This is due to the different sources of stochastic uncertainty or noise in the estimate of A_{cc}, which we have described at length in section 3.2. The likelihood for the amplitude given the cosmological parameters reads

$$P\left(\widehat{\xi}^{g,T} \,|\, A_{cc}, p\right) = G\left(\widehat{\xi}^{g,T} - A_{cc}\,\xi^{g,T}, C_{cc}\right) . \tag{3.29}$$

A commonly used estimator of the amplitude A_{cc} is the maximum likelihood amplitude

$$\begin{aligned} \widehat{A}_{cc} &= \frac{\xi^{g,T\,\dagger}\; C_{cc}^{-1}\; \widehat{\xi}^{g,T}}{\xi^{g,T\,\dagger}\; C_{cc}^{-1}\; \xi^{g,T}} \\ &= \frac{\sum_l (2l+1) \sum_{i,j} C_l^{gi,T} (C_l^{cc})^{-1}(i,j)\, \widehat{C}_l^{gj,T}}{\sum_l (2l+1) \sum_{i,j} C_l^{gi,T} (C_l^{cc})^{-1}(i,j)\, C_l^{gj,T}} , \end{aligned} \tag{3.30}$$

where in the second line we have used the representation of the cross-correlation functions in spherical harmonics space. The maximum likelihood amplitude is an unbiased estimator (if the underlying probability distribution is Gaussian), hence for the fiducial model we have for the average over all cosmic realisations

$$\langle \widehat{A}_{cc} \rangle_{\mathrm{all}} = 1 , \tag{3.31}$$

since $\langle \widehat{C}_l^{gi,\,T} \rangle_{\mathrm{all}} = C_l^{gi,\,T}$ by definition of the latter quantity. Note that here we have assumed that the data are taken in a universe where the underlying cosmology is actually the fiducial model. This will be assumed in the rest of this work as well.

The variance in $\widehat{A}_{cc}$ is given by

$$\begin{aligned} \sigma_{cc}^2 &\equiv \langle \left(\widehat{A}_{cc} - \langle \widehat{A}_{cc} \rangle_{\mathrm{all}}\right)^2 \rangle_{\mathrm{all}} \\ &= \left(\xi^{g,T\,\dagger}\; C_{cc}^{-1}\; \xi^{g,T}\right)^{-1} \\ &= \left[\sum_l (2l+1) \sum_{i,j} C_l^{gi,T} (C_l^{cc})^{-1}(i,j)\, C_l^{gj,T} \right]^{-1} . \end{aligned} \tag{3.32}$$

In the standard literature, an estimated significance is given to the detection of the amplitude, the estimated signal-to-noise ratio

$$\left(\frac{\widehat{S}}{N}\right)_{cc}^2 \equiv \left(\frac{\widehat{A}_{cc}}{\sigma_{cc}}\right)^2 = \frac{\left(\sum_l (2l+1) \sum_{i,j} C_l^{gi,T} (C_l^{cc})^{-1}(i,j)\, \widehat{C}_l^{gj,T}\right)^2}{\sum_l (2l+1) \sum_{i,j} C_l^{gi,T} (C_l^{cc})^{-1}(i,j)\, C_l^{gj,T}} . \tag{3.33}$$

However, since the real signal is $A_{cc} = 1$, the actual signal-to-noise ratio or detection significance is given by

$$\left(\frac{S}{N}\right)_{cc}^2 \equiv \frac{1}{\sigma_{cc}^2} = \sum_l (2l+1) \sum_{i,j} C_l^{gi,T} (C_l^{cc})^{-1}(i,j)\, C_l^{gj,T} , \tag{3.34}$$

and is therefore independent of the data.

3.3.2 Analysis of error-contributions

In this section, we analyse the different sources of noise that contribute to the total variance in eq. (3.32). In order to simplify this task, we assume that there is no shot noise in the galaxy distribution, that is, we set $\epsilon = 0$ in eq. (3.18), which means that the galaxies trace the matter distribution perfectly. Furthermore, we work with the ideal case that we have a galaxy survey that covers the whole sky and goes out to redshift 2. With these two assumptions, we have a perfect knowledge of the matter distribution δ_m relevant for the ISW effect.

For sufficiently narrow bins, the integration kernels for ISW and galaxy density contrast are approximately constant over the bin and hence $a_{lm}^{\mathrm{isw}(i)} = \mathrm{const(i)} \times a_{lm}^{gi}$. In eqs (3.30), (3.32), and (3.34), we can therefore substitute every index gi by the index isw(i), since the constant factor cancels out. Now, if one uses the ISW kernel, working with several narrow bins that cover the whole volume relevant for the ISW effect is equivalent to working with only one bin covering the same volume. This is because the ISW integrated over the whole relevant volume is exactly the information about the ISW effect contained in the CMB. Thus, one does not gain anything by working with bins if using the correct kernel. We outline the proof for this in Appendix B.1. In what follows, we therefore consider only one bin, which significantly simplifies the form of eqs (3.30), (3.32), and (3.34).

Furthermore, we note that, since the ISW effect is uncorrelated with the primordial CMB fluctuations, we have $C_l^{\mathrm{isw},T} = C_l^{\mathrm{isw}}$. The index 'all' now indicates an average over the probability distribution $\mathcal{P}(T_{\mathrm{isw}}, T_{\mathrm{prim}} \,|\, p) = \mathcal{P}(T_{\mathrm{isw}} \,|\, p)\mathcal{P}(T_{\mathrm{prim}} \,|\, p)$ (cf. section 3.2). Under the above assumptions, eq. (3.30) for the estimated amplitude reads

$$\widehat{A}_{cc} = \frac{\sum_l (2l+1) \frac{\widehat{C}_l^{\mathrm{isw},T}}{C_l^{\mathrm{isw}} + C_l^T}}{\sum_l (2l+1) \frac{C_l^{\mathrm{isw}}}{C_l^{\mathrm{isw}} + C_l^T}} , \tag{3.35}$$

with the variance (eq. 3.32)

$$\sigma_{cc}^2 = \left(\sum_l (2l+1) \frac{C_l^{\mathrm{isw}}}{C_l^{\mathrm{isw}} + C_l^T} \right)^{-1} , \tag{3.36}$$

and the signal-to-noise ratio in eq. (3.34) simplifies to

$$\left(\frac{S}{N}\right)_{cc}^2 = \sum_l (2l+1) \frac{C_l^{\mathrm{isw}}}{C_l^{\mathrm{isw}} + C_l^T} . \tag{3.37}$$

The signal-to-noise ratio as a function of the maximum summation index l_{max} for our fiducial model is depicted in the top panel of Fig. 3.1, for which we have modified CMBEASY in order to obtain C_l^{isw} and C_l^T. There are contributions to the signal-to-noise up to roughly $l = 100$. Note, though, that our assumptions of Gaussianity of the matter realisation δ_m and the assumption of Q being a linear operator do not hold on small scales where structure growth has become non-linear. However, this issue will not be addressed here and it will not affect our main results, which are due to advantages of our method on the very large scales, which are most affected by cosmic variance.

The above estimator for the amplitude is only unbiased when averaging over the joint distribution

$$\langle \widehat{A}_{cc} \rangle_{\mathrm{all}} \equiv \langle \langle \widehat{A}_{cc} \rangle_{\mathrm{prim}} \rangle_{\mathrm{isw}} = 1\,. \tag{3.38}$$

Here, we indicate averages over $\mathcal{P}(T_{\mathrm{prim}}\,|\,p)$ and $\mathcal{P}(T_{\mathrm{isw}}\,|\,p)$ by the indices 'prim' and 'isw', respectively. This means that both the primordial CMB fluctuations and the realisation of the local matter distribution are included in the error budget. We call the latter the local variance, indicating that it originates in the realisation of the matter distribution in our observed Universe. Let us now estimate the contribution of the local variance to the total variance of $\widehat{A}_{cc}$. To this end, we split the variance in eq. (3.36) into two parts

$$\begin{aligned}
\sigma_{cc}^2 &\equiv \langle \langle \left(\widehat{A}_{cc} - 1 \right)^2 \rangle_{\mathrm{prim}} \rangle_{\mathrm{isw}} \\
&= \langle \langle \left(\widehat{A}_{cc} - \langle \widehat{A}_{cc} \rangle_{\mathrm{prim}} \right)^2 \rangle_{\mathrm{prim}} \rangle_{\mathrm{isw}} \\
&\quad + \langle \left(\langle \widehat{A}_{cc} \rangle_{\mathrm{prim}} - 1 \right)^2 \rangle_{\mathrm{isw}} \\
&\equiv \sigma_{\mathrm{prim}}^2 + \sigma_{\mathrm{loc}}^2\,,
\end{aligned} \tag{3.39}$$

where we have defined the contributions to the variance coming from primordial CMB fluctuations and the local variance as σ_{prim}^2 and σ_{loc}^2, respectively. Both can be easily calculated, and the second contribution turns out to be

$$\sigma_{\mathrm{loc}}^2 = 2\, \frac{\sum_l (2l+1) \frac{\left(C_l^{\mathrm{isw}}\right)^2}{\left(C_l^T + C_l^{\mathrm{isw}}\right)^2}}{\left(\sum_l (2l+1) \frac{C_l^{\mathrm{isw}}}{C_l^T + C_l^{\mathrm{isw}}} \right)^2}\,. \tag{3.40}$$

In the bottom panel of Fig. 3.1, we plot the relative contribution of the local to the total variance, $\sigma_{\mathrm{loc}}^2/\sigma_{cc}^2$, against the maximum l that we consider in the analysis for our fiducial cosmological model. For a maximum multipole $l_{\max} = 100$, this relative contribution amounts to

$$\frac{\sigma_{\mathrm{loc}}^2}{\sigma_{cc}^2} \approx 11\%\,. \tag{3.41}$$

This estimate agrees with Cabré et al. (2007), who compare different error estimates for the standard cross-correlation method. They compare what they call the MC1 method, which only takes into account the variance in the CMB and ignores the variance in the galaxy overdensity, with their MC2 method, which includes also the variance in the galaxy overdensity. Both methods rely on performing Monte Carlo (MC) simulations of the CMB, and of the galaxy overdensity in the case of MC2, and the simulations used to compare the different error estimates have converged with an accuracy of about 5 per cent, as stated in the paper. The result is that, compared to the MC2 method, the MC1 method underestimates the error by about 10 per cent, which agrees well with our estimate.

In eq. (3.39), we have seen that the contributions to the variance of the measured amplitude of the cross-correlation come from primordial temperature fluctuations (σ_{prim}^2) as well as the specific realisation of the LSS in our Universe (σ_{loc}^2). In the following two sections, we will show that it is possible to reduce both σ_{loc}^2 and σ_{prim}^2 by working conditional on available information about the LSS and the primordial temperature fluctuations. We will start with reducing σ_{loc}^2 in the next section.

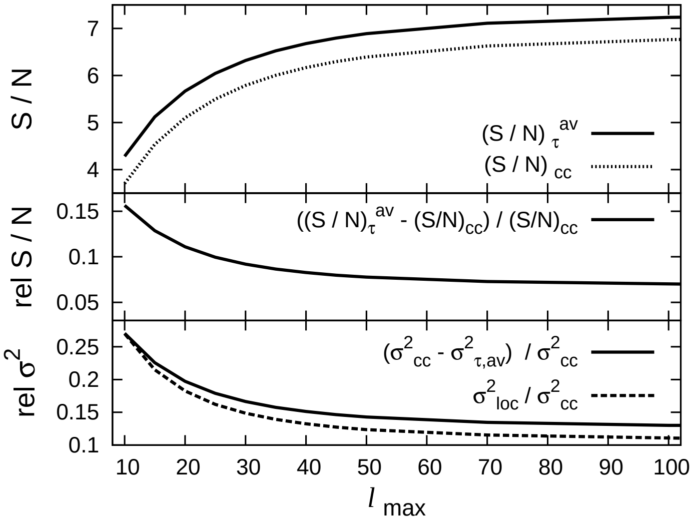


Figure 3.1: Comparison of the average signal-to-noise ratio and variance of the optimal method with the ones of the standard method for $z_{\max} = 2$. **Top panel:** Average signal-to-noise ratio of the optimal method (solid) and signal-to-noise ratio of the standard method (dashed) versus the maximal multipole considered in the analysis. **Middle panel:** Relative improvement of the average signal-to-noise ratio in the optimal method. **Bottom panel:** Average relative improvement of the variance in the optimal method (solid) and relative contribution of the local variance to the total variance in the standard method (dashed).

3.4 Optimal temperature method

In the last section, we have seen that the local variance contributes about 11 per cent to the total variance in the standard method for ISW detection, which is quite a considerable contribution. The reason is, that the ISW effect is created by the decay of the gravitational potential coming from the structures on the largest scales, i.e. from structures that have not yet undergone significant gravitational collapse and are still not decoupled from the expansion of the Universe. These largest scales are most affected by cosmic variance. Therefore, when comparing the observed (local) cross-correlation function to its cosmic mean value, the realisation of the matter distribution in our vicinity acts as a source of systematic noise in the estimation of the cross-correlation, thus leading to a biased detection significance, due to cosmic variance.

From the surveys mentioned above, the local matter distribution is known to a certain degree, and hence the local variance can be reduced by working conditional on that information. We present a generic technique of how to include the knowledge of the matter distribution into the detection of the ISW via cross-correlation, thus reducing the sources of noise to the unknown part of the matter distribution and the primordial CMB fluctuations. We define the systematic noise that comes from the known part of the matter distribution as bias, for it can be removed by working conditional on the LSS data. The method presented here is referred to as the *optimal temperature method*, in contrast to the *standard method* for ISW detection described above and the *optimal polarization method* explained below.

The main idea of the optimal temperature method is to create an ISW template from a Wiener filter reconstruction of the LSS. We then estimate the amplitude of that template in the CMB data, using a maximum likelihood estimator. Since we assume everything to be Gaussian distributed, this maximum likelihood estimator is equivalent to the estimator we obtain from an optimal matched filter approach. Our amplitude estimate is unbiased already when averaging conditional on the observed galaxy density contrast, δ_g. Since we work conditional on the LSS, the variance in the estimated amplitude and the signal-to-noise ratio both depend on the actual realisation of the matter in the Universe. For an ideal LSS survey, we show that the average variance in the detected amplitude is reduced by 13 per cent in the optimal temperature method. As we show in this section, in the framework of the optimal temperature method it is straightforward to correct for the magnification bias due to gravitational lensing, as described by Loverde et al. (2007). Furthermore, there is no need to estimate the covariance matrix by Monte Carlo simulations as in the standard method. This saves time and increases the accuracy of the method.

Note that a different attempt to make the detection of the ISW unbiased by the realisation of the local matter distribution was made by Zhang (2006a). It involves comparing CMB-galaxy and lensing-galaxy cross-correlation functions, and thus relies on nowadays still-difficult lensing measurements. Another work which does not suffer from local variance is by Hernández-Monteagudo (2008). He implements an optimal matched filter in spherical harmonics space, and finds by numerical comparison that it always performs better than or equally well as the standard method. However, Hernández-Monteagudo (2008) works directly on the sphere, without using a Wiener filter reconstruction of the LSS, and is therefore slightly sub-optimal in exploiting the available three-dimensional information on galaxy positions.

In principle, the idea of working conditional on the LSS data can also be used to decrease the variance of the detection of other secondary effects on the CMB, such as the kinetic Sunyaev-Zel'dovich effect, the Rees-Sciama effect or gravitational lensing. However, in our derivation we assume a Gaussian data-model, which is very well suited for the ISW effect, because the very large scales on which the ISW effect is created are still Gaussian. The assumption of Gaussianity breaks down on smaller scales, though, due to non-linear structure growth. For the detection of secondary effects on non-linear scales, our method would therefore need to be modified to account for the non-linearities. This is possible using information field theory presented in section 2.3 and in Enßlin et al. (2009), however, it is beyond the scope of this work, and we leave the extension of our methods into the non-Gaussian regime for future work.

3.4.1 Derivation of the posterior distribution

Let us first ask the question what the observed galaxy density contrast tells us about the matter distribution δ_m. The data model in eq. (3.18) is the same as the one in eq. (2.16), and signal and noise are Gaussian distributed, cf. eqs (3.1) and (3.21). We therefore obtain the posterior distribution

$$\mathcal{P}(\delta_m \,|\, \delta_g, p) = \mathcal{G}(\delta_m - D_m j, D_m)\,, \tag{3.42}$$

where we have defined the Wiener variance

$$D_m \equiv (R^\dagger {N_\epsilon}^{-1} R + {S_m}^{-1})^{-1}\,, \tag{3.43}$$

and the response over noise weighted data

$$j \equiv R^{\dagger} N_{\epsilon}{}^{-1} \delta_g \,. \tag{3.44}$$

From eq. (3.42), we can directly read off the posterior mean of the matter distribution δ_m,

$$\delta_m^{\mathrm{rec}} \equiv D_m j = (R^{\dagger} N_{\epsilon}{}^{-1} R + S_m{}^{-1})^{-1} R^{\dagger} N_{\epsilon}{}^{-1} \delta_g \,. \tag{3.45}$$

This is the Wiener filter applied to the galaxy-overdensity (Wiener 1950; Zaroubi et al. 1995; Zaroubi 1995; Fisher et al. 1995; Erdoğdu et al. 2004; Kitaura & Enßlin 2008). We call this estimator a *reconstruction* of the matter distribution from the galaxy survey, thus the symbol δ_m^{rec}.

With this knowledge of the matter distribution, let us now find the posterior distribution for $T = T_{\mathrm{isw}} + T_{\mathrm{prim}} + T_{\mathrm{det}}$, given the observed galaxy density contrast δ_g. The probability distribution for T_{isw}, obtained from the one for δ_m, eq. (3.42), reads[5]

$$\mathcal{P}(T_{\mathrm{isw}} \,|\, \delta_g, p) = \mathcal{G}(T_{\mathrm{isw}} - T_{\tau}, \mathcal{Q}\, D_m \mathcal{Q}^{\dagger}) \,, \tag{3.46}$$

where we have defined the ISW template

$$T_{\tau} \equiv \mathcal{Q}\, \delta_m^{\mathrm{rec}} \,. \tag{3.47}$$

Since the uncertainty in the reconstructed matter distribution is not related to the primordial CMB fluctuations (cf. section 3.2.1), the joint probability distribution for T_{isw}, T_{prim}, and T_{det} given δ_g factorises:

$$\begin{aligned} \mathcal{P}(T_{\mathrm{isw}}, T_{\mathrm{prim}}, T_{\mathrm{det}} \,|\, \delta_g, p) &= \mathcal{P}(T_{\mathrm{isw}} \,|\, \delta_g, p)\, \mathcal{P}(T_{\mathrm{prim}} \,|\, p) \\ &\quad\; \mathcal{P}(T_{\mathrm{det}} \,|\, p) \,. \end{aligned} \tag{3.48}$$

Note that in the above equation we have used the fact that the primordial CMB fluctuations do not depend on the galaxy distribution. We now again use the fact that the sum of stochastically independent Gaussian distributed random variables is again Gaussian distributed with the sum of the covariance matrices. We then obtain the posterior distribution for T, given the LSS data:

$$\mathcal{P}(T \,|\, \delta_g, p) = \mathcal{G}(T - T_{\tau}, C_{\Delta T}) \,. \tag{3.49}$$

Here, we have used the probability distributions for T_{prim} and T_{det}, eqs (3.5) and (3.12), and the covariance matrix for the total noise is

$$C_{\Delta T} \equiv \langle \Delta T \Delta T^{\dagger} \rangle = \mathcal{Q}\, D_m \mathcal{Q}^{\dagger} + C_{\mathrm{prim}} + C_{\mathrm{det}} \,, \tag{3.50}$$

where we have defined $\Delta T \equiv T - T_{\tau}$, and the average has to be taken over the distribution given in eq. (3.49). As we have already said before, we will neglect the detector noise in the rest of this section ($C_{\mathrm{det}} \approx 0$) and only include it when deriving the likelihood in section 3.7.

---

[5] note that we again use that linear transformations of Gaussian distributed random vectors are again Gaussian distributed, cf. eq. (3.11)

3.4.2 Estimation of the ISW amplitude

We can now ask the same question as before, namely if it is at all possible to detect a non-zero amplitude A_τ that we put in front of our ISW template in eq. (3.49). Again, we can write down the likelihood function for the amplitude

$$\mathcal{P}(T\,|\,A_\tau, \delta_g, p) = \mathcal{G}(T - A_\tau T_\tau, C_{\Delta T})\,, \tag{3.51}$$

and estimate the amplitude by a maximum likelihood estimator

$$\widehat{A_\tau} = \frac{T^\dagger C_{\Delta T}^{-1} T_\tau}{{T_\tau}^\dagger C_{\Delta T}^{-1} T_\tau} = \frac{\sum_l (2l+1)\frac{\widehat{C_l^{T_\tau,T}}}{C_l^{\Delta T}}}{\sum_l (2l+1)\frac{\widehat{C_l^{T_\tau}}}{C_l^{\Delta T}}}\,, \tag{3.52}$$

where we have assumed $f_{\rm sky} = 1$ in the second step, and defined the estimator $\widehat{C_l^{T_\tau}}$ of the ISW power spectrum analogous to the cross-correlation estimator in eq. (3.25):

$$\widehat{C_l^{T_\tau}} \equiv \frac{1}{2l+1} \sum_m |a_{lm}^{T_\tau}|^2\,. \tag{3.53}$$

This maximum likelihood amplitude is again an unbiased estimator, but now with respect to the probability distribution conditional on δ_g,

$$\langle \widehat{A_\tau} \rangle_{\rm cond} = 1\,, \tag{3.54}$$

where the index ‘cond’ on the average denotes an average over the distribution $\mathcal{P}(T\,|\,A_\tau, \delta_g, p)$.

In other words, we have eliminated the noise component coming from the realisation of the known part of δ_m, thus reducing the sources of noise to the unknown part of δ_m and the primordial CMB fluctuations. The variance in $\widehat{A_\tau}$ is

$$\begin{aligned} \sigma_\tau^2 &\equiv \langle \left(\widehat{A_\tau} - \langle \widehat{A_\tau} \rangle_{\rm cond}\right)^2 \rangle_{\rm cond} \\ &= \left({T_\tau}^\dagger C_{\Delta T}^{-1} T_\tau\right)^{-1} = \left(\sum_l (2l+1) \frac{\widehat{C_l^{T_\tau}}}{C_l^{\Delta T}}\right)^{-1}\,, \end{aligned} \tag{3.55}$$

(again we have assumed $f_{\rm sky} = 1$ in the second step), and we obtain the signal-to-noise ratio for a full-sky analysis

$$\left(\frac{S}{N}\right)_\tau^2 \equiv \frac{1}{\sigma_\tau^2} = \sum_l (2l+1) \frac{\widehat{C_l^{T_\tau}}}{C_l^{\Delta T}}\,. \tag{3.56}$$

Note that the error estimate (and hence the signal-to-noise ratio) of the optimal temperature method depends on the concrete LSS realisation, and how well it is suited to detect the ISW effect. In a universe where by chance the local LSS does/does not permit a good ISW detection, the error is small/large, as it should be.

We would like to point out that in our method there is no need to estimate the covariance matrices from Monte Carlo simulations, since for a given set of cosmological parameters, the matter covariance matrix (power spectrum) S_m can be calculated analytically using the fitting formula provided by Bardeen et al. (1986), since it is still linear on the scales we are interested in. $C_{\rm prim}$ can be obtained from Boltzmann codes such as CMBEASY, and the noise covariance N_ϵ can be estimated from the data.

3.4.3 Comparison of signal-to-noise ratios and biasing

In order to compare our method to the standard one, let us now again make the simplifying assumption that there is no shot noise in the galaxy distribution, and that we have a perfect galaxy survey, as we did in section 3.3.2. At the end of this section, we will approximately look at the effects of a galaxy survey that is incomplete in redshift, i.e. that goes out to a maximal redshift $z_{\rm max} < 2$. For the perfect survey, the shot noise covariance matrix N_ϵ is zero, and hence the posterior for δ_m in eq. (3.42) is infinitely sharply peaked around the reconstruction $\delta_m^{\rm rec}$ (eq. 3.45), which turns into

$$\begin{aligned} \delta_m^{\rm rec} &= (R^\dagger {N_\epsilon}^{-1} R)^{-1} R^\dagger {N_\epsilon}^{-1} \delta_g \\ &= R^{-1}\delta_g \, . \end{aligned} \tag{3.57}$$

Here, R^{-1} should be read as the pseudo-inverse of R, e.g. as defined in terms of Singular Value Decomposition (see Press et al. (1992) and Zaroubi et al. (1995)).

The posterior for δ_m in eq. (3.42) is therefore now a Dirac delta function

$$\mathcal{P}(\delta_m \,|\, \delta_g, p) = \delta(\delta_m - R^{-1}\delta_g) \, , \tag{3.58}$$

which makes our ISW template exact, and the noise covariance matrix due to the error in the reconstruction is zero, $\mathcal{Q} D_m \mathcal{Q}^\dagger = 0$, thus leaving us with $C_{\Delta T} = C_{\rm prim} = C_T - C_{\rm isw}$. Since our perfect LSS survey covers the complete volume relevant for the ISW effect, our template is now equal to the ISW-temperature fluctuations, $T_\tau = T_{\rm isw}$. We can then substitute all indices τ in eqs (3.52)-(3.56) by the index isw, and the estimated amplitude becomes

$$\widehat{A}_\tau = \frac{\sum_l (2l+1) \frac{\widehat{C}_l^{{\rm isw},T}}{C_l^T - C_l^{\rm isw}}}{\sum_l (2l+1) \frac{\widehat{C}_l^{\rm isw}}{C_l^T - C_l^{\rm isw}}} \, , \tag{3.59}$$

with the variance

$$\sigma_\tau^2 = \left(\sum_l (2l+1) \frac{\widehat{C}_l^{\rm isw}}{C_l^T - C_l^{\rm isw}} \right)^{-1} \, , \tag{3.60}$$

and the signal-to-noise ratio

$$\left(\frac{S}{N}\right)_\tau^2 = \sum_l (2l+1) \frac{\widehat{C}_l^{\rm isw}}{C_l^T - C_l^{\rm isw}} \, . \tag{3.61}$$

As we mentioned before, the variance, and thus the signal-to-noise ratio of the optimal temperature method, depend on the actual realisation of the matter distribution in our observed Universe. In Fig. 3.2, we plot the probability distribution of our signal-to-noise ratio for $l_{\rm max} = 100$ and $z_{\rm max} = 2$, which we have inferred from the distribution of $T_{\rm isw}$ using the central limit theorem

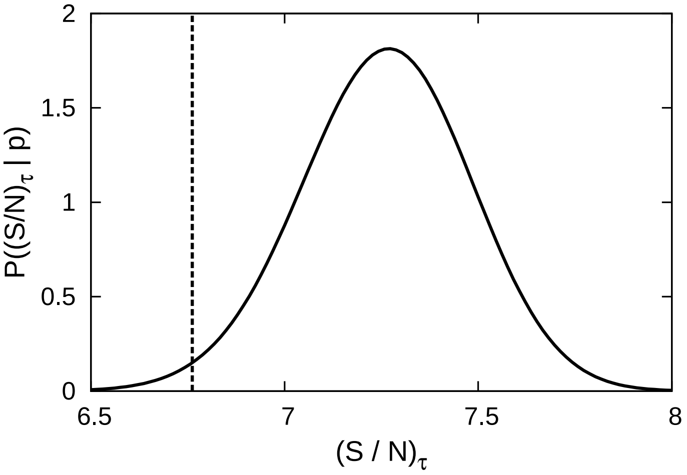


Figure 3.2: Probability distribution of the signal-to-noise ratio in the optimal temperature method (solid) for $l_{\max} = 100$ and $z_{\max} = 2$. The vertical line (dashed) shows the signal-to-noise ratio of the standard method for comparison.

for $(S/N)^2_\tau$, and from that deriving the distribution for $(S/N)_\tau$.[6] We have also checked the validity of the central limit theorem in this case by comparing with the correct probability distribution of the signal-to-noise ratio given by an expansion into Laguerre polynomials as derived, e.g., in Castaño-Martínez & López-Blázquez (2007). The probability distribution is such that the signal-to-noise ratio can easily differ by $\Delta\,(S/N)_\tau \approx 1$ for two different realisations of the matter distribution.

The mean signal-to-noise ratio, $(S/N)^{\mathrm{av}}_\tau \equiv 1/\sqrt{\sigma^2_{\tau,\,\mathrm{av}}} \equiv 1/\sqrt{\langle\sigma^2_\tau\rangle_{\mathrm{isw}}}$, increases with $l_{\max}$ as it did for the standard method. For every $l_{\max}$, we compare the mean signal-to-noise ratio of the optimal temperature method to the signal-to-noise ratio of the standard method (cf. eq. 3.37) in the top panel of Fig. 3.1, again for $z_{\max} = 2$. Note that in our formula for the signal-to-noise ratio, eq. (3.61), there is now a minus sign between C^T_l and C^{isw}_l, in contrast to the signal-to-noise ratio of the standard method in eq. (3.37), which has a plus sign instead. Thus we take advantage of the LSS instead of moving it into the error budget. The absolute enhancement of the signal-to-noise ratio in our method is therefore independent of $l_{\max}$, since the main advantage of working conditional on the LSS arises on the very large scales, where the contribution of the ISW to the CMB is highest. The average relative improvement of the signal-to-noise is depicted in the middle panel of Fig. 3.1. It amounts to about 7 per cent for $l_{\max} = 100$. In the bottom panel of Fig. 3.1, we compare the mean relative improvement $(\sigma^2_{cc} - \sigma^2_{\tau,\,\mathrm{av}})/\sigma^2_{cc}$ of the variance in the optimal temperature method with the contribution of the local to the total variance in the standard method. The variance is reduced by about 13 per cent in the optimal temperature method, as compared to the standard method.

Note that the maximal average signal-to-noise ratio we can hope for when trying to detect the ISW via cross-correlation, given a perfect LSS survey, is $(S/N)^{\mathrm{av}}_\tau \approx 7.3$, with a variance as depicted in Fig. 3.2. Hence, if we are lucky and live in an environment that allows for a high signal-to-noise ratio, we can maximally obtain a detection significance of about $(7.5-8)\,\sigma$.

Let us now look at the effect of an incomplete galaxy survey. Incomplete galaxy surveys can be treated generically with our method, because the dark matter field, and thus the ISW effect, are split

---

[6] This will provide accurate results for multipoles $l \gg 1$, however, is a coarse approximation in the regime $l \sim 1$.

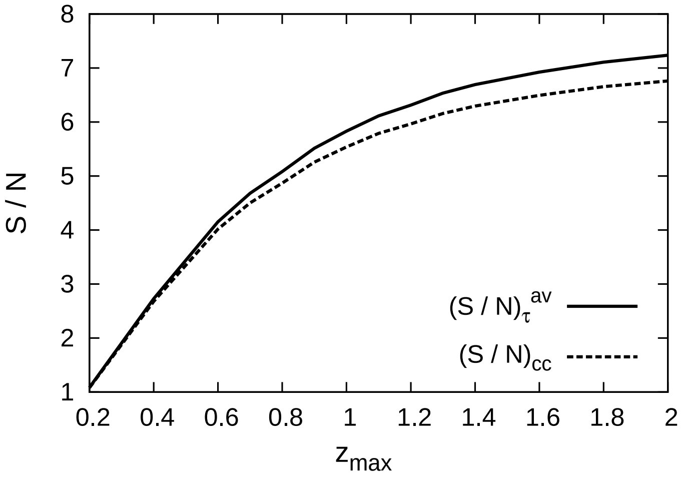


Figure 3.3: Average signal-to-noise ratio of the optimal temperature method (solid) and signal-to-noise ratio of the standard method (dashed) versus z_{max} for $l_{\mathrm{max}} = 100$.

into a known part (the reconstruction) and an unknown part (an additive noise term uncorrelated with the reconstruction). However, for now we only want to give a rough estimate of the consequences of an incomplete survey. Therefore, we introduce a sharp cut-off in redshift, z_{max}, and we simply redefine $T_{\mathrm{isw}} \equiv T_{\mathrm{isw}}(< z_{\mathrm{max}})$ to be the part of the ISW effect created at $z < z_{\mathrm{max}}$. The part of the ISW that has been created at $z > z_{\mathrm{max}}$ is then considered part of the primordial temperature fluctuations T_{prim}. The power-spectra C_l^{isw} and C_l^{prim} are redefined accordingly. With this redefinition, we have introduced a correlation between what we consider the ISW effect and primordial fluctuations, which we would not have if we had used the reconstruction for redefining T_{isw}. However, for getting the picture, we ignore this subtlety for the moment.[7]

In Fig. 3.3, we plot the signal-to-noise ratio of the standard method together with the average signal-to-noise ratio of the optimal temperature method versus z_{max} for $l_{\mathrm{max}} = 100$, where we have used the above-described redefinition of C_l^{isw} in eqs (3.61) and (3.37). With decreasing maximal redshift of the LSS survey, the total signal-to-noise ratio in both methods goes down, as does its relative enhancement of the optimal temperature method as compared to the standard one. The relative contribution of the local to the total variance in the standard method goes down with decreasing survey depth as well.

Currently available full-sky LSS surveys (which effectively cover about 70-80 per cent of the sky after masking out the galaxy, see Giannantonio et al. (2008)) are either very shallow in redshift (Two-Micron All-Sky Survey, Jarrett et al. (2000)), or they have an uncertain redshift distribution which results in a large uncertainty in the ISW template (NRAO VLA Sky Survey (Condon et al. 1998), High Energy Astrophysics Observatory (Boldt 1987)). The SDSS luminous red galaxy (LRG) sample covers a redshift range up to $z \sim 0.7$, and the QSO sample even reaches out to $z \sim 2.5$. However, the SDSS only covers about a quarter of the sky, and thus at most a quarter of the ISW-relevant volume. Even with future surveys such as EUCLID[8], we will effectively (after masking out the galaxy) only observe an area of about $20,000\,\mathrm{deg}^2$, which is roughly half of the sky. We would then *on average* still only obtain an enhancement of the detection significance by at most 3.5 per cent.

---

[7] The ratio of this neglected coupling to the template strength gets large for small z_{max}. Our estimates are therefore less accurate in this regime.

[8] `http://sci.esa.int/science-e/www/area/index.cfm?fareaid=102`

However, as we stated in section 3.3.2, the amplitude-estimate of the standard method is biased when the averaging is performed conditional on the galaxy-data δ_g, due to local variance. This leads to an over- or underestimation of the detection significance, since the estimated amplitude is used when estimating the signal-to-noise ratio from the data. As we have shown, the contribution of the local to the total variance of the estimator is quite small, about 11 per cent for an ideal galaxy survey and even smaller for a shallower survey or a survey that covers only a fraction of the sky. We thus expect the biasing effect in general to be quite weak. However, we could be unlucky and live in an unlikely realisation of the matter distribution, given the power spectrum, which would enhance the effect of the biasing.

With the method we presented in this section, the local variance effect is reduced. If we knew the local matter distribution perfectly, we would not be affected by local variance at all, as we have shown. Unfortunately, we have to rely on reconstructions of the matter distribution from LSS surveys, which suffer from shot noise and the effects of mask and selection function. However, the reconstruction treats mask and selection function in an optimal way, and extracts the maximum amount of information from the LSS data which can then be used in the ISW detection.

3.5 Optimal polarization method

With the optimal temperature method introduced in section 3.4, we were able to reduce the low redshift cosmic variance effect in the estimate of the ISW amplitude, i.e. we reduced the noise coming from the specific realisation of LSS in our Universe, σ^2_{loc}. However, in both the standard method and the optimal temperature method, the main source of uncertainty in the detection of the ISW effect comes from chance correlations of primordial CMB fluctuations with the LSS, σ^2_{prim} (cf. section 3.3.2). In this section, we show how σ^2_{prim} can be reduced by including polarization data in the analysis. CMB polarization is correlated with the temperature fluctuations, and can thus be used to obtain information about the latter. We use the observed E-mode polarization map, which we translate into a temperature map using the TE cross-power spectrum. The obtained temperature map is then subtracted from the observed temperature map, and thus no longer contributes to the noise budget of the detected signal. In other words, we perform our amplitude estimate of the ISW template conditional on the part of the temperature fluctuations which is correlated with the polarization map.

Again we work with a Gaussian data model, as before when deriving the optimal temperature method. This assumption is very well-suited for the ISW effect. Using *information field theory* (Enßlin et al. 2009), it is possible to extend the optimal polarization method into the non-Gaussian regime, in order to use it for detecting secondary effects on smaller scales, but again we leave those extensions for future work.

Using polarization data to reduce the noise in the detection of secondary effects was first proposed by Robert Crittenden, following a suggestion of Lyman Page (Crittenden 2006). He already derived the uncorrelated temperature power spectrum, which we show in Figure 3.4, and roughly estimates the improvement of the signal-to-noise ratio for ISW detection to be around 20 per cent, which we confirm with our calculations.

3.5.1 Uncorrelated temperature maps

In order to use the information contained in the polarization data in our analysis, we enlarge our data vector d to include the observed polarization map P:

$$d \equiv \begin{pmatrix} T \\ P \end{pmatrix}, \tag{3.62}$$

where the polarization map P contains the maps of the Stokes Q and U parameters,

$$P \equiv \begin{pmatrix} Q \\ U \end{pmatrix}. \tag{3.63}$$

As we have seen in chapter 1, we represent the polarization in spherical harmonics space in terms of the so-called E and B modes, and thus

$$a^d_{lm} \equiv \begin{pmatrix} a^T_{lm} \\ a^E_{lm} \\ a^B_{lm} \end{pmatrix}. \tag{3.64}$$

In principle, it is possible that the secondary effect we are looking for is also present as a small signal in the polarization data. If the temperature anisotropies created by the secondary effect exhibit a quadrupole component at the time of reionization, this quadrupole will be re-scattered by free electrons and create a polarization signal (Zaldarriaga 1997). However, for the ISW this effect has been proved to be small (Cooray & Melchiorri 2006). It should also be small for the RS effect, lensing and the kinetic SZ effect, the highest contributions of which are on relatively small scales and are mostly created after reionization. Thus, as a first approximation we assume that the polarization data do not carry any signal of the effect we want to detect. Given the signal template for the temperature, T_τ, our signal template τ is then

$$\begin{aligned} \tau &\equiv \begin{pmatrix} T_\tau \\ 0 \end{pmatrix}, \\ a^\tau_{lm} &\equiv \begin{pmatrix} a^{T_\tau}_{lm} \\ 0 \\ 0 \end{pmatrix}, \end{aligned} \tag{3.65}$$

and the data model becomes

$$d = \begin{pmatrix} T \\ P \end{pmatrix} = \begin{pmatrix} T_\tau + \Delta T \\ P \end{pmatrix}, \tag{3.66}$$

with $\Delta T \equiv T - T_\tau$ as before. The observed polarization map, $P = P_{\mathrm{cmb}} + P_{\mathrm{fg}} + P_{\mathrm{det}}$, consists of the cosmological polarization signal P_{cmb}, residual Galactic foregrounds after foreground removal P_{fg}, and the detector noise P_{det}. Assuming again Gaussianity, we can write down the likelihood

$$\mathcal{P}(d \,|\, \tau, p) = \mathcal{G}(d - \tau, \tilde{C}), \tag{3.67}$$

where the covariance matrix $\tilde{C}$ is

$$\tilde{C} \equiv \langle (d-\tau)(d-\tau)^\dagger \rangle_{\rm cond} = \begin{pmatrix} C_{\Delta T} & C_{\Delta T,P} \\ C_{P,\Delta T} & C_P \end{pmatrix}, \tag{3.68}$$

with

$$\begin{aligned} C_{\Delta T} &\equiv \langle \Delta T \Delta T^\dagger \rangle_{\rm cond}\,, && (3.69)\\ C_{\Delta T,P} &\equiv \langle \Delta T P^\dagger \rangle_{\rm cond}\,, && (3.70)\\ C_P &\equiv \langle P P^\dagger \rangle_{\rm cond}\,, && (3.71) \end{aligned}$$

and we have redefined the index ‘cond’ to denote the average over the probability distribution in eq. (3.67).

By using block-wise inversion of the covariance matrix $\tilde{C}$, it is possible to rewrite the likelihood as a product of the part of the temperature map which is uncorrelated with polarization, and a polarization part. To this end, let us define the uncorrelated temperature map and covariance

$$\begin{aligned} T^{\rm rec}_{\rm uncorr} &\equiv T - C_{\Delta T,P} C_P^{-1} P \\ &= T - T^{\rm rec}_{\rm corr}\,, \\ C_{\rm uncorr} &\equiv C_{\Delta T} - C_{\Delta T,P} C_P^{-1} C_{P,\Delta T} \\ &= C_{\Delta T} - C_{\rm corr}\,, \end{aligned} \tag{3.72}$$

where we have introduced the definitions

$$\begin{aligned} T^{\rm rec}_{\rm corr} &\equiv C_{\Delta T,P} C_P^{-1} P\,, \\ C_{\rm corr} &\equiv C_{\Delta T,P} C_P^{-1} C_{P,\Delta T}\,. \end{aligned} \tag{3.73}$$

We have added the index ‘rec’ of $T^{\rm rec}_{\rm corr}$ and $T^{\rm rec}_{\rm uncorr}$ to indicate that under certain assumptions $T^{\rm rec}_{\rm corr}$ actually corresponds to a Wiener reconstruction of the polarization map, translated to a temperature map. We prove this in appendix C.1. Effectively, what we have done is the following. We have a polarization map P, which is correlated with the temperature fluctuations ΔT via $C_{\Delta T,P}$. That is, the polarization map contains information about the temperature map, which we can translate into the part of the temperature map which is correlated with the polarization, using the prescription $T^{\rm rec}_{\rm corr} \equiv C_{\Delta T,P} C_P^{-1} P$. This correlated part of the temperature map is subtracted from the observed one, in order to obtain the part of the temperature map which is uncorrelated with the polarization, $T^{\rm rec}_{\rm uncorr}$.

With the above definitions, the likelihood becomes

$$\begin{aligned} \mathcal{P}(d\,|\,\tau,p) &= \mathcal{P}(T\,|\,P,T_\tau,p)\,\mathcal{P}(P\,|\,p) \\ &= \mathcal{G}(T^{\rm rec}_{\rm uncorr} - T_\tau, C_{\rm uncorr})\,\mathcal{G}(P,C_P)\,, \end{aligned} \tag{3.74}$$

as we prove in Appendix B.2. Now, our goal is to find the signal template T_τ in the CMB data. The polarization part of the above likelihood, $\mathcal{G}(P,C_P)$, does not depend on the signal template, nor does the uncorrelated temperature part explicitely depend on P. In other words, the observed polarization map does not contain relevant information any more after introducing the uncorrelated temperature fluctuations. Thus, we can marginalise over it, and continue only with the likelihood of the uncorrelated temperature map

$$\mathcal{P}(T^{\rm rec}_{\rm uncorr}\,|\,T_\tau, p) \quad \equiv \quad \mathcal{G}(T^{\rm rec}_{\rm uncorr} - T_\tau, C_{\rm uncorr})\,. \tag{3.75}$$

The uncorrelated temperature map, $T^{\rm rec}_{\rm uncorr}$, fluctuates around our signal template T_τ only with the variance $C_{\rm uncorr}$, which is smaller than the full variance $C_{\Delta T}$ of the observed temperature map. This reduced variance is the uncertainty going into our signal detection problem now, rather than the full variance of the original temperature fluctuations. Note that it is straightforward to derive the factorised likelihood also for the case that we do have a non-zero signal template P_τ for the polarization part. In that case, the covariance matrix $\tilde{C}$ is slightly changed, as well as the definitions of the uncorrelated temperature map and covariance matrix, and we can no longer neglect the polarization part of the likelihood. Please refer to Appendix B.2 for details.

In multipole space, only the E-mode, a^E_{lm}, is correlated with the temperature fluctuations, whereas the B-mode is not (cf. section 1.4.4). In the isotropic case, transforming $T^{\rm rec}_{\rm corr}$ and $C_{\rm corr}$ into multipole space results in

$$\begin{aligned} a^{T^{\rm rec}_{\rm corr}}_{lm} &= \frac{C^{\Delta T,E}_l}{C^E_l} a^E_{lm}\,, \\ C^{T^{\rm rec}_{\rm corr}}_l &= \frac{\left(C^{\Delta T,E}_l\right)^2}{C^E_l}\,, \end{aligned} \tag{3.76}$$

where we have defined $C^{\Delta T,E}_l \equiv \langle a^{\Delta T}_{lm} a^{E*}_{lm}\rangle$, and the corresponding quantities for $T^{\rm rec}_{\rm uncorr}$ and $C_{\rm uncorr}$ are

$$\begin{aligned} a^{T^{\rm rec}_{\rm uncorr}}_{lm} &= a^T_{lm} - a^{T^{\rm rec}_{\rm corr}}_{lm}\,, \\ C^{\rm uncorr}_l &= C^{\Delta T}_l - C^{T^{\rm rec}_{\rm corr}}_l\,, \end{aligned} \tag{3.77}$$

with $C^{\Delta T}_l \equiv \langle a^{\Delta T}_{lm} a^{\Delta T*}_{lm}\rangle$. Note that if isotropy does not hold, for example, if we cut out a part of the sky by applying a mask or if we have inhomogeneous noise, the covariance matrices $C_{\rm corr}$ and $C_{\rm uncorr}$ are no longer diagonal in spherical harmonics space.

In the next section, we estimate by how much the variance in the detection of secondary signals can actually be reduced by working conditional on the correlated part of the temperature fluctuations. We will perform this estimate for an ideal scenario, where we have noiseless data.

3.5.2 Reduction of variance

We now attempt to get a feeling for the reduction of the variance in the detection of secondary signals, obtained from including polarization data. We look at the best case scenario, in which we have full-sky polarization maps, and there is no contamination of the CMB polarization signal $P_{\rm cmb}$ by foregrounds or detector noise. Note that we are still far from this scenario with our currently available polarization data. The WMAP satellite has provided us with full-sky polarization maps, but since WMAP was primarily designed to measure temperature fluctuations, the detector noise in the polarization data is very high. Even more problematic is the contamination of the CMB polarization by synchrotron and dust emission from our Galaxy. Therefore, the WMAP team has masked out about 25 per cent of the sky, but even in the remaining parts of the sky the Galactic foregrounds dominate over the CMB signal. In chapter 4, we will obtain the correlated and uncorrelated temperature maps for the WMAP data, and have a look at the noise contamination. With the *Planck* mission (Tauber 2000), the detector noise will be reduced by a considerable amount, however, lots of work has still to be done in order to understand and remove the Galactic foregrounds. Therefore, our estimate of the reduction of variance has to be understood as an upper limit of what we can gain from polarization data.

As in section 3.4, let us put an amplitude in front of the signal template in eq. (3.75), and estimate it from the data using a maximum likelihood estimator:

$$\widehat{A_\tau} = \frac{{T^{\mathrm{rec}}_{\mathrm{uncorr}}}^\dagger C^{-1}_{\mathrm{uncorr}} T_\tau}{{T_\tau}^\dagger C^{-1}_{\mathrm{uncorr}} T_\tau} = \frac{\sum_l (2l+1) \frac{\widehat{C}_l^{T^{\mathrm{rec}}_{\mathrm{uncorr}},T_\tau}}{C_l^{\mathrm{uncorr}}}}{\sum_l (2l+1) \frac{\widehat{C}_l^{T_\tau}}{C_l^{\mathrm{uncorr}}}} \,. \tag{3.78}$$

Here, the last expression is in spherical harmonics space. The variance of $\widehat{A_\tau}$ is now

$$\sigma_\tau^2 = \left({T_\tau}^\dagger C^{-1}_{\mathrm{uncorr}} T_\tau\right)^{-1} = \left(\sum_l (2l+1) \frac{\widehat{C}_l^{T_\tau}}{C_l^{\mathrm{uncorr}}}\right)^{-1} \,, \tag{3.79}$$

and thus the signal-to-noise ratio becomes

$$\begin{aligned} \left(\frac{S}{N}\right)^2_{\mathrm{pol}} &= {T_\tau}^\dagger C^{-1}_{\mathrm{uncorr}} T_\tau = \sum_l (2l+1) \frac{\widehat{C}_l^{T_\tau}}{C_l^{\mathrm{uncorr}}} \\ &= \sum_l \frac{(2l+1)\,\widehat{C}_l^{T_\tau}}{C_l^{\Delta T} - \left(C_l^{\Delta T,E}\right)^2 / C_l^E} \,. \end{aligned} \tag{3.80}$$

Note that we have used the index ‘pol’ to indicate that this is the signal-to-noise ratio one obtains when using the polarization data to reduce the variance. Comparing the signal-to-noise ratio in eq. (3.80) with the one in eq. (3.56), we see that by including the information contained in the polarization data, we reduce the variance in every mode by the term $\left(C_l^{\Delta T,E}\right)^2 / C_l^E$.

Let us now get an impression of how much the variance gets reduced for the different multipoles. To this end, we neglect the detector noise E_{det} and the foreground noise E_{fg}[9] (note that we have neglected the detector and foreground noise for the temperature data, too), which allows us write

$$C_l^{\Delta T,E} \approx C_l^{T,E_{\mathrm{cmb}}} - C_l^{T_\tau,E_{\mathrm{cmb}}} \tag{3.81}$$

$$C_l^{\Delta T} \approx C_l^{T} - 2C_l^{T,T_\tau} + C_l^{T_\tau} \tag{3.82}$$

$$C_l^{E} \approx C_l^{E_{\mathrm{cmb}}} \,. \tag{3.83}$$

We furthermore neglect the cross-term $C_l^{T_\tau,E_{\mathrm{cmb}}}$. For the ISW effect, we have verified numerically that it is negligible, as shown in Fig. 3.5. For the kinetic SZ effect and the RS effect, the template itself is so small that we can also certainly neglect $C_l^{T_\tau,E_{\mathrm{cmb}}}$. Then, the uncorrelated temperature

---

[9] In reality, Galactic E-mode foregrounds E_{fg} are likely to be the limiting factor in the improvement of the detection significance coming from including polarization data. We comment on this at the end of this section.

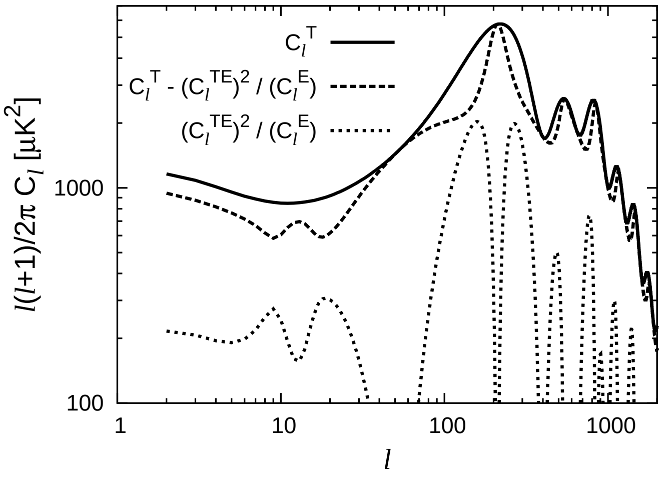


Figure 3.4: Reduction of the variance in the detection of secondary temperature signals by using the information contained in polarization data. Shown are the CMB temperature power spectrum C_l^T (solid), and the template-free part of the uncorrelated temperature power spectrum $C_l^T - \left(C_l^{T,E_{\rm cmb}}\right)^2 / C_l^{E_{\rm cmb}}$ (dashed), together with the part of the CMB power spectrum coming from the correlated part of the temperature fluctuations which we infer from the polarization map, $\left(C_l^{T,E_{\rm cmb}}\right)^2 / C_l^{E_{\rm cmb}}$ (dotted).

power spectrum defined in eq. (3.77) becomes

$$C_l^{\rm uncorr} \approx C_l^T - 2C_l^{T,T_\tau} + C_l^{T_\tau} - \frac{\left(C_l^{T,E_{\rm cmb}}\right)^2}{C_l^{E_{\rm cmb}}} . \tag{3.84}$$

In Fig. 3.4, we plot the template-free part of the uncorrelated temperature power spectrum, $C_l^T - \left(C_l^{T,E_{\rm cmb}}\right)^2 / C_l^{E_{\rm cmb}}$ (note that we have not included the template-dependent terms $-2C_l^{T,T_\tau}$ and $C_l^{T_\tau}$ in the plot), which gives us an impression of how the variance coming from primordial temperature fluctuations is being reduced by including polarization data. The variance will be further reduced by working conditional on the signal template T_τ, which is encoded in the terms $-2C_l^{T,T_\tau}$ and $C_l^{T_\tau}$, and already described in section 3.4. We also plot the original CMB power spectrum C_l^T and the difference to the uncorrelated one for comparison. We have assumed a flat ΛCDM model with the parameter values given by Komatsu et al. (2009), table 1 ($\Omega_b h^2 = 0.02265, \Omega_\Lambda = 0.721$, $h = 0.701$, $n_s = 0.96$, $\tau = 0.084$, $\sigma_8 = 0.817$), and used CMBEASY (Doran 2005) for obtaining the respective spectra.

In Fig. 3.6, we plot a realisation of the original temperature map T (top panel), the uncorrelated temperature map $T^{\rm rec}_{\rm uncorr}$ (middle panel) and the correlated temperature map, $T^{\rm rec}_{\rm corr}$, for comparison (bottom panel). The realisations were created using the HEALPix package (Górski et al. 2005).

Note that all of what we have done works equally well for reducing the polarization map when trying to detect a secondary signal contained in the polarization data. One has to simply exchange the roles of T and P in the derivation. This was partly already done by Jaffe (2003), who used the information contained in the CMB temperature map for predicting a polarization map from it. The equivalent plot to Fig. 3.4 for this scenario is given in Fig. 3.7. The likelihood for the case of simultaneously detecting a temperature template T_τ and a polarization template P_τ is derived in

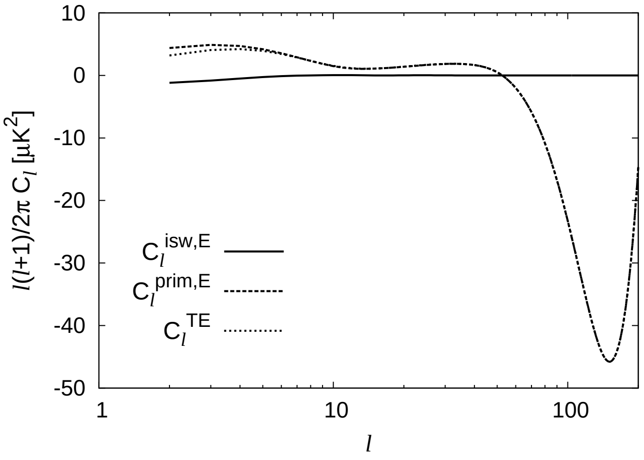


Figure 3.5: Plotted are the cross-power spectra $C_l^{\mathrm{isw},E}$, $C_l^{\mathrm{prim},E}$ and $C_l^{T,E}$. We see that T_{isw} and E are hardly correlated, $C_l^{\mathrm{isw},E} \approx 0$.

Appendix B.2.

In practice, the accuracy to which we can measure the polarization map is limited by Galactic foregrounds P_{fg}, the most important of which are synchrotron radiation and dust emission of the Milky Way. Uncertainty in the measured polarization map makes the reduction of the temperature power spectrum less efficient, because the power contained in the foreground noise, $C_l^{E_{\mathrm{fg}}}$, enhances the observed E-mode power spectrum $C_l^E \approx C_l^{E_{\mathrm{cmb}}} + C_l^{E_{\mathrm{fg}}} + C_l^{E_{\mathrm{det}}}$. In section 3.6, we estimate the improvement of the signal-to-noise ratio for an ISW measurement for such a realistic scenario.

3.5.3 Application to the ISW effect

Let us now apply our method to the ISW effect. That is, our signal template T_τ is now an ISW template which we obtain from a Wiener filter reconstruction of the LSS, as described in section 3.4. Again we assume the best-case scenario of having perfect (noiseless) LSS and CMB data. In other words, we neglect detector noise and residual Galactic foregrounds, as well as the shot-noise in the observed galaxy distribution, and assume that we have an ideal galaxy survey that covers the whole sky and goes out to a redshift of at least two. Recall that our signal template is exact in that case, $T_\tau = T_{\mathrm{s}} \equiv T_{\mathrm{isw}}$, and the residual $(T - T_{\mathrm{isw}}) \equiv T_{\mathrm{prim}}$ is simply given by the primordial CMB fluctuations, which are created at the surface of last scattering (we have ignored other secondary effects here). Since we assume T_{isw} to be uncorrelated with the primordial fluctuations T_{prim}, we can write $C_l^{T,T_\tau} \equiv C_l^{T,\mathrm{isw}} = C_l^{\mathrm{isw}}$.

The signal-to-noise ratio for the detection of the ISW signal, eq. (3.80), then reduces to

$$\left(\frac{S}{N}\right)^2_{\mathrm{pol}} = \sum_l \frac{(2l+1)\,\widehat{C}_l^{\mathrm{isw}}}{C_l^{\mathrm{prim}} - \left(C_l^{\mathrm{prim},E_{\mathrm{cmb}}}\right)^2 / C_l^{E_{\mathrm{cmb}}}} \,. \tag{3.85}$$

As explained in section 3.4, the signal-to-noise ratio depends on the specific LSS realisation in our Universe via $\widehat{C}_l^{\mathrm{isw}}$. Again, we can infer its probability distribution from the distribution of T_{isw} by using the central limit theorem for the distribution of $(S/N)^2$ and deriving the distribution for S/N

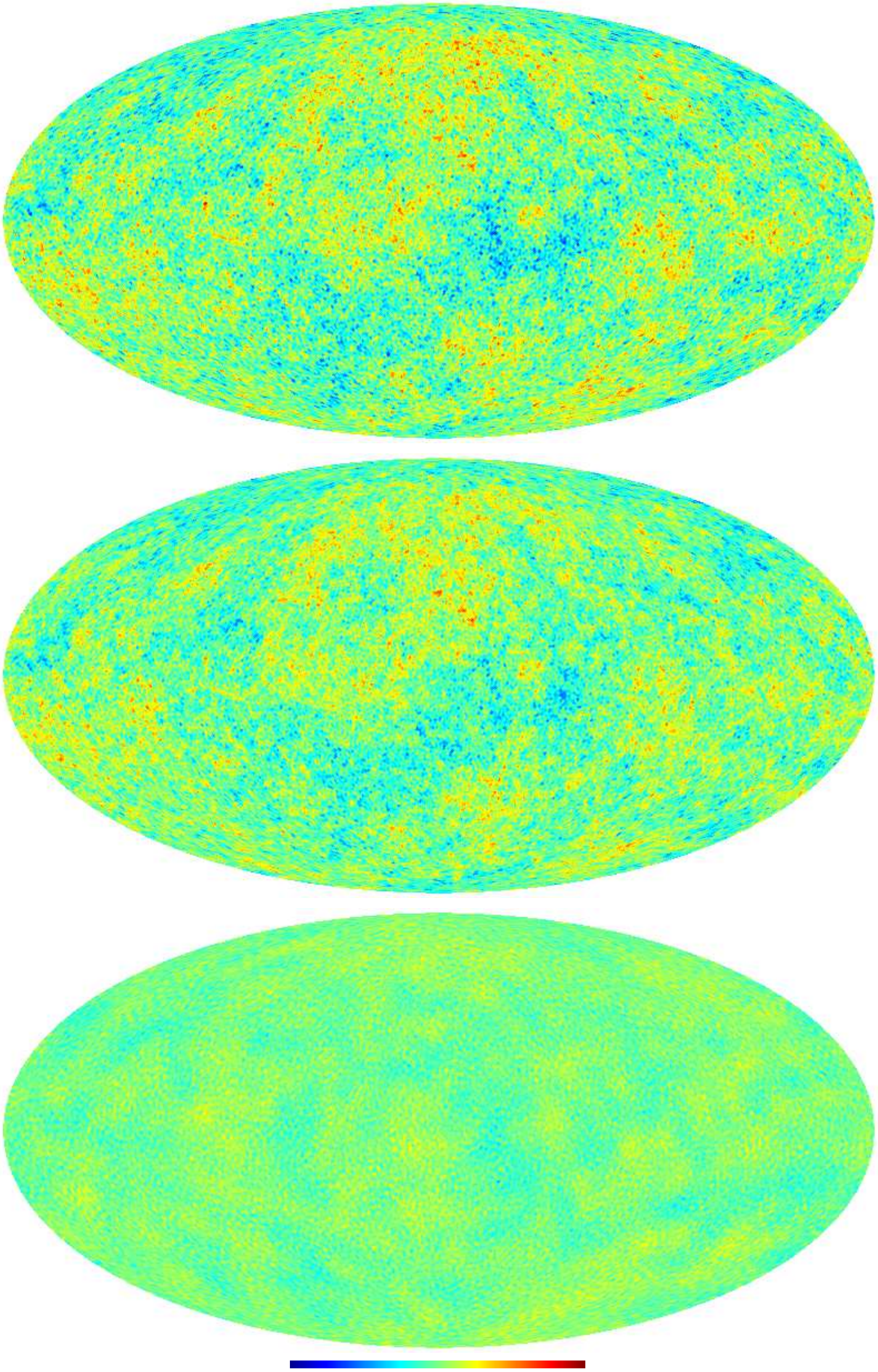


Figure 3.6: Realisation of the original CMB temperature map T (top panel), the uncorrelated temperature map $T^{\mathrm{rec}}_{\mathrm{uncorr}}$ (middle panel) and the difference between the two, $T^{\mathrm{rec}}_{\mathrm{corr}}$, for comparison (bottom panel) in μK. We have chosen the same colour range from $-500\mu K$ to $500\mu K$ for all maps.

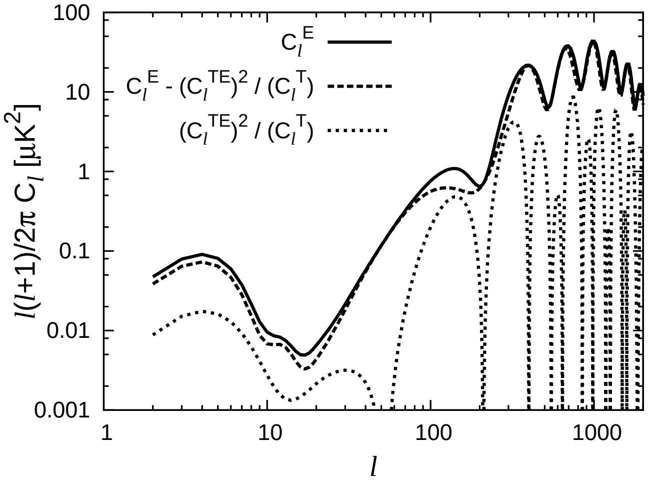


Figure 3.7: Reduction of the variance in the detection of secondary polarization signals by using the information contained in temperature data. Shown are the CMB E-mode power spectrum $C_l^{E_{\rm cmb}}$ (solid), and the template-free part of the uncorrelated E-mode power spectrum $C_l^{E_{\rm cmb}} - \left(C_l^{T,E_{\rm cmb}}\right)^2 / C_l^T$ (dashed), together with the part of the CMB power spectrum coming from the correlated part of the E-mode fluctuations which we infer from the temperature map, $\left(C_l^{T,E_{\rm cmb}}\right)^2 / C_l^T$ (dotted).

from that. We then average the signal-to-noise ratio over this probability distribution in order to compare it to the signal-to-noise ratio of the standard method, eq. (3.37), and the average signal-to-noise ratio of the optimal temperature method, eq. (3.61). Recall that the signal-to-noise ratio obtained for the standard method is given by

$$\left(\frac{S}{N}\right)_{cc}^2 = \sum_l \frac{(2l+1)\,C_l^{\rm isw}}{C_l^T + C_l^{\rm isw}}\,. \tag{3.86}$$

The cumulative signal-to-noise ratios versus the maximal multipole $l_{\rm max}$ used in the analysis are plotted in Fig. 3.8. We see that including the polarization data in the analysis increases the signal-to-noise ratio by 16 per cent as compared to the optimal temperature method, and by 23 per cent as compared to the standard method. Note that we only included the linear ISW effect in Fig. 3.8. Beyond a multipole of about $l \approx 100$, non-linear effects start to play a crucial role (Cooray 2002b), which could change the plot for $l > 100$. However, we see that for the linear ISW effect, there is hardly any contribution from such high multipoles.

Let us now look at the enhancement of the signal-to-noise ratio for shallower LSS surveys. We use the same approximation as in section 3.4, i.e., we introduce a sharp cut-off in redshift and redefine everything beyond that redshift as primordial fluctuations. We plot the redshift-dependence of the signal-to-noise ratios of the three methods in Fig. 3.9. We also plot the ratio of the signal-to-noise of the optimal polarization method with the one of the standard method (solid) and with the one of the optimal temperature method (dashed). Note that the enhancement of the signal-to-noise ratio with respect to the optimal temperature method is almost constant in redshift. This is quite clear from the fact that we have reduced the *primordial* noise with the polarization data, and neither the primordial noise nor the reduction of the latter depend on redshift. Therefore, the reduction of the noise from including polarization data is always the same, independent of how deep in redshift our

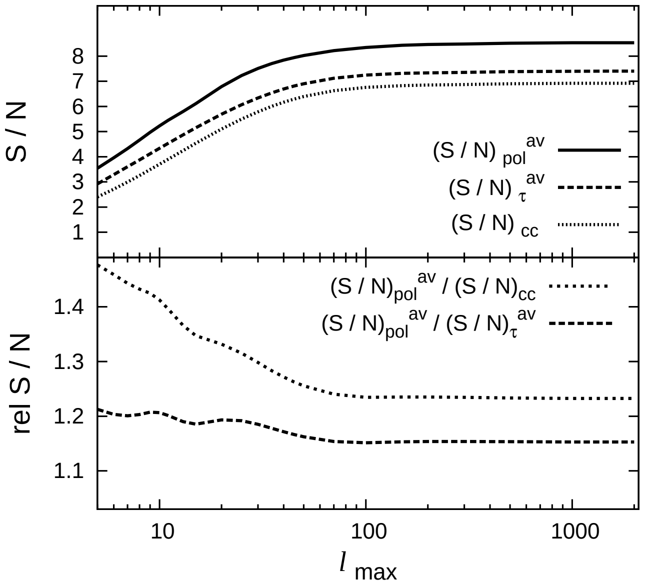


Figure 3.8: Comparison of the cumulative signal-to-noise ratios for $z_{\max} = 2$. **Top panel:** Average signal-to-noise ratio of the optimal polarization method $(S/N)^{\text{av}}_{\text{pol}}$ (solid), of the optimal temperature method $(S/N)^{\text{av}}_{\tau}$ (dashed), and signal-to-noise ratio of the standard method $(S/N)_{cc}$ (dotted) versus the maximal multipole considered in the analysis. **Bottom panel:** Ratio of the signal-to-noise of the optimal polarization method with the one of the standard method (dotted) and with the one of the optimal temperature method (dashed).

survey goes, and the signal-to-noise ratio is already significantly enhanced for currently available surveys. For example, for a maximal redshift of $z_{\max} \approx 0.3$, which is the maximal redshift for the SDSS main galaxy sample, we have a better signal-to-noise by about 16 per cent as compared to the standard method. The additional enhancement for higher redshifts of our signal-to-noise ratio with respect to the standard method comes from working conditional on the galaxy data, as we have described in detail in section 3.4.

3.6 Improvement for currently available data

Let us now find out how much subtracting the correlated temperature map, $T^{\text{rec}}_{\text{corr}}$, from the original CMB temperature map would improve ISW measurements with currently available CMB and LSS data. In chapter 4, we will compute $T^{\text{rec}}_{\text{corr}}$ and $T^{\text{rec}}_{\text{uncorr}}$ together with the covariance of the correlated temperature map, C_{corr}, for the WMAP data, and we will already use these results here. We transform the covariance matrix C_{corr} to multipole space, which gives us the non-diagonal matrix $C^{\text{corr}}_{l,m,l',m'}$ since our problem is not isotropic. This is due to the inhomogeneous noise in the polarization map and the fact that we we have to mask out the Galactic plane in the polarization data (cf. chapter 4). However, we only consider the diagonal, $C^{\text{corr}}_{l,m,l,m}$, and take the average over all multipole components m for every fixed l:

$$C^{\text{corr}}_{ll} \equiv \frac{1}{2l+1} \sum_{m} C^{\text{corr}}_{l,m,l,m} \,. \tag{3.87}$$

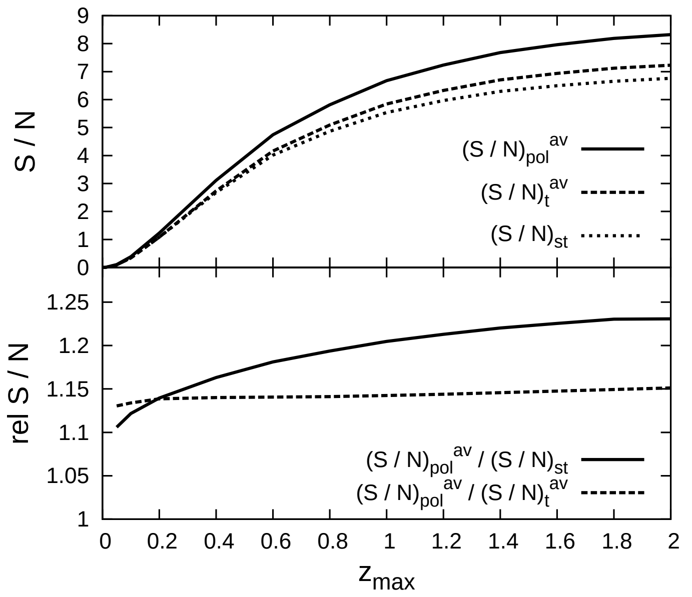


Figure 3.9: Comparison of the signal-to-noise ratios versus the maximal redshift $z_{\rm max}$ of the galaxy survey. **Top panel:** Average signal-to-noise ratio of the optimal polarization method $(S/N)^{\rm av}_{\rm pol}$ (solid), of the optimal temperature method $(S/N)^{\rm av}_{\tau}$ (dashed) and signal-to-noise ratio of the standard method $(S/N)_{cc}$ (dotted). **Bottom panel:** Ratio of the signal-to-noise of the optimal polarization method with the one of the standard method (dotted) and with the one of the optimal temperature method (dashed). We see that with polarization data included, the signal-to-noise is significantly enhanced even for low redshifts.

In the upper panel of Fig. 3.10, we plot this ‘averaged power’ of the observed $T^{\rm rec}_{\rm corr}$ together with the theoretical power spectrum of the correlated temperature map, $C_l^{T^{\rm rec}_{\rm corr}} \approx (C_l^{T,E_{\rm cmb}})^2/C_l^{E_{\rm cmb}}$, and the power in the CMB temperature fluctuations, C_l^T. The theoretical power spectrum, $(C_l^{T,E_{\rm cmb}})^2/C_l^{E_{\rm cmb}}$, is the upper limit of the reduction of the variance that we can obtain per multipole (cf. Fig. 3.4). The ratio $C_{ll}^{\rm corr}/[(C_l^{T,E_{\rm cmb}})^2/C_l^{E_{\rm cmb}}]$ is shown in the bottom panel of Fig. 3.10. For WMAP, this ratio is around 50 per cent for the lowest multipoles and falls off to a value below 1 per cent above $l \approx 10$.

We now compute the signal-to-noise ratio we would obtain with the WMAP data for our optimal polarization method and compare it to the one of the standard method. We do not create an ISW template from LSS surveys to compute the correct signal-to-noise ratio, but we rather perform a crude estimate in multipole space. To this end, we estimate the signal-to-noise ratio of the optimal polarization method for WMAP data, $(S/N)^{\rm av}_{{\rm pol},WMAP}$, by simply substituting the term $(C_l^{{\rm prim},E_{\rm cmb}})^2/C_l^{E_{\rm cmb}}$ in eq. (3.85) by $C_{ll}^{\rm corr}$. We plot the ratio of $(S/N)^{\rm av}_{{\rm pol},WMAP}$ to the signal-to-noise ratio of the standard method, $(S/N)_{cc}$, for a given depth of the LSS survey in Fig. 3.11. This is analogous to what is plotted in the bottom panel of Fig. 3.9. Note that we have again neglected the shot-noise in the galaxy distribution. We see that for galaxy surveys with $z_{\rm max} \lesssim 0.6$, we do not gain more than 5 per cent in detection significance for WMAP. With currently available data, our optimal polarization method and the standard method for ISW detection thus still yield approximately the same detection significance. Given that the standard method has been applied to all currently available LSS data sets (see Ho et al. 2008; Giannantonio et al. 2008, and references therein), we decide not to proceed and apply our method to WMAP data.

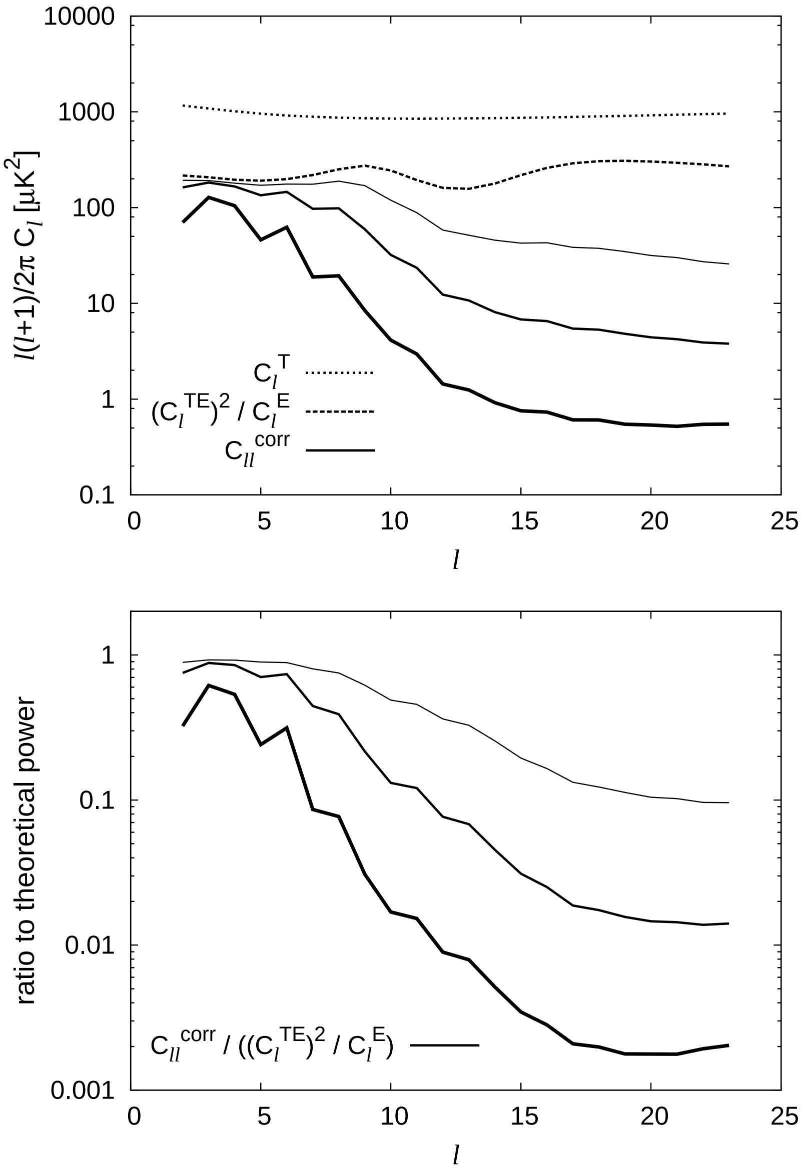


Figure 3.10: Reduction of the variance of ISW detection from including polarization data for WMAP and for 2 limiting estimates for *Planck*, which are explained in the text. **Top panel:** C_l^T (dotted), the theoretical reduction of variance per multipole, $(C_l^{T,E_{\rm cmb}})^2/C_l^{E_{\rm cmb}}$ (dashed), and the variance reduction achieved by WMAP/*Planck* $C_{ll}^{\rm corr}$ (solid). The three cases shown for $C_{ll}^{\rm corr}$ are: The estimate for *Planck* with the WMAP noise covariance scaled down by a factor of 1 per cent (upper thin line), the same for a scale-factor of 10 per cent (middle line), and the estimate for WMAP (bottom thick line). **Bottom panel:** The ratio $C_{ll}^{\rm corr}/((C_l^{T,E_{\rm cmb}})^2/C_l^{E_{\rm cmb}})$, which is roughly the ratio of the variance reduction achieved by WMAP/*Planck* over the theoretically achievable reduction of variance per multipole. The three cases shown are the same as in the top panel.

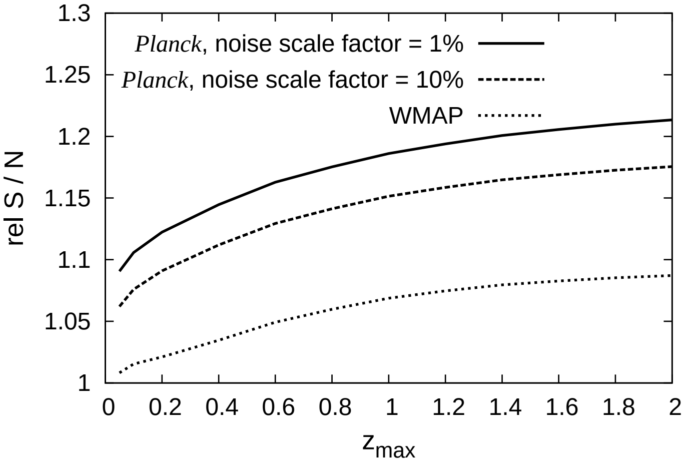


Figure 3.11: Ratio of the signal-to-noise ratio of the optimal polarization method to the one of the standard method, for WMAP (dotted line) and for the two limiting estimates for *Planck* as explained in the text: WMAP noise covariance scaled down by a factor of 10 per cent (dashed line, $\alpha = 0.7$) and by a factor of 1 per cent (solid line, $\alpha = 0.9$). For *Planck*, we will obtain an enhancement of the detection significance of at least 10 per cent, even for the more conservative estimate.

However, the *Planck* satellite (Tauber 2000), which was launched in May 2009, will provide us with much more accurate polarization measurements than WMAP. Let us therefore do a similar rough estimate of the potential improvement of ISW detection with *Planck* data. We first need to find a way to estimate C_{ll}^{corr} for *Planck*, which we attempt to do based on the polarization noise covariance matrix of WMAP. For *Planck*, the Galactic foregrounds will certainly be the limiting factor for the quality of the polarization data. We thus assume the detector noise to be negligible in comparison to the residual foregrounds. We can furthermore expect the foreground removal for *Planck* data to be more accurate than the one for WMAP, due to the broader range of frequencies covered by *Planck*. In the following, we therefore assume the covariance due to residual foregrounds for *Planck* to be between 5 and 50 per cent of the one for WMAP. For the WMAP polarization data, the foregrounds contribute about 20 per cent to the diagonal of the noise covariance matrix N_P in pixel space. In order to obtain a noise estimate for the two limiting cases of the *Planck* foreground removal, we do not change the shape of the total noise covariance, N_P, but simply scale it down with a factor of 1 per cent or 10 per cent, respectively. This will only give us a very crude estimate, of course, since the covariance matrix of the residual foregrounds should differ strongly from the detector-noise dominated covariance of WMAP. In particular, the contribution of the foregrounds to the total noise covariance matrix of WMAP should depend on the multipole, which we ignore by just downscaling the total noise covariance. Nevertheless, for the very rough estimate we are trying to obtain, this assumption should be good enough. We show the resulting power spectra and their ratio in Fig. 3.10. The ratio $C_{ll}^{\mathrm{corr}}/[(C_l^{T,E_{\mathrm{cmb}}})^2/C_l^{E_{\mathrm{cmb}}}]$ for the lowest multipoles is around 0.7 for the WMAP noise covariance scaled down by a factor of 10 per cent, and around 0.9 for the noise covariance scaled down by a factor of 1 per cent. For the higher multipoles, this ratio falls off quite rapidly, which is simply due to the fact that the WMAP polarization data contain so little information at the higher multipoles. For *Planck*, we assume that the quality of the polarization data does not notably drop until $l \approx 100$. That is, we assume a constant ratio

$$\alpha \equiv C_{ll}^{\mathrm{corr}} / [(C_l^{T,E_{\mathrm{cmb}}})^2 / C_l^{E_{\mathrm{cmb}}}] \,, \tag{3.88}$$

and we take α to be 0.7 and 0.9, for the two limiting cases described above. For these two cases, we again compute the signal-to-noise ratio of the optimal polarization method by substituting $(C_l^{\mathrm{prim},E_{\mathrm{cmb}}})^2/C_l^{E_{\mathrm{cmb}}}$ in eq. (3.85) by $\alpha\,(C_l^{\mathrm{prim},E_{\mathrm{cmb}}})^2/C_l^{E_{\mathrm{cmb}}}$, and plot its ratio to the signal-to-noise ratio of the standard method in Fig. 3.11. The improvement we obtain is already around 10 per cent for low redshift surveys of $z_{\mathrm{max}} \approx 0.3$ for the more conservative estimate. We can thus expect that the improvement of the ISW detection significance for the optimal polarization method will be at least 10 per cent for *Planck*, even with currently available LSS surveys.

3.7 Likelihood for cosmological parameters

Many of the above-mentioned cross-correlation studies, which use the standard method described in section 3.3 to detect the ISW effect, attempt to constrain cosmological parameters using the ISW. However, the biasing effect of the signal-to-noise ratio we have described in section 3.4.3 is of course also present when moving from the pure detection of the ISW to the task of constraining cosmological parameters. This problem can already be seen in the likelihood function for the cosmological parameters of the standard method in eq. (3.26). The estimator of the cross-correlation function, $\widehat{\xi}^{g,T}$, could be quite different from the theoretical prediction with the underlying parameter values, just because we are living in an unlikely realisation of the matter distribution, given the power spectrum. Then, the likelihood in eq. (3.26) would favour cosmological parameter values for which the theoretical prediction of the cross-correlation function is closer to its estimator, thus biasing the parameter estimates. Furthermore, to our knowledge, there is no straightforward way of combining the likelihood from the cross-correlation in eq. (3.26) with the likelihoods for CMB and LSS data, as e.g. given by Verde et al. (2003b), Percival et al. (2004) and Cole et al. (2005).

Usually, when combining CMB with LSS data for deriving constraints on cosmological parameters, it is assumed that the two data-sets are stochastically independent, i.e. that $\mathcal{P}(T, P, \delta_g \,|\, p) = \mathcal{P}(T, P \,|\, p)\mathcal{P}(\delta_g \,|\, p)$ (see, e.g., Tegmark et al. 2004; Spergel & et al 2007; Komatsu et al. 2009). But the ISW effect (and also other effects as, e.g., the Sunyaev-Zel'dovich-effect) introduces a small stochastic dependence of the CMB data on the LSS data. That is, instead of assuming that the joint likelihood factorises, one should consider

$$\mathcal{P}(T, P, \delta_g \,|\, p) = \mathcal{P}(T, P \,|\, \delta_g, p)\mathcal{P}(\delta_g \,|\, p) \,, \tag{3.89}$$

in which we insert eq. (A.4) for the data model given in eq. (3.18), and eq. (3.74), obtaining

$$\mathcal{P}(T, P, \delta_g \,|\, p) = \mathcal{G}(T_{\mathrm{uncorr}}^{\mathrm{rec}} - T_\tau, C_{\mathrm{uncorr}})\; \mathcal{G}(P, C_P)\; \mathcal{G}(\delta_g, R S_m R^\dagger + N_\epsilon) \,, \tag{3.90}$$

and we recall for convenience the definition of $T_{\mathrm{uncorr}}^{\mathrm{rec}}$ and C_{uncorr}, eq. (3.72),

$$\begin{aligned} T_{\mathrm{uncorr}}^{\mathrm{rec}} &\equiv T - C_{\Delta T,P} C_P^{-1} P \\ C_{\mathrm{uncorr}} &\equiv C_{\Delta T} - C_{\Delta T,P} C_P^{-1} C_{P,\Delta T} \,, \end{aligned}$$

and of T_τ, eq. (3.47),

$$T_\tau \equiv \mathcal{Q}\,\delta_m^{\mathrm{rec}}\,.$$

Note that here the detector noise T_{det} has to be included in C_{uncorr}, because we also consider higher multipoles, in which cosmic variance no longer dominates the uncertainty. Eq. (3.90) is the generic expression for the joint likelihood $\mathcal{P}(T, P, \delta_g\,|\,p)$ for the cosmological parameters p, given CMB and LSS data, consistently including the small coupling between the two data-sets introduced by the ISW effect. The quantities depending on the cosmological parameters are $C_{\Delta T}$, $C_{\Delta T,\Delta P}$, $C_{\Delta P}$, $\mathcal{Q}$, R, S_m, and, in general, N_ϵ. Multiplying the likelihood by a prior $\mathcal{P}(p)$ for the cosmological parameters, one can then sample the parameter space and derive constraints on the cosmological parameters from the posterior distribution $\mathcal{P}(p\,|\,T, P, \delta_g) \propto \mathcal{P}(T, P, \delta_g\,|\,p)\,\mathcal{P}(p)$. Note that our likelihood function remains valid if galaxy bias variations, position dependent noise, and other non-linear effects of galaxy formation are taken into account, as long as the variance of the reconstruction, $D_m \equiv \langle(\delta_m - \delta_m^{\mathrm{rec}})\,(\delta_m - \delta_m^{\mathrm{rec}})^\dagger\rangle$, is estimated consistently (see Enßlin et al. (2009) for methods to treat such complications).

Using the joint likelihood given by eq. (3.90) in parameter sampling studies, rather than assuming the likelihoods of CMB and LSS to be independent, can affect the constraints on dark-energy related parameters. If the curvature of the Universe is set to zero in parameter estimation studies, we do not expect a notable difference in Ω_Λ, since $\Omega_\Lambda = 1 - \Omega_m$ and Ω_m is well constrained by the acoustic peaks in the CMB. However, if the curvature is used as additional parameter, Ω_Λ is no longer fixed by Ω_m and thus sensitive to changes in the ISW effect. The constraints on Ω_Λ using our likelihood should then differ from the ones obtained when neglecting the coupling between the data-sets. The constraints on the EoS parameter of dark energy should change as well. Note that in order to see such a difference, we will need a LSS survey that covers enough volume relevant for the ISW effect. The ideal survey would be EUCLID, however, Ho et al. (2008) and Giannantonio et al. (2008) claim to see effects in the parameter constraints when including information from the ISW effect already when combining the currently available surveys listed in section 3.4.3.

3.8 Conclusions

Due to the obscuration by primordial CMB fluctuations, the detection of the ISW effect and other secondary effects in the CMB is a rather difficult task, and has to be performed by cross-correlating the CMB temperature fluctuations with the large-scale structure. The standard method for doing so compares the observed cross-correlation function to its theoretical prediction, which is by construction an ensemble average over all realisations of the primordial CMB fluctuations and matter distributions. Therefore, both the specific realisation of the primordial temperature fluctuations and the LSS in our Universe act as sources of noise in the detection of secondary effects in the CMB. In this chapter, we have derived methods for ISW detection which reduce both of these sources of uncertainty by working conditional on the LSS distribution and on the CMB polarization data.

First, we have presented a generic technique of how to include the knowledge of the matter distribution into ISW detection in an optimal way, thus reducing the effect of the local variance. This optimal temperature method requires a three-dimensional Wiener filter reconstruction of the LSS, including an estimator of the full uncertainty covariance matrix of the reconstruction. Note that also other reconstruction techniques that provide an estimator of the uncertainty covariance can easily be included into our method. The conditionality on the LSS data results in a dependence of the variance in the detected signal on the specific realisation of the LSS in the observed Universe. The average variance in the optimal temperature method is reduced by about 13 per cent as compared to the standard method, in the case of an ideal LSS survey. This reduction of the noise translates into an average enhancement of the signal-to-noise or detection significance by about 7 per cent for the optimal temperature method. However, note that also the signal-to-noise ratio depends on the actual realisation of the matter distribution. Therefore, even if the average enhancement of the detection significance only amounts to 7 per cent, we could be lucky (or unlucky) and live in an unlikely realisation of the matter distribution given the power spectrum, so that the enhancement in our Universe could be higher (or lower) than the average value. The fact that the standard method for ISW detection does not work conditional on the LSS distribution causes the ISW estimate to be biased by the latter. This effect becomes stronger the more unlikely the specific LSS realisation in our Universe is.

We have then presented a way of reducing the noise coming from primordial temperature fluctuations by simply subtracting from the temperature map the part which is correlated with the polarization data. When doing so, only the uncorrelated part of the temperature fluctuations contributes to the variance of the signal estimate. We calculated the achievable reduction in primordial noise for perfect (noiseless) data, and obtained a signal-to-noise ratio of up to 8.5. This corresponds to an enhancement of the signal-to-noise ratio by 16 per cent as compared to our optimal temperature method, independent of the depth of the LSS survey. In comparison to the standard method, the signal-to-noise ratio is enhanced by 23 per cent for a full-sky galaxy survey which goes out to a redshift of at least two. For the upcoming polarization data from the *Planck* mission, a very crude estimate yields an enhancement of the detection significance of at least 10 per cent. This depends strongly on the residual foregrounds present in the polarization maps. We would like to point out that in our methods there is no need to estimate the covariance matrix by Monte Carlo simulations, which safes time and increases the accuracy of the method (using 1000 Monte Carlo simulations to estimate the standard covariance matrix of the cross-correlation function only reaches an accuracy of about 5 per cent, as stated by Cabré et al. (2007)).

In order to consistently include the information encoded in the ISW effect in cosmological parameter estimation studies, we have derived the joint likelihood $P(T, P, \delta_g \,|\, p)$ for the cosmological parameters p, given CMB and LSS data, within the linear regime of structure formation. We expect small changes in the dark-energy related cosmological parameters when using this joint likelihood rather than assuming that the likelihoods of CMB and LSS data factorise.

The variance reduction achieved with the presented methods will significantly improve the detection of all kinds of secondary effects on the CMB, for which a spatial signal template can be constructed from non-CMB data. Note, however, that we have used a Gaussian approximation for the uncertainty in the signal template, which may not be optimal for effects on smaller scales such as the RS effect, the kinetic SZ effect, or gravitational lensing. The extension to non-Gaussian data models is beyond the scope of this work, but can be done using the framework introduced in section 2.3, which was developed in Enßlin et al. (2009). The work presented here stresses the importance of accurate measurements of the LSS distribution and of CMB polarization fluctuations even for signals that are not directly contained in these two data-sets. The *Planck Surveyor* mission, as well as more future experiments like PolarBeaR[10] or CMBPol[11] will soon allow us to benefit from polarization for the detection of secondary CMB signals in the way presented here.

---

[10] `http://bolo.berkeley.edu/polarbear/index.html`

[11] Baumann et al. (2008), `http://cmbpol.uchicago.edu`

Chapter 4

The axis of evil - a polarization perspective

Note: This chapter, as well as appendix C, have been submitted for publication in MNRAS (Frommert & Enßlin 2009b).

4.1 Introduction

A major assumption of modern day cosmology is the cosmological principle, which states that the Universe is homogeneous and isotropic on large scales. The observed isotropy of the Cosmic Microwave Background (CMB) is one of the strongest evidences supporting the cosmological principle.

However, in recent years, there have been claims of anomalies detected in the CMB temperature map with considerable significance, which seem to break statistical isotropy of the temperature fluctuations and thus to question the cosmological principle. Several groups (de Oliveira-Costa et al. 2004; Abramo et al. 2006; Land & Magueijo 2007; Samal et al. 2008; Rakić & Schwarz 2007) claim to have found a strong alignment between the preferred axes of the quadrupole and the octopole, which is commonly referred to as the *axis of evil*. Others (Bernui 2008; Eriksen et al. 2007; Hoftuft et al. 2009) have found a significant power asymmetry between the northern and southern ecliptic hemisphere, and some weaker anomalies have been found for the low multipoles beyond the octopole (Copi et al. 2004; Land & Magueijo 2005; Abramo et al. 2006; Pereira & Abramo 2009). However, the existence of such an isotropy breaking in the CMB temperature map is strongly under debate, and also negative results have been published (Souradeep et al. 2006; Magueijo & Sorkin 2007). The claims of the existence of a preferred direction in the CMB temperature map have led to a discussion about whether this is simply due to a chance fluctuation in the CMB temperature map, if it can be blamed on local structures or on systematics in the measurement, or whether it is actually due to a preferred direction intrinsic to our Universe (Copi et al. 2007; Dolag et al. 2005; Maturi et al. 2007; Samal et al. 2009; Groeneboom & Eriksen 2009; Morales & Sáez 2008; Vielva et al. 2007; Inoue & Silk 2007; Gao 2009; Ackerman et al. 2007; Copi et al. 2006; Schwarz et al. 2004; Hansen et al. 2004, 2009; Prunet et al. 2005; Jaffe et al. 2005, 2006; Bernui et al. 2006; Wiaux et al. 2006).

The polarization fluctuations of the CMB, just as its temperature fluctuations, have their origin in the primordial gravitational potential. The polarization should thus exhibit similar peculiarities as the temperature, provided they are due to some preferred direction intrinsic to the geometry of the primordial Universe. Note that this is not generic to every model creating anomalies in the

map	explanation	eq.	multipole	(l, b)	σ
P_{corr}	”$T \rightarrow P$”	(4.13)	quadr	$(-117^\circ, 60^\circ)$	-
			oct	$(-124^\circ, 66^\circ)$	-
$P^{\mathrm{rec}}_{\mathrm{uncorr}}$	”$P - P_{\mathrm{corr}}$”	(4.16)	quadr	$(-79^\circ, 36^\circ)$	42°
			oct	$(-17^\circ, 0^\circ)$	48°
$T^{\mathrm{rec}}_{\mathrm{corr}}$	”$P \rightarrow T$”	(4.9)	quadr	$(-73^\circ, 42^\circ)$	42°
			oct	$(-17^\circ, -19^\circ)$	37°
$T^{\mathrm{rec}}_{\mathrm{uncorr}}$	”$T - T^{\mathrm{rec}}_{\mathrm{corr}}$”	(4.12)	quadr	$(-107^\circ, 42^\circ)$	33°
			oct	$(-112^\circ, 54^\circ)$	10°

Table 4.1: Axes and their uncertainties for the four different maps in Galactic coordinates. The large errors are due to the effects of the mask, residual foregrounds and the detector noise in the WMAP polarization data.

temperature map. For example, if the peculiarities in the temperature maps are due to a secondary effect on the CMB such as the integrated Sachs-Wolfe effect, we would not expect them to be present in the polarization maps (Dvorkin et al. 2008). The search for anomalies in the CMB polarization map is still in its initial stage, due to the high noise-level in the available full-sky polarization map from the *Wilkinson Microwave Anisotropy Probe* (WMAP). Souradeep et al. (2006) have found some evidence for anisotropies in the WMAP polarization data using the method proposed in Basak et al. (2006). However, they state that the anisotropies are likely due to observational artifacts such as foreground residuals, and that further work is required in order to confirm a possible cosmic origin.

Given that the polarization map is correlated with the temperature map, it is not a statistically independent probe of the anomalies which have been found in the temperature map. If the observed anomalies were due to a chance fluctuation in the temperature map, this chance fluctuation could also be present in the polarization map, due to the correlation between the two. Therefore, we split the WMAP polarization map into a part correlated with the temperature map, P_{corr}, and a part uncorrelated with the latter, $P^{\mathrm{rec}}_{\mathrm{uncorr}}$. We obtain the part of the polarization map which is correlated with the temperature map by simply translating the temperature map into a polarization map, using their cross-correlation. The part of the polarization map which is uncorrelated with the temperature map serves as a truly independent probe of the above-mentioned anomalies. Chance fluctuations in the temperature maps do not affect the uncorrelated polarization map, so that a detection of the anomalies in the latter would be a hint to an actual cosmological origin of them. Note, though, that this does not have the power to exclude residual foregrounds or systematics as potential origins for the anomalies. Similarly, we split the WMAP temperature map into a part correlated with the polarization map, $T^{\mathrm{rec}}_{\mathrm{corr}}$, and an uncorrelated map, $T^{\mathrm{rec}}_{\mathrm{uncorr}}$. If the anomalies detected in the CMB temperature map are of genuine cosmological origin, they should be present in the correlated and the uncorrelated parts of both the temperature and polarization map. For convenience, the four resulting maps are summarised and briefly described in Table 4.1.

We focus on using the uncorrelated polarization map to probe the axis of evil. In order to define the preferred axis of the multipoles, we use a statistic proposed by de Oliveira-Costa et al. (2004), which is the axis around which the angular momentum dispersion is maximised for a given multipole l. We note that we have to mask out about 25 per cent of the sky in the WMAP polarization data in order to reduce Galactic foregrounds. Furthermore, the polarization data are highly contaminated by detector noise and residual foregrounds even outside the mask. We therefore perform a Wiener filtering of the polarization data before determining the preferred axes, in order to reduce the noise contained in the maps. However, we still expect a large uncertainty in the axes, which we obtain by running Monte Carlo (MC) simulations conditional on the data. The uncertainty in our axes amounts to $\sigma \approx 45°$.

We find that, for all four of the maps, the preferred axes of the quadrupole all point in the same direction, within our measurement precision. However, the preferred axis of the octopole of the uncorrelated polarization map does not align with the one of the quadrupole. The same holds for the correlated temperature map. In order to assess our result, we ask the following question. We take the axes measured in the temperature map as given, and assume that the axes of the uncorrelated polarization map are distributed isotropically and independently of each other. We then ask how likely it is that at least one of these axes lies such that the axis of the temperature map lies inside its 1σ region. This probability amounts to about 50 per cent for currently available polarization data. This high probability is due to the large uncertainties we have in the axes of the uncorrelated polarization map. The main contribution of this uncertainty comes from the high noise-level in the polarization data rather than from the mask. We can therefore hope that the *Planck* polarization data (Tauber 2000) will yield much stronger constraints on the axes than the WMAP data.

Note that our approach to probing the axis of evil in polarization is phenomenological, since not all theoretical models of the primordial Universe exhibiting anomalies in the CMB temperature map show the same behaviour in the uncorrelated polarization map. We outline a more thorough analysis, taking into account the predictions of the specific models for the uncorrelated polarization map, in the conclusions of this chapter.

This chapter is organised as follows. In sections 4.2 and 4.3, we explain in detail the splitting of the WMAP temperature and polarization maps, respectively. Section 4.4 is devoted to determining the preferred axes for the quadrupole and octopole for our four maps. We conclude in section 4.5.

4.2 Splitting of the temperature map

In this section, we split the WMAP temperature map into a part correlated with the WMAP polarization map, $T^{\mathrm{rec}}_{\mathrm{corr}}$, and a part which is not, $T^{\mathrm{rec}}_{\mathrm{uncorr}}$. To this end, we translate the polarization map into the correlated part of the temperature map, using the cross-correlation between the two. However, as we have already mentioned in the introduction, the WMAP polarization data are highly contaminated by detector noise and Galactic foregrounds. The observed polarization map we use is the linear combination of the maps of the Ka, Q, and V frequency bands (corresponding to 33, 41, and 61 GHz), which is used for determining the low-l polarization likelihood in the 5 year WMAP likelihood code (Hinshaw et al. 2009). By using the linear combination of the maps, we combine the information from different frequency bands rather than using only the information contained in a particular band. Therefore, the linear combination is less contaminated by noise than the original maps per frequency band. We use the P06 mask (Page et al. 2007) to mask out the Galactic plane in the polarization map. The linear combination maps for the Stokes Q and U parameters are shown in Fig. 4.1 in Galactic coordinates.

In order to reduce the noise level, we perform a Wiener filtering of the observed polarization map before translating it into the part of the temperature map which is correlated with the polarization data. Similarly, we will perform a Wiener filtering of the part of the polarization map which is uncorrelated with the temperature map, as we will describe in detail later on. Our data model for

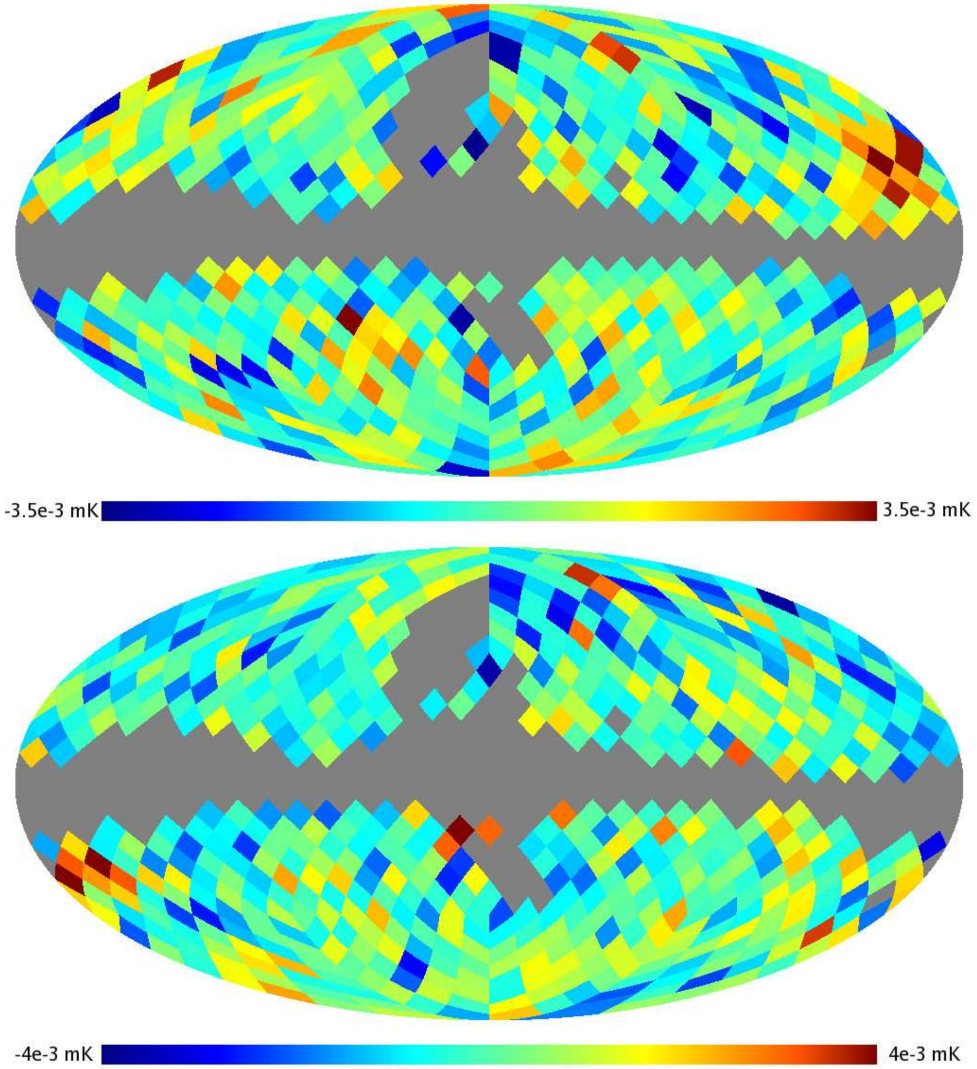

-3.5e-3 mK
3.5e-3 mK
-4e-3 mK
4e-3 mK

Figure 4.1: Observed polarization maps (linear combination of Ka, Q, and V band maps). Stokes Q map (top panel) and Stokes U map (bottom panel).

the observed polarization map P, which contains the Stokes Q and U maps shown in Fig. 4.1, is

$$P \equiv \begin{pmatrix} Q \\ U \end{pmatrix} \equiv W\left(P_{\mathrm{cmb}} + P_{\mathrm{det}} + P_{\mathrm{fg}}\right) . \tag{4.1}$$

Here, P_{cmb} is the intrinsic CMB polarization, P_{det} and P_{fg} denote the detector noise and residual foregrounds, respectively, and we have introduced the window W in order to describe the mask.

Let us define the signal covariance matrix of the CMB polarization given the cosmological parameters p,

$$S_P \equiv \langle P_{\mathrm{cmb}} P_{\mathrm{cmb}}{}^{\dagger} \rangle_{\mathcal{P}(P_{\mathrm{cmb}}\,|\,p)} , \tag{4.2}$$

and the noise covariance matrices for the detector noise and the residual foregrounds:

$$\begin{aligned} N_{\mathrm{det}} &\equiv \langle P_{\mathrm{det}} P_{\mathrm{det}}{}^{\dagger} \rangle_{\mathcal{P}(P_{\mathrm{det}})} , \\ N_{\mathrm{fg}} &\equiv \langle P_{\mathrm{fg}} P_{\mathrm{fg}}{}^{\dagger} \rangle_{\mathcal{P}(P_{\mathrm{fg}})} . \end{aligned} \tag{4.3}$$

The signal power spectrum (and thus S_P) has been computed using CMBEASY (Doran 2005) for the Maximum Likelihood cosmological model from Dunkley et al. (2009): $\{\Omega_b h^2 = 0.0227, \Omega_\Lambda = 0.751, h = 0.724, \tau = 0.089, n_s = 0.961, \sigma_8 = 0.787\}$.

In order to derive the Wiener filter for P, let us define the noise,

$$n \equiv W\left(P_{\mathrm{det}} + P_{\mathrm{fg}}\right) , \tag{4.4}$$

for which the noise covariance is then

$$N_P \equiv \langle n\, n^{\dagger} \rangle_{\mathcal{P}(n)} = W(N_{\mathrm{det}} + N_{\mathrm{fg}})W^{\dagger} , \tag{4.5}$$

where we have assumed that P_{det} and P_{fg} are uncorrelated. We take the total noise covariance, N_P, for the observed polarization map from the WMAP code. We further identify P_{cmb} with the signal s, the mask W with the response R, and P with the data d. With these definitions, we have translated our data model, eq. (4.1), into the one given in eq. (2.16). If we assume the noise n and the signal P_{cmb} to be Gaussian distributed[1], we therefore obtain the posterior distribution for the signal

$$\mathcal{P}(P_{\mathrm{cmb}}\,|\,P, p) = \mathcal{G}\left(P_{\mathrm{cmb}} - P^{\mathrm{rec}}_{\mathrm{cmb}}, D_p\right) , \tag{4.6}$$

with

$$P^{\mathrm{rec}}_{\mathrm{cmb}} \equiv (S_P^{-1} + W^{\dagger} N_P^{-1} W)^{-1} W^{\dagger} N_P^{-1} P , \tag{4.7}$$

which is the Wiener reconstruction of the polarization map, and

$$D_p \equiv (S_P^{-1} + W^{\dagger} N_P^{-1} W)^{-1} , \tag{4.8}$$

which denotes the Wiener variance. We show the Stokes Q and U maps of the Wiener filtered polarization map $P^{\mathrm{rec}}_{\mathrm{cmb}}$ in the top panels of Fig. 4.2 and Fig. 4.3, respectively. Note that only the low l modes survive the Wiener filtering, whereas the higher l modes are strongly suppressed due to the high noise-level they contain.

---

[1] The assumption of Gaussianity holds well for the detector noise P_{det} and the signal P_{cmb}. For the residual Galactic foregrounds, this assumption is probably less accurate.

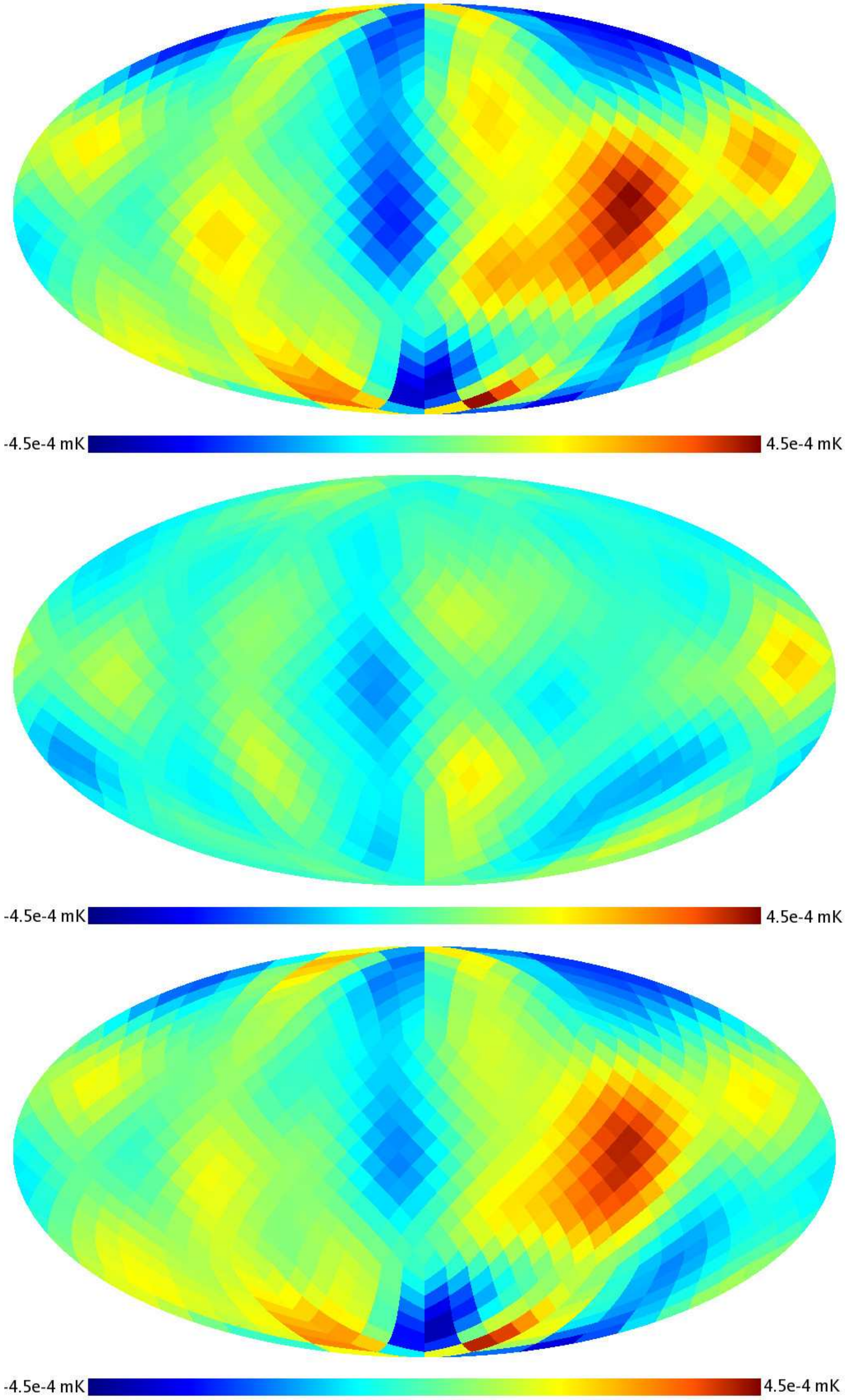


Figure 4.2: Stokes Q part of the following polarization maps: **Top panel:** Wiener filtered polarization map, $P^{\mathrm{rec}}_{\mathrm{cmb}}$. **Middle panel:** Part of the polarization map correlated with the temperature map, P_{corr}. **Bottom panel:** Part of the polarization map uncorrelated with the temperature map, $P^{\mathrm{rec}}_{\mathrm{uncorr}}$. The colour scale is the same in all maps.

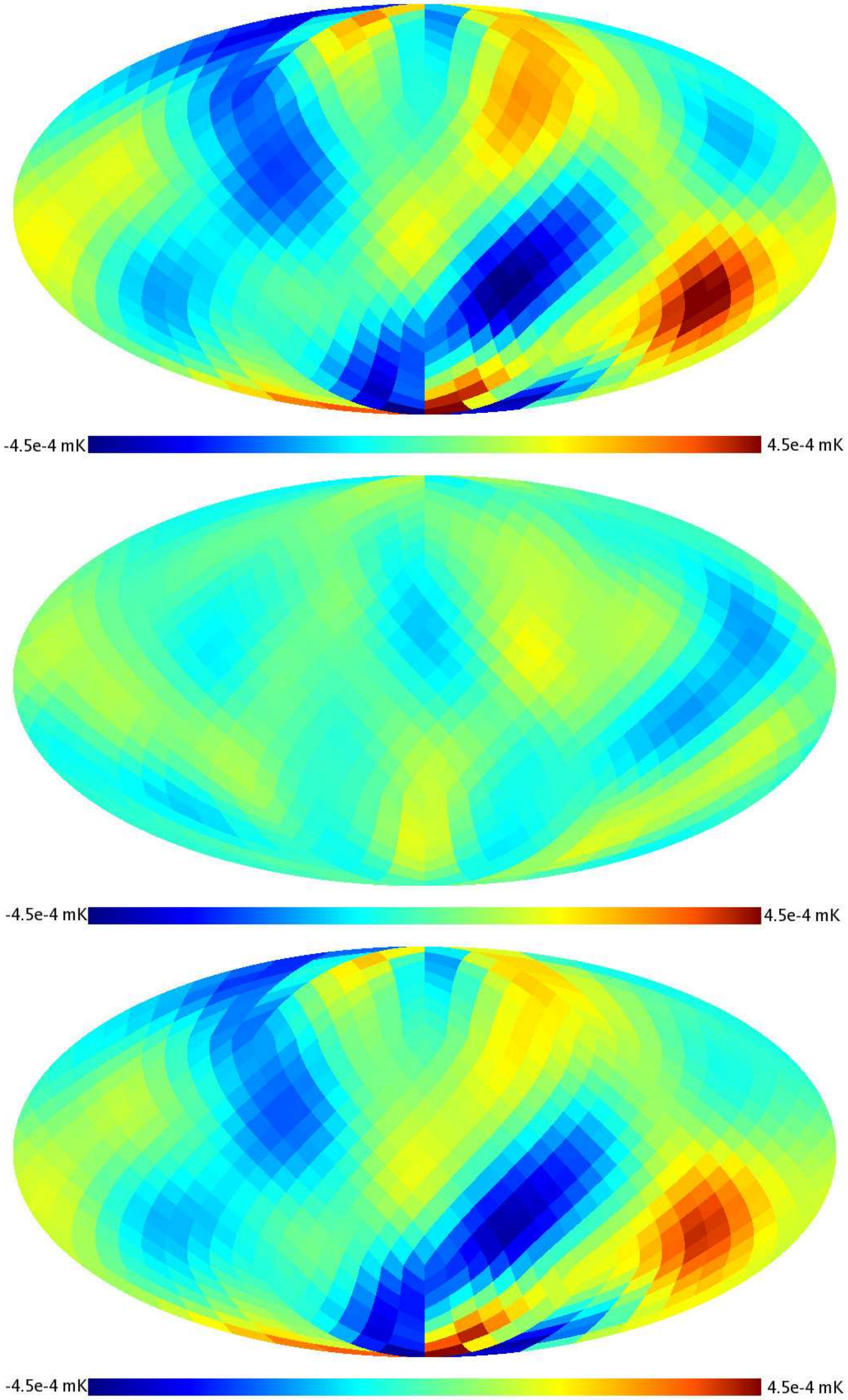


Figure 4.3: Stokes U part of the following polarization maps: **Top panel:** Wiener filtered polarization map, $P^{\mathrm{rec}}_{\mathrm{cmb}}$. **Middle panel:** Part of the polarization map correlated with the temperature map, P_{corr}. **Bottom panel:** Part of the polarization map uncorrelated with the temperature map, $P^{\mathrm{rec}}_{\mathrm{uncorr}}$. The colour scale is the same in all maps.

We now split the WMAP temperature map into a part correlated with the polarization map, $T^{\rm rec}_{\rm corr}$, and a part uncorrelated with the latter, $P^{\rm rec}_{\rm uncorr}$. We use the Wiener filtered polarization map $P^{\rm rec}_{\rm cmb}$, which is of resolution NSIDE=8, and the internal linear combination (ILC) temperature map (Gold et al. 2009), which we have smoothed with a Gaussian beam of FWHM=18.3° and downgraded to the same resolution. Among the different WMAP temperature maps, the ILC is the one for which the alignment of the low multipoles is least contaminated by Galactic foregrounds (Gruppuso & Burigana 2009). When working on large scales, we can safely neglect the detector noise in the temperature data (Afshordi 2004). Furthermore, we decide to neglect residual foregrounds in the temperature map.

We translate the Wiener filtered polarization map, $P^{\rm rec}_{\rm cmb}$, into the correlated part of the temperature map, using the cross-correlation between the two:

$$T^{\rm rec}_{\rm corr} \equiv S_{T,P}\, S_P^{-1} P^{\rm rec}_{\rm cmb}\,, \tag{4.9}$$

where the signal covariance matrices given the cosmological parameters, p, are defined as

$$\begin{aligned} S_{P,T} &\equiv \langle P_{\rm cmb} T^{\dagger} \rangle_{\mathcal{P}(T,P_{\rm cmb}\,|\,p)}\,, \qquad (4.10)\\ S_{T} &\equiv \langle T\, T^{\dagger} \rangle_{\mathcal{P}(T\,|\,p)}\,. \qquad (4.11) \end{aligned}$$

The uncorrelated temperature map $T^{\rm rec}_{\rm uncorr}$ is then obtained by simply subtracting $T^{\rm rec}_{\rm corr}$ from T:

$$T^{\rm rec}_{\rm uncorr} \equiv T - T^{\rm rec}_{\rm corr}\,. \tag{4.12}$$

In Appendix C.1, we prove that $T^{\rm rec}_{\rm corr}$ and $T^{\rm rec}_{\rm uncorr}$ are indeed uncorrelated, and that the definitions of $T^{\rm rec}_{\rm corr}$ in eq. (4.9) and in eq. (3.73) are equivalent. In other words, the splitting of the CMB temperature map performed here is the same splitting that we have already used in chapter 3 in order to reduce the noise in ISW measurements.

We plot T, $T^{\rm rec}_{\rm corr}$, and $T^{\rm rec}_{\rm uncorr}$ in the top, middle, and bottom panel of Fig. 4.4, respectively. Let us first concentrate on $T^{\rm rec}_{\rm corr}$, and try to assess whether some of its structures could come from Galactic foregrounds rather than being intrinsic CMB fluctuations. Note that this is just meant to be a quick glance on what we can immediately pick out by eye. Comparing $T^{\rm rec}_{\rm corr}$ with the overview over the Galactic foregrounds published in Hinshaw et al. (2007), Fig. 7, makes us suspect that the warm region in the middle of the northern hemisphere might be associated with the North Galactic Spur. A part of this region is already masked out, but it is well possible that the mask should be bigger in order to better mask out this foreground. One might also think that the big red blob on the right hand side of $T^{\rm rec}_{\rm corr}$, close to the Galactic plane, could be due to the Gum Nebula. However, plotting the two maps on top of each other reveals that the Gum Nebula lies further to the East than our red blob. Therefore we exclude that the blob comes from that particular foreground.

Let us now compare the three maps T, $T^{\rm rec}_{\rm corr}$, and $T^{\rm rec}_{\rm uncorr}$. In the northern Galactic hemisphere, all maps look quite similar, apart from the hot region around the North Galactic Spur, which is more prominent in $T^{\rm rec}_{\rm corr}$ than in the other two maps, and which we have already commented on. However, in the western part of the southern hemisphere, we obtain a strong deviation of $T^{\rm rec}_{\rm corr}$ from the ILC map. In fact, the features in $T^{\rm rec}_{\rm corr}$ have the opposite sign to the structures in the ILC map. This enhances the amplitudes of the features in the western part of the southern hemisphere in $T^{\rm rec}_{\rm uncorr}$ as compared to the ILC map. In particular, the so-called *cold spot*, which has been found to have non-Gaussian characteristics by Vielva et al. (2004), turns out to be even colder in $T^{\rm rec}_{\rm uncorr}$ than in the

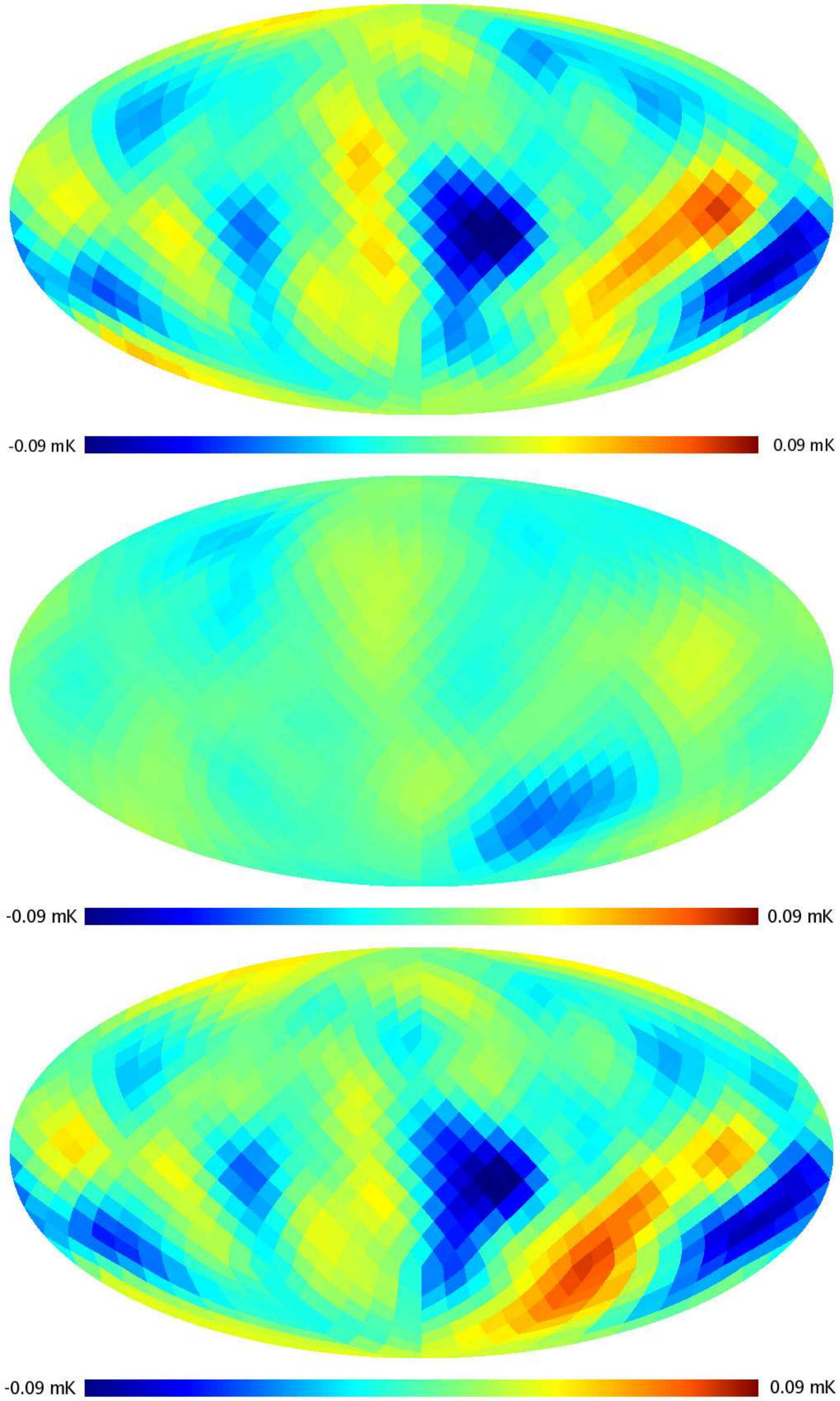


Figure 4.4: **Top panel:** ILC map, smoothed with a beam of 18.3° and downgraded to a resolution of NSIDE=8. **Middle panel:** Part of the temperature map which is correlated with the polarization map, $T^{\rm rec}_{\rm corr}$. **Bottom panel:** Part of the temperature map which is uncorrelated with the polarization map, $T^{\rm rec}_{\rm uncorr}$. The colour scale is the same in all maps.

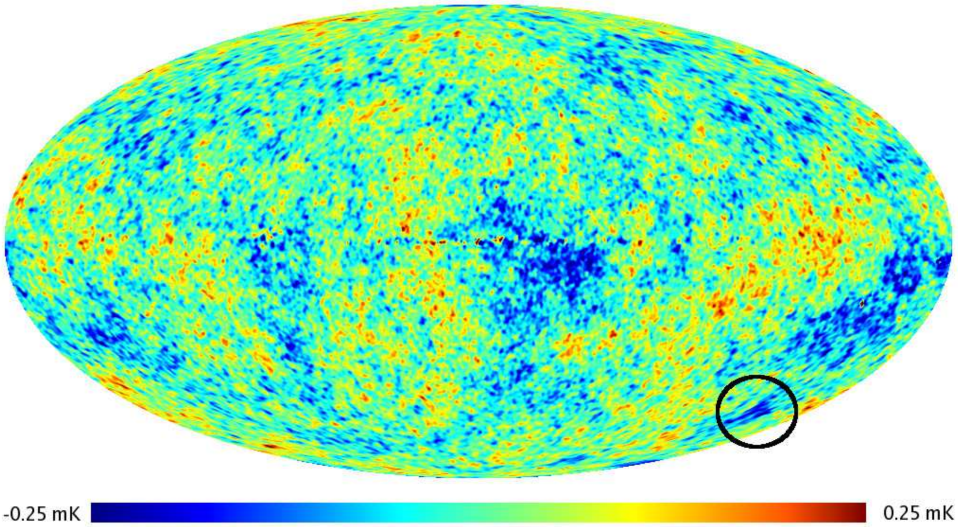


Figure 4.5: The *cold spot*, which has been found to have non-Gaussian characteristics, is marked in the ILC map shown here by a black circle.

ILC map. The cold spot, which we mark in the ILC map in Fig. 4.5 by a black circle, has later been confirmed to have non-Gaussian characteristics by many others (see, e.g. , Martínez-González et al. 2006; Cruz et al. 2006; Naselsky et al. 2007). It would be interesting to redo the above-mentioned analyses of the cold spot with the high-resolution version of $T^{\mathrm{rec}}_{\mathrm{uncorr}}$, in order to see whether the significance of the non-Gaussian features is even higher in that map. A thorough analysis of the characteristics of the cold spot is beyond the scope of this work, though, and we leave this exciting question for future work. Lastly, we notice that on the large scales we are looking at, we have much stronger deviations of the temperature towards the cold end of the temperature spectrum than towards the warm end, for all three of the maps.

4.3 Splitting of the polarization map

We now split the WMAP polarization map into a part correlated with the WMAP temperature map, P_{corr}, and a part uncorrelated with that, $P^{\mathrm{rec}}_{\mathrm{uncorr}}$. As before, we obtain the correlated polarization map by simply translating the temperature map into a polarization map:

$$P_{\mathrm{corr}} \equiv S_{P,T} S_T^{-1} T \, , \tag{4.13}$$

The Stokes Q and U maps of P_{corr} are shown in the middle panels of Fig. 4.2 and Fig. 4.3, respectively.

In order to obtain the uncorrelated map, we would like to subtract P_{corr} from P_{cmb}:

$$P_{\mathrm{uncorr}} \equiv P_{\mathrm{cmb}} - P_{\mathrm{corr}} \, . \tag{4.14}$$

However, we do not know P_{cmb} because we only observe P, which is highly contaminated by noise. Subtracting P_{corr} from the Wiener filtered polarization map, $P^{\mathrm{rec}}_{\mathrm{cmb}}$, does not result in uncorrelated maps. We therefore subtract $WP_{\rm corr}$ from the observed polarization map, P:

$$\begin{aligned} P^{\rm raw}_{\rm uncorr} &\equiv P - WP_{\rm corr} \\ &= WP_{\rm uncorr} + n\,, \end{aligned} \tag{4.15}$$

where the noise n is the same as in section 4.2. We then compute the Wiener reconstruction of the signal $P_{\rm uncorr}$, with the data being $P^{\rm raw}_{\rm uncorr}$:

$$P^{\rm rec}_{\rm uncorr} = [(S_P - S_{P,T}S_T^{-1}S_{T,P})^{-1} + W^\dagger N_P^{-1}W]^{-1}W^\dagger N_P^{-1}P^{\rm raw}_{\rm uncorr}\,. \tag{4.16}$$

Here, we have used the signal covariance

$$\begin{aligned} &\langle P_{\rm uncorr}P_{\rm uncorr}{}^\dagger\rangle_{\mathcal{P}(P_{\rm cmb},T\,|\,p)} \\ = \;& \langle P_{\rm cmb}P_{\rm cmb}{}^\dagger\rangle - \langle P_{\rm cmb}T^\dagger\rangle S_T^{-1}S_{T,P} \\ & -S_{P,T}S_T^{-1}\langle TP_{\rm cmb}{}^\dagger\rangle + S_{P,T}S_T^{-1}\langle TT^\dagger\rangle S_T^{-1}S_{T,P} \\ = \;& S_P - S_{P,T}S_T^{-1}S_{T,P}\,. \end{aligned} \tag{4.17}$$

$P^{\rm rec}_{\rm uncorr}$ given in eq. (4.16) is uncorrelated with $P_{\rm corr}$, as we prove in Appendix C.2. The posterior of $P_{\rm uncorr}$ is given by

$$\mathcal{P}(P_{\rm uncorr}\,|\,T,P,p) = \mathcal{G}\left(P_{\rm uncorr} - P^{\rm rec}_{\rm uncorr}, D_{\rm uncorr}\right)\,, \tag{4.18}$$

with the Wiener variance

$$D_{\rm uncorr} \equiv [(S_P - S_{P,T}S_T^{-1}S_{T,P})^{-1} + W^\dagger N_P^{-1}W]^{-1}\,. \tag{4.19}$$

Note that here we use the index "uncorr" to denote the uncorrelated polarization map, whereas in section 3.5 this index referred to the uncorrelated temperature map.

We show the Stokes Q and U maps of the uncorrelated polarization map, $P^{\rm rec}_{\rm uncorr}$, in the bottom panels of Fig. 4.2 and Fig. 4.3, respectively. Note that the symbols for the correlated and uncorrelated parts of temperature and polarization maps are listed and briefly explained in Table 4.1.

4.4 The axis of evil

We now search for the axis of evil in the four maps $P_{\rm corr}$, $P^{\rm rec}_{\rm uncorr}$, $T^{\rm rec}_{\rm corr}$, and $T^{\rm rec}_{\rm uncorr}$. Note that $P_{\rm corr}$ and $T^{\rm rec}_{\rm corr}$ have of course the same axes as the original temperature and polarization maps, T and $P^{\rm rec}_{\rm cmb}$, respectively. To define the preferred axis, we use a statistic proposed by de Oliveira-Costa et al. (2004), which has been introduced in order to quantify the preferred direction that can be picked out in the smoothed temperature map by eye. When looking at the smoothed ILC map in Fig. 4.4, most of the hot and cold blobs seem to be lying on the same plane. The quadrupole and octopole extracted from the ILC map show the same behaviour (see, e.g., de Oliveira-Costa et al. 2004), and the planes are roughly the same for the two multipoles. In order to quantify this alignment, de Oliveira-Costa et al. (2004) came up with the following statistic. The temperature maps are expanded into spherical harmonics, which are eigenfunctions of the square and the z-component of the angular momentum operator $\boldsymbol{L}$:

$$T(\hat{\boldsymbol{n}}) = \sum_l T_l(\hat{\boldsymbol{n}}) \equiv \sum_{l,m} a^T_{lm} Y_{lm}(\hat{\boldsymbol{n}})\,. \tag{4.20}$$

Then, for every multipole l, one determines the z-axis $\hat{\boldsymbol{n}}$ for which the expectation value of the z-component of $\boldsymbol{L}$, $\hat{\boldsymbol{n}} \cdot \boldsymbol{L}$, is maximised:

$$\langle T_l \,|\, (\hat{\boldsymbol{n}} \cdot \boldsymbol{L})^2 \,|\, T_l \rangle = \sum_m m^2 \,|\, a^T_{lm}(\hat{\boldsymbol{n}}) \,|^{\,2}\,, \tag{4.21}$$

Here, $a^T_{lm}(\hat{\boldsymbol{n}})$ denotes the spherical harmonic coefficient a^T_{lm} obtained in a coordinate system with the z-axis pointing in $\hat{\boldsymbol{n}}$-direction. We determine the axis $\hat{\boldsymbol{n}}$ by simply rotating the z-axis into every pixel centre and checking for the maximum, which is well feasible at our resolution. Neighbouring pixel centres in our map differ by approximately 7°, but we will soon see that the uncertainties in our axes are so large that it is sufficient to check only the pixel centres as potential z-axes. We have done the same exercise allowing the axes to point to all pixel centres of NSIDE=16 instead of NSIDE=8, and our results are robust under this change.

As we have already mentioned, the mask, residual foregrounds and detector noise in the polarization data will result in an uncertainty in the preferred axes. The posterior distribution of the real CMB polarization map, $P_{\rm cmb}$, given the one we observe, P, is given by eq. (4.6). $P_{\rm cmb}$ fluctuates around our Wiener reconstruction, $P^{\rm rec}_{\rm cmb}$, with the Wiener variance D_P.

In order to obtain the uncertainties in the axes of $T^{\rm rec}_{\rm corr}$ and $T^{\rm rec}_{\rm uncorr}$, we have run Monte Carlo (MC) simulations, drawing realisations of $P_{\rm cmb}$ from its posterior distribution. From these, we obtain realisations of

$$\begin{aligned} T_{\rm corr} &\equiv S_{T,P}\, S_P^{-1} P_{\rm cmb}\,, \\ T_{\rm uncorr} &\equiv T - T_{\rm corr}\,, \end{aligned} \tag{4.22}$$

for which we then determine the preferred axes. The uncertainty in the axes of $P^{\rm rec}_{\rm uncorr}$ is obtained similarly, using the posterior distribution of $P_{\rm uncorr}$ given in eq. (4.18). Note that T and thus $P_{\rm corr}$ are assumed to have no contributions from residual foregrounds or detector noise, and thus no uncertainty in the preferred axes.

For drawing realisations from the probability distribution in eq. (4.6), we have computed the Wiener variance D_P given in eq. (4.8). We have then computed the Cholesky factorisation L of D_P, which is a particular form of the square-root of a positive definite matrix:

$$D_P = LL^\dagger. \tag{4.23}$$

In order to obtain our realisation, $P_{\rm cmb}$, we apply L to a map n_w of white noise, i.e. a map where the temperature at every pixel is independently drawn from a Gaussian distribution with unit variance, and add the mean value $P^{\rm rec}_{\rm cmb}$: $P_{\rm cmb} \equiv L\, n_w + P^{\rm rec}_{\rm cmb}$. This results in a map which is drawn from the distribution in eq. (4.6), as one can easily see:

$$\begin{aligned} &\langle (P_{\rm cmb} - P^{\rm rec}_{\rm cmb})(P_{\rm cmb} - P^{\rm rec}_{\rm cmb})^\dagger \rangle_{\mathcal{P}(n_w)} \\ = \;& L\, \langle n_w n_w^\dagger \rangle_{\mathcal{P}(n_w)}\, L^\dagger = LL^\dagger = D_P. \end{aligned} \tag{4.24}$$

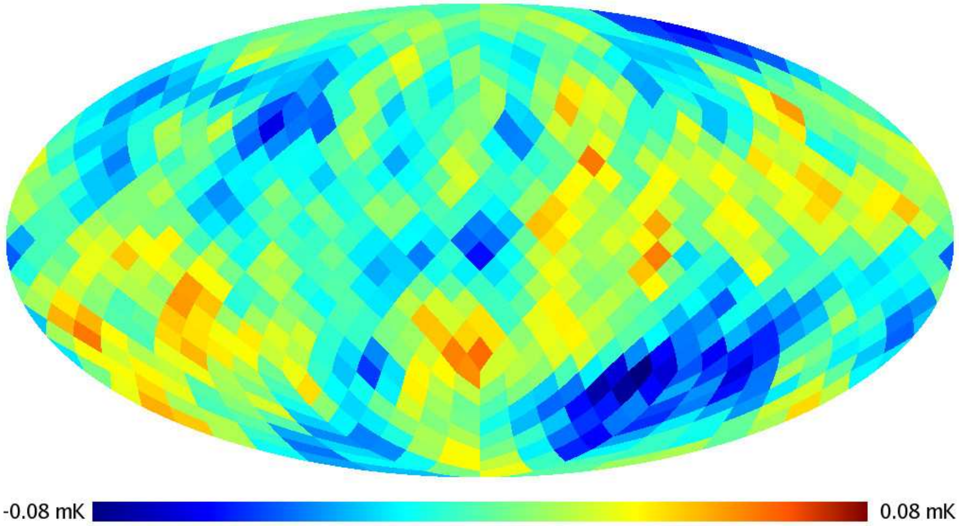


Figure 4.6: Wiener realisation of T_{corr}

An example of a Wiener realisation of T_{corr} in shown in Fig. 4.6.[2]

We plot the axes and their uncertainties for the different maps in Figs 4.7 – 4.10. Both ends of every axis are marked by a cross in the maps, and the colour coding counts how many times the preferred axis came to lie on the respective pixels in 5000 MC samples.

All axes and their standard deviations σ, which we obtained from the MC simulations, are summarised in Table 4.1. For P_{corr}, and thus the ILC map, we reproduce the results from de Oliveira-Costa et al. (2004) within our measurement precision: the axes of the quadrupole and the octopole of P_{corr} point in the same direction, which is roughly $(l, b) \approx (-120°, 63°)$, where l and b denote Galactic longitude and latitude, respectively (de Oliveira-Costa et al. (2004) found $(l, b) \approx (-110°, 60°)$). For $T^{\text{rec}}_{\text{uncorr}}$, again both axes point in the same direction as the axes of P_{corr} within our measurement precision.

For $P^{\text{rec}}_{\text{uncorr}}$, the preferred axis of the quadrupole has an angular distance to the average axis of the ILC map of 37°. That means that the latter lies inside its 1σ region. The same holds for $T^{\text{rec}}_{\text{corr}}$ (and thus $P^{\text{rec}}_{\text{cmb}}$), for which the axis of the quadrupole has an angular distance to the average axis of the ILC map of 34°. The axes of the octopole of $P^{\text{rec}}_{\text{uncorr}}$ and $T^{\text{rec}}_{\text{corr}}$, though, do not align with the axis of evil. What can we learn from this result? The significance of the alignment between the axes of the quadrupole and octopole in the temperature map has been assessed extensively in earlier works. Here, we only look at the additional information we obtain from the axes of $P^{\text{rec}}_{\text{uncorr}}$. To this end, let us take the preferred axis in the temperature map T as given, and assume that the axes of $P^{\text{rec}}_{\text{uncorr}}$ are distributed isotropically over the sky and independently from each other. In Appendix C.3, we work out the probability for at least one of the axes of $P^{\text{rec}}_{\text{uncorr}}$ being such that the axis of the temperature map is included in the 1σ region around it. This probability amounts to about 50 per cent, due to the large 1σ regions we have.

In order to assess whether the mask or the noise in the WMAP polarization maps is the main

---

[2] We had to regularise the Wiener variances, eqs (4.8) and (4.19), by adding Gaussian noise in order to make them positive definite. This is required by the Cholesky factorisation. However, since the noise was added mostly on small scales, the quadrupole and octopole remained completely unaffected by this. In fact, our results remained unchanged under varying the variance of the added Gaussian noise over 5 orders of magnitude.

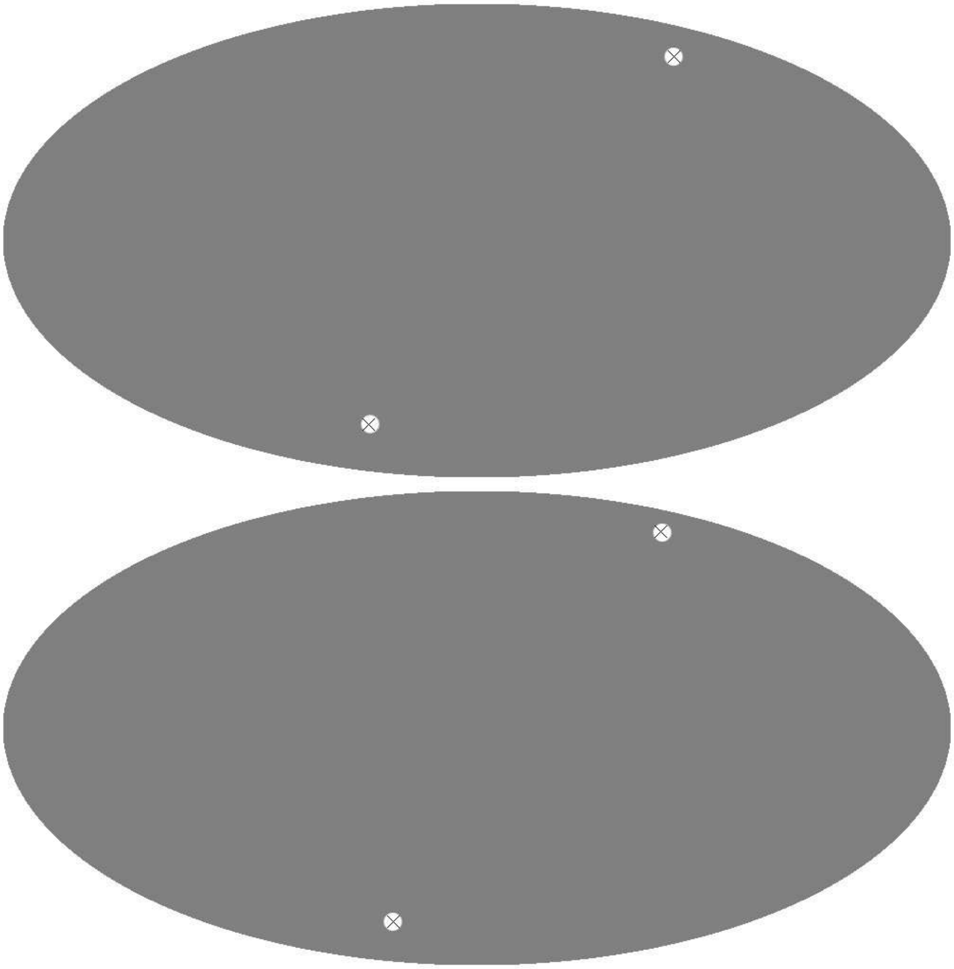


Figure 4.7: Preferred axis of the quadrupole (top panel) and the octopole (bottom panel) for P_{corr} and thus for the ILC map. We reproduce the results of de Oliveira-Costa et al. (2004) within our measurement precision. The axes of quadrupole and octopole point in the same direction, which has been named the axis of evil.

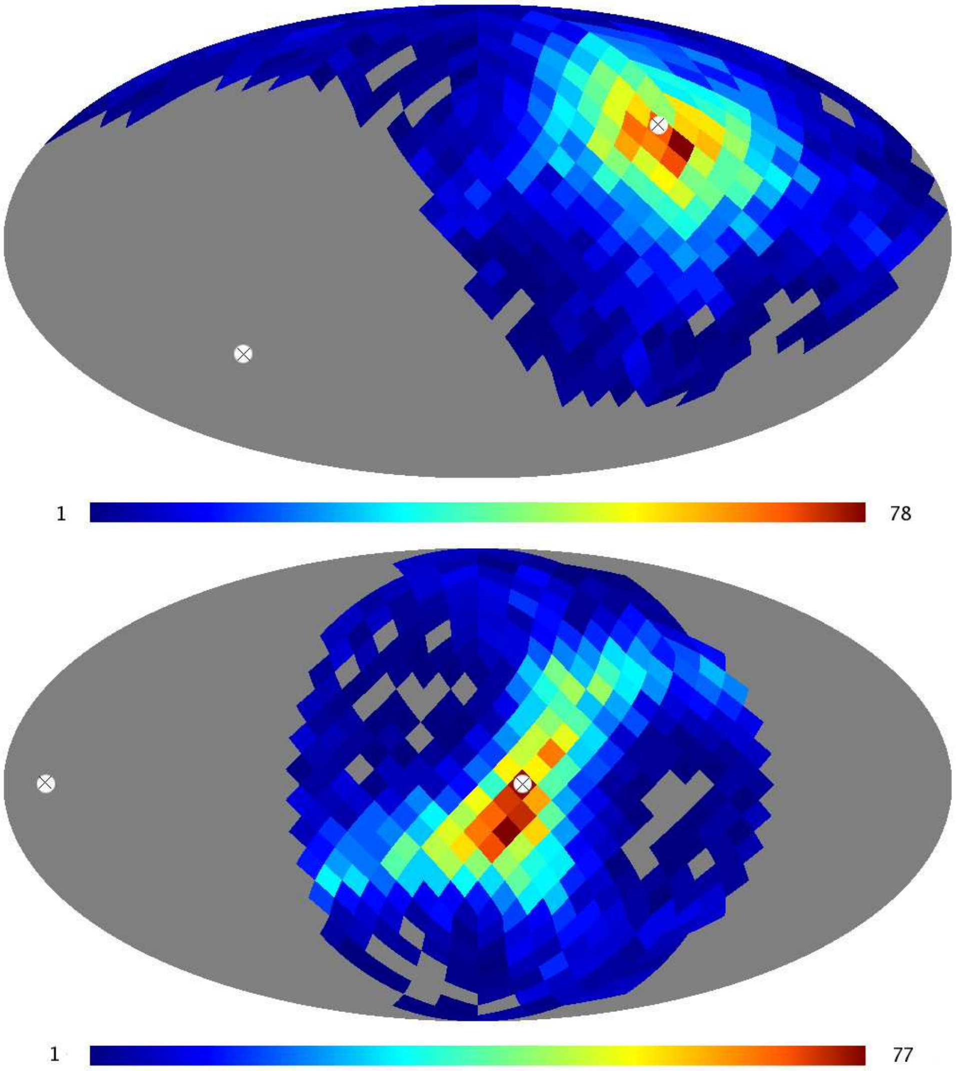


Figure 4.8: Preferred axis of the quadrupole (top panel) and the octopole (bottom panel) for $P^{\mathrm{rec}}_{\mathrm{uncorr}}$. The colour coding counts the number of MC samples whose axis came to lie on the respective pixel. The axis of the quadrupole aligns with the axis of evil within our measurement precision, whereas the axis of the octopole does not.

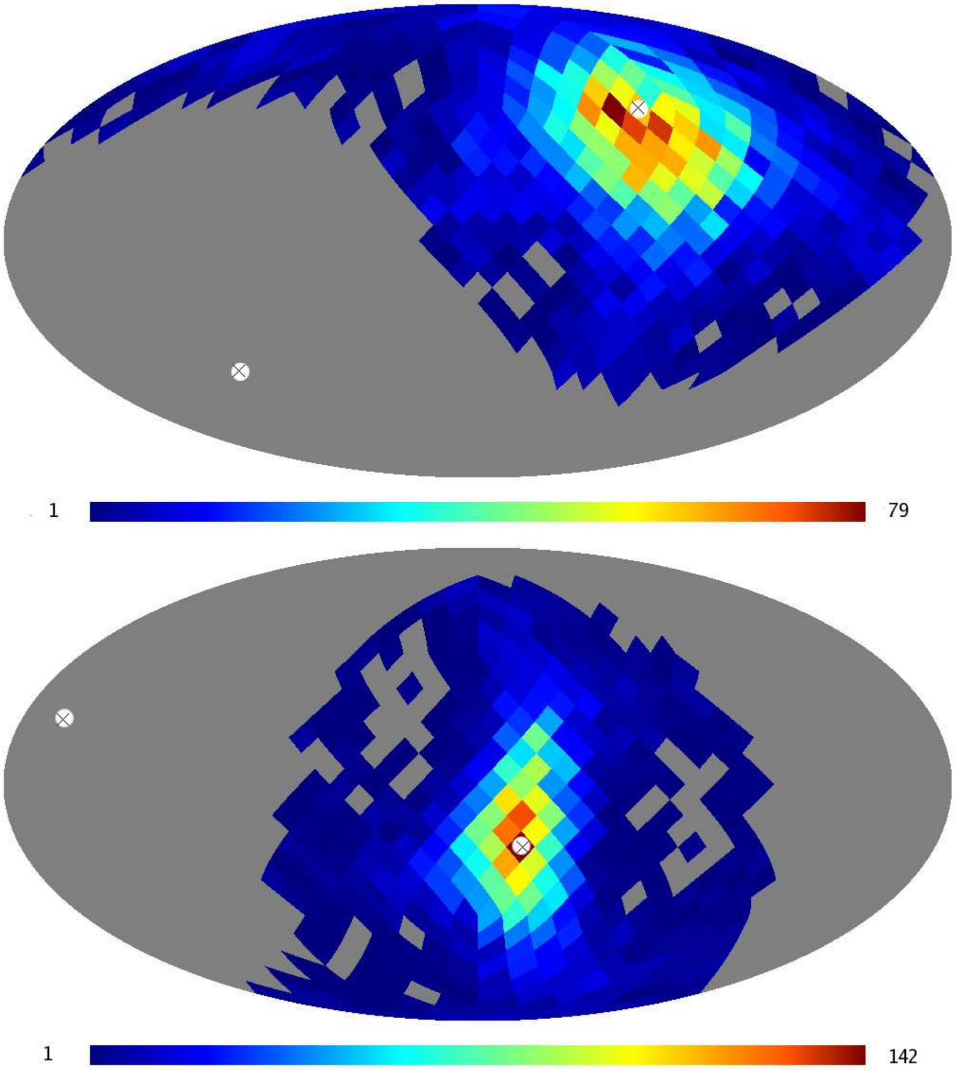


Figure 4.9: Preferred axis of the quadrupole (top panel) and the octopole (bottom panel) for $T_{\mathrm{corr}}^{\mathrm{rec}}$ and thus for $P_{\mathrm{cmb}}^{\mathrm{rec}}$. The axis of the quadrupole aligns with the axis of evil within our measurement precision, whereas the axis of the octopole does not.

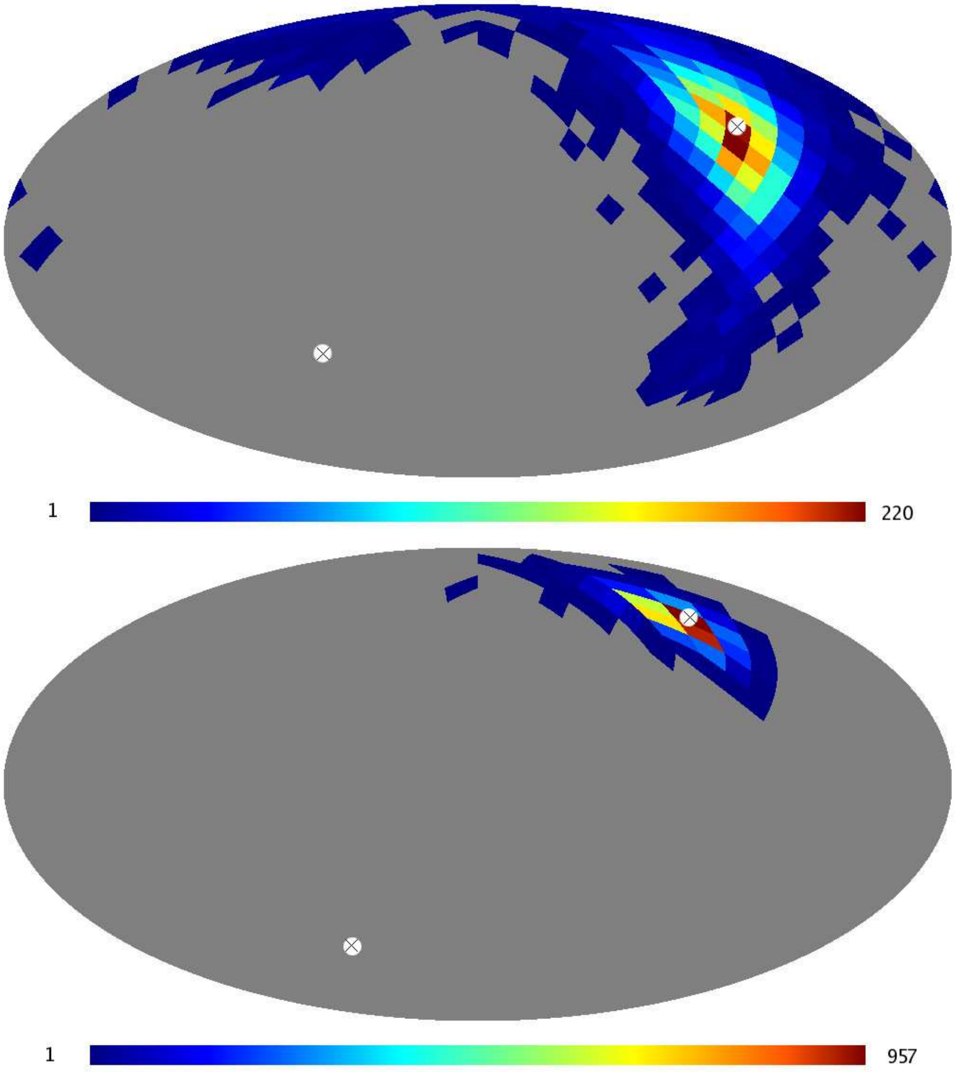


Figure 4.10: Preferred axis of the quadrupole (top panel) and the octopole (bottom panel) for $T^{\mathrm{rec}}_{\mathrm{uncorr}}$. The axes of the quadrupole and the octopole both align with the axis of evil within our measurement precision.

source of uncertainty in the axes, we have determined the uncertainty with the amplitude of the noise covariance matrix rescaled to 10 per cent of the original one. This yields an uncertainty of about 20° in the axes. We have done the same exercise for the noise amplitude downscaled to 1 per cent of the original one, which results in an uncertainty of $7^\circ - 10^\circ$ in the axes. This means that the noise is actually the main source of uncertainty in our analysis rather than the mask. Soon, the *Planck Surveyor* mission (Tauber 2000) will provide us with polarization measurements that have a noise-level which is significantly below the one in the WMAP data. The main problem will then be the contamination of the polarization data by Galactic foregrounds. In the WMAP polarization data, the foregrounds contribute about 20 per cent to the diagonal of the noise covariance matrix N_P in pixel space. With *Planck*, we will be able to determine the foregrounds better than with WMAP, due to the broader frequency range covered by *Planck*. If we assume that the covariance due to residual foregrounds for *Planck* will be between 5 and 50 per cent of the one for WMAP, we will get the uncertainty on the axes down to about 10° and 20°, respectively. With this, we will have a powerful test to probe the axis of evil in polarization.

4.5 Conclusions

In the last few years, a preferred axis has been found in the CMB temperature map, posing a challenge to the cosmological principle. This so-called *axis of evil* denotes the unusual alignment of the preferred axes of the quadrupole and the octopole in the temperature map.

We have split the CMB temperature and polarization maps from WMAP into a part correlated with the respective other map, and an uncorrelated part. If the axis of evil were due to some preferred direction intrinsic to the geometry of the primordial Universe, we would expect its signature to be present in all four of these maps, although this is not true for all theoretical models creating an axis in the temperature map. In particular, the part of the polarization map which is uncorrelated with the temperature map serves as a statistically independent probe of the axis of evil. In order to reduce the noise contained in the polarization maps, we have Wiener filtered the maps before computing the axes. We have then determined the preferred axes of the quadrupole and the octopole for the four maps. In order to assess the uncertainty in the axes coming from the mask, detector noise and residual foregrounds in the polarization maps, we have run MC simulations conditional on the observational data.

For the part of the polarization map which is correlated with the temperature map, P_{corr}, we find that the axes of quadrupole and octopole point in the same direction, confirming earlier results by de Oliveira-Costa et al. (2004). The part of the temperature map which is uncorrelated with the polarization map, $T^{\mathrm{rec}}_{\mathrm{uncorr}}$, exhibits the same alignment of the axes within our measurement precision. For the part of the polarization map which is uncorrelated with the temperature map, $P^{\mathrm{rec}}_{\mathrm{uncorr}}$, we find that only the axis of the quadrupole aligns with the axis of evil, whereas the axis of the octopole does not. The same holds for the correlated part of the temperature map, $T^{\mathrm{rec}}_{\mathrm{corr}}$. We have computed the probability that a rough alignment with the axis of evil, as we find it for the axis of the quadrupole of $P^{\mathrm{rec}}_{\mathrm{uncorr}}$, happens by chance if the axes are distributed isotropically. This probability amounts to 50 per cent for currently available polarization data, due to the large uncertainties in the axes. We are thus looking forward to redoing this analysis with polarization maps from *Planck*, which will yield much more significant results. Of course, similar analyses can be carried out for all other anomalies that have been found in the CMB temperature maps. Note that, instead of working in pixel space as we have done, one could implement the analysis in spherical harmonics space, which would help to separate the E modes we are working with from contamination by B modes.

The approach we have chosen here is a phenomenological approach, since in principle one should take into account that different models causing anomalies in the temperature map predict different signatures in the polarization map. Thus, for a more thorough analysis, one would need to consider particular models of the primordial Universe creating anomalies in the temperature maps, and compute the statistical properties of the uncorrelated polarization map for these. This can be done by modifying a Boltzmann code such as CMBEASY or by simulations as in Dvorkin et al. (2008). One can then try to find these predicted signatures in the uncorrelated polarization map via Bayesian model selection. Such an analysis would truly go beyond the usual a posteriori analysis of finding anomalies in the temperature map, since we would use an actual model to make predictions for the uncorrelated polarization map and then compare these predictions with observations. We leave this promising analysis for future work.

Chapter 5

Efficient cosmological parameter sampling using sparse grids

Note: This chapter, as well as appendix D.1, will be submitted for publication in MNRAS. Section 5.2 as well as parts of the other sections have been written by Dirk Pflüger.

5.1 Introduction

The main two bottlenecks in cosmological parameter estimation using the power spectrum of the cosmic microwave background (CMB) are the calculation of the theoretical C_l-spectrum using Boltzmann codes such as CMBFAST (Seljak & Zaldarriaga 1996), CAMB (Lewis et al. 2000), or CMBEASY (Doran 2005) and the evaluation of the likelihood using the official WMAP likelihood code[1]. There exist several methods to speed up the calculation of the power spectrum (Jimenez et al. 2004; Kaplinghat et al. 2002; Habib et al. 2007) or the WMAP likelihood function $\mathcal{L}$ (Sandvik et al. 2004; Fendt & Wandelt 2007; Auld et al. 2008). These methods are based on different techniques, such as analytic approximations, polynomial fits, and neural networks, which are all trained using a set of training points, for which the real power spectra and likelihood values have to be calculated. Once the codes are trained for a particular cosmological model, they can be used to evaluate the power spectrum or the likelihood function in every subsequent parameter estimation run, which significantly speeds up the Markov Chain Monte Carlo (MCMC) simulations used for parameter estimation. Due to the ever-growing amount of available data, a fast evaluation of the likelihood is becoming of increasing importance, especially when combining CMB data with data-sets whose likelihood is less expensive to evaluate. The *Planck Surveyor* mission (Tauber 2000) will be the upcoming challenge in this respect.

We approximate the WMAP log-likelihood function $\ln\mathcal{L}$ in the spirit of CMBfit (Sandvik et al. 2004) and Pico (Fendt & Wandelt 2007), which work with polynomial fits, and CosmoNet (Auld et al. 2008), an approach based on neural networks. In contrast to the fitting functions constructed therein, we introduce the technique of *sparse grids* in this context to construct an interpolation of $\ln\mathcal{L}$, returning the exact function values at a set of sampling points.

Most straightforward interpolation techniques are based on sets of sampling points in each dimension, typically based on (uniform) grid structures—consider, e. g., piecewise d-linear or piecewise d-polynomial interpolation schemes. Unfortunately, grid-based methods are only feasible in low-dimensional settings, as they suffer from the so-called *curse of dimensionality*: Spending $\tilde{N}$ function evaluations or grid points in one dimension leads to $\tilde{N}^d$ grid points in d dimensions. The exponential dependency on the dimensionality imposes severe restrictions on the number of dimensions that can be handled. Sparse grids, as introduced by Zenger (1991), allow to overcome the curse of dimensionality to some extent, at least for sufficiently smooth functions as it is the case in our setting. Sparse grid interpolation is based on an a priori selection of grid points, requiring significantly fewer grid points than conventional interpolation on a full grid, while preserving the asymptotic error decay of a full-grid interpolation with increasing grid resolution up to a logarithmic factor. This permits us to compute higher-dimensional interpolations and approximations than before. A very good overview about sparse grids, discussing general properties, can be found in Bungartz & Griebel (2004).

---

[1] `http://lambda.gsfc.nasa.gov/product/map/dr3/likelihood_get.cfm`

The sparse grid technique is a completely general approach, not tailored to a single application, and can therefore be used to interpolate any function which is sufficiently smooth. Additionally, as it allows for arbitrary adaptive refinement schemes, the general, fast convergence rates can be improved even further, by adapting to the special characteristics of the underlying target function.

We obtain excellent results, which are competitive to fitting procedures using polynomials (Fendt & Wandelt 2007; Sandvik et al. 2004) or neural networks (Auld et al. 2008) in speed and accuracy. Furthermore, we believe that the interpolation based on sparse grids has several advantages over these approaches. First of all, we can use the results of sparse grid approximation quality (Bungartz & Griebel 2004), guaranteeing the convergence of the interpolation towards the original function with increasing number of grid points.

Second, once we have chosen the volume in which we want to interpolate the function in question, the sparse grid structure itself determines a priori the location of potential sampling points (which can additionally be refined in an adaptive manner a posteriori). This makes it unnecessary to assemble a set of training points beforehand (by running MCMCs as it is done by Fendt & Wandelt (2007), e.g.). The generation of the sampling points and the construction of the interpolant can be strongly parallelised, which makes the sparse grid approach an ideal candidate for computational grid projects such as the AstroGrid[2]. The time needed to construct the interpolant is determined almost only by the time it takes to evaluate the likelihood at the sampling points. We do not need additional training time as in the case of Auld et al. (2008).

Furthermore, polynomial fits to a set of training points run the risk of creating unphysical wiggles if the polynomial degree of the fitting function is chosen too high with respect to the amount of training points available. Using the sparse grids approach, piecewise polynomials of low degree are sufficient, as we are not fitting certain evaluation points, but rather interpolating a function. Increasing the accuracy is therefore equivalent to evaluating at more sampling points.

Sparse grids are based on a hierarchical formulation of the underlying basis functions, which can be used to obtain a generic estimate of the current approximation error while evaluating more and more sampling points. This can be directly used as a criterion for adaptive refinement as well as to stop further refinement.

Another advantage is that the projection of sparse grid interpolations can be done in a very fast and simple way. This would make sparse grids in principle a good candidate for sampling posteriors and projecting them directly, without having to use a Markov Chain approach in order to marginalise the posterior. Given that MCMCs need to determine the points sequentially and can therefore not be parallelised (apart from running several chains at the same time), it would be highly desirable to find alternatives that can be run in parallel.

---

[2] `http://www.d-grid.de/index.php?id=45&L=1`

We have attempted to use sparse grids in order to substitute the MCMCs in cosmological parameter estimations. In order to directly project the posterior distribution we would need to sample the posterior rather than its logarithm. Since, in general, the logarithm of a probability density function is considerably more well-behaved than the function itself, Sandvik et al. (2004), Fendt & Wandelt (2007), and Auld et al. (2008) all operate in log-space to speed up the generation of MCMCs instead. As the convergent phase of the interpolation with sparse grids sets in rather late when interpolating Gaussian functions (and thus the WMAP likelihood, which is close to a d-dimensional Gaussian), we restricted ourselves to the log-likelihood as well.

This chapter is organised as follows. First, we describe the basics of sparse grids in section 5.2, introducing a modification of the standard sparse grid approach, thus adapting the latter to our problem. In section 5.3, we then present the interpolation of the WMAP likelihood for two different sets of parameters in both six and seven dimensions. We show that the results obtained for regular (non-adaptive) sparse grids are already competitive to other approaches and demonstrate how adaptive refinement can further improve the results. Section 5.4 finally concludes this chapter.

5.2 Basics of sparse grids

Standard grid-based approaches of interpolating a function f exhibit the curse of dimensionality, a term going back to Bellman (1961): Any straightforward discretisation scheme which employs $\tilde{N}$ grid points (or, equivalently, degrees of freedom) in one dimension leads to $\tilde{N}^d$ grid points in d dimensions. For reasonable $\tilde{N}$, the exponential dependency on the number of dimensions typically does not allow to handle more than four-dimensional problems.

Sparse grids are able to overcome this hurdle to some extent, requiring significantly fewer grid points than a full grid, while preserving the asymptotic error decay of full grid interpolation with increasing grid resolution up to a logarithmic factor. Sparse grids have originally been developed for the solution of partial differential equations (Zenger 1991) and have meanwhile been applied to various problems, see Bungartz & Griebel (2004) and the references cited therein. Recent work on sparse grids includes stochastic and non-stochastic partial differential equations in various settings (von Petersdorff & Schwab 2006; Ganapathysubramanian & Zabaras 2007; Widmer et al. 2008), as well as applications in economics (Reisinger & Wittum 2007; Holtz 2008), regression (Garcke & Hegland 2009; Garcke 2006), classification (Bungartz et al. 2008; Garcke et al. 2001), fuzzy modelling (Klimke et al. 2006), and more. Note that (non-adaptive) sparse grids are closely related to the technique of hyperbolic crosses (Temlyakov 1993).

In this section, we provide a brief overview of sparse grids for interpolation. For a detailed derivation of the characteristics of sparse grids, we refer to Bungartz & Griebel (2004). We start by formulating the interpolation on a conventional full grid using hierarchical basis functions, from which we then derive the interpolation on a sparse grid by omitting the basis functions contributing least to the interpolation.

5.2.1 General idea of interpolating a function on a full grid

We consider the piecewise d-linear interpolation of a function $f : \Omega \to \mathbb{R}$ which is given only algorithmically, i.e., we have no closed form of f but we can only evaluate f at arbitrary points

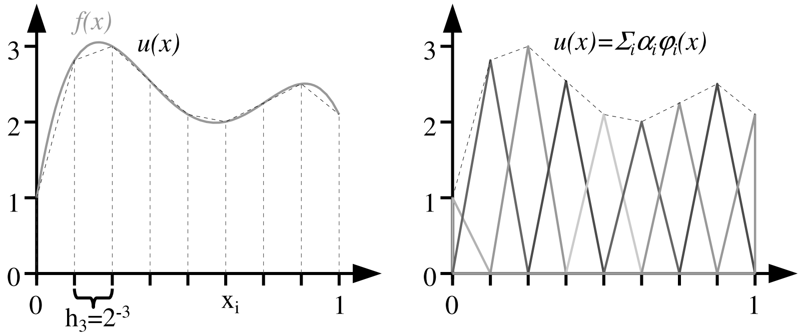


Figure 5.1: One-dimensional piecewise linear interpolation $u(x)$, dashed, of a function $f(x)$, solid, (left) by a linear combination of hat basis functions (right).

using a numerical code. As we want to discretise our domain of interest Ω, we restrict Ω to a compact sub-volume of $\mathbb{R}^d$; here, $\Omega \equiv [0,1]^d$, the d-dimensional unit-hypercube. (For the standard approach of sparse grids techniques, we only consider functions that are zero on the boundary of the volume on which they are defined. This assumption will be dropped when we come to the interpolation of the log-likelihood of WMAP.)

To construct an interpolant u of f, we discretise Ω via a regular grid, obtaining equidistant grid points $\boldsymbol{x}_i$, with mesh width $h_n = 2^{-n}$ for some discretisation or refinement level n, at which we evaluate and interpolate f. If we define a suitable set of piecewise d-linear basis functions $\varphi_i(\boldsymbol{x})$, we can obtain $u(\boldsymbol{x})$ from the space of continuous, piecewise d-linear functions V_n by combining them adequately as a weighted sum of basis functions, i.e.

$$f(\boldsymbol{x}) \approx u(\boldsymbol{x}) \equiv \sum_i \alpha_i \varphi_i(\boldsymbol{x})$$

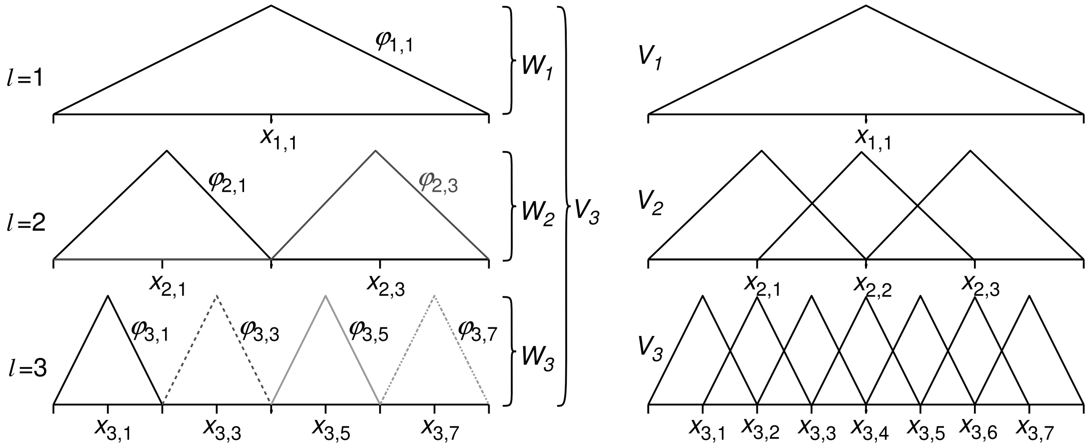
with coefficients α_i. Fig. 5.1 sketches the idea for a one-dimensional example, using the standard nodal basis.

The curse of dimensionality, encountered when using a full grid, can be circumvented by a suitable choice of basis functions: We need a basis where the relevant information is represented by as few basis functions as possible. Most basis functions can then be omitted as they contribute only little to the interpolation of f, reducing a full grid to a sparse grid and allowing us to handle higher-dimensional functions than before. A suitable basis can be found by a hierarchical construction as introduced in the following section.

5.2.2 Hierarchical basis functions in one dimension

Sparse grids depend on a hierarchical decomposition of the underlying approximation spaces. Therefore, and first considering only the one-dimensional case which we will later extend to d dimensions, we use the standard hat function,

$$\varphi(x) = \max(1 - |x|, 0)\,, \tag{5.1}$$



$l=1$
$\varphi_{1,1}$
$x_{1,1}$
W_1
$l=2$
$\varphi_{2,1}$
$\varphi_{2,3}$
$x_{2,1}$
$x_{2,3}$
W_2
V_3
$l=3$
$\varphi_{3,1}$
$\varphi_{3,3}$
$\varphi_{3,5}$
$\varphi_{3,7}$
$x_{3,1}$
$x_{3,3}$
$x_{3,5}$
$x_{3,7}$
W_3
V_1
V_2
V_3
$x_{1,1}$
$x_{2,1}$
$x_{2,2}$
$x_{2,3}$
$x_{3,1}$
$x_{3,2}$
$x_{3,3}$
$x_{3,4}$
$x_{3,5}$
$x_{3,6}$
$x_{3,7}$

Figure 5.2: One-dimensional basis functions $\varphi_{l,i}$ and corresponding grid points $x_{l,i}$ up to level $n = 3$ in the hierarchical basis (left) and the common nodal point basis (right).

from which we derive one-dimensional hat basis functions by dilatation and translation,

$$\varphi_{l,i}(x) \equiv \varphi(2^l x - i)\,, \tag{5.2}$$

which depend on a level l and an index i, $0 < i < 2^l$. The basis functions have local support and are centred at grid points $x_{l,i} = 2^{-l} i$, at which we will interpolate f. Introducing the hierarchical index sets

$$I_l \equiv \left\{ i \in \mathbb{N} : 1 \leq i \leq 2^l - 1, i \text{ odd} \right\}\,, \tag{5.3}$$

we obtain a set of hierarchical subspaces W_l,

$$W_l \equiv \operatorname{span}\{\varphi_{l,i}(x) : i \in I_l\}\,. \tag{5.4}$$

We can then formulate the space of piecewise linear functions V_n on a full grid with mesh width h_n for a given level n as a direct sum of W_l,

$$V_n = \bigoplus_{l \leq n} W_l\,, \tag{5.5}$$

see Fig. 5.2. Note that all basis functions of the same subspace W_l have the same size, shape, and compact support, that their supports are non-overlapping, and that together they cover the whole domain.

The interpolation $u(x) \in V_n$ can then be written as a finite sum,

$$u(x) = \sum_{l \leq n, i \in I_l} \alpha_{l,i} \varphi_{l,i}(x)\,, \tag{5.6}$$

where the so-called (hierarchical) surplusses $\alpha_{l,i}$ are uniquely indexed by the same level and index as the corresponding basis functions.

5.2.3 Higher-dimensional interpolation on a full grid

The basis functions are extended to the d-dimensional case via a tensor product approach,

$$\varphi_{\boldsymbol{l},\boldsymbol{i}}(\boldsymbol{x}) \equiv \prod_{j=1}^{d} \varphi_{l_j,i_j}(x_j)\,, \tag{5.7}$$

with the d-dimensional multi-indices $\boldsymbol{l}$ and $\boldsymbol{i}$ indicating level and index for each dimension. The other one-dimensional notations can be transferred to the arbitrary-dimensional case as well, consider, e.g. , the index set $I_{\boldsymbol{l}}$,

$$I_{\boldsymbol{l}} \equiv \left\{ \boldsymbol{i} : 1 \leq i_j \leq 2^{l_j} - 1, i_j \text{ odd}, 1 \leq j \leq d \right\}, \tag{5.8}$$

the subspaces $W_{\boldsymbol{l}}$, the space V_n of piecewise d-linear functions with mesh width h_n in each dimension,

$$V_n = \bigoplus_{|\boldsymbol{l}|_\infty \leq n} W_{\boldsymbol{l}}\,, \tag{5.9}$$

leading to a full grid with $(2^n - 1)^d$ grid points, and to the interpolant $u(\boldsymbol{x}) \in V_n$,

$$u(\boldsymbol{x}) = \sum_{|\boldsymbol{l}|_\infty \leq n, \boldsymbol{i} \in I_{\boldsymbol{l}}} \alpha_{\boldsymbol{l},\boldsymbol{i}} \varphi_{\boldsymbol{l},\boldsymbol{i}}(\boldsymbol{x})\,. \tag{5.10}$$

Here and later on, we need the l_1-norm $|\boldsymbol{l}|_1 = \sum_{j=1}^{d} l_j$ and the maximum-norm $|\boldsymbol{l}|_\infty = \max_{1 \leq j \leq d} l_j$ of multi-indices $\boldsymbol{l}$. Fig. 5.3 shows some 2-dimensional examples for the basis functions of the subspaces $W_{\boldsymbol{l}}$, which correspond to anisotropic sub-grids with mesh-width h_{l_j} in dimension j characterised by the multi-index $\boldsymbol{l}$.

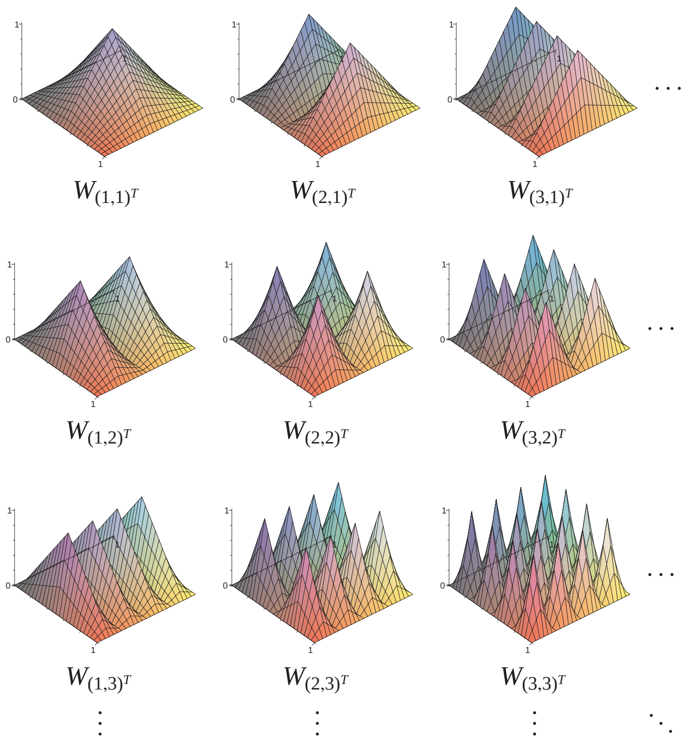
5.2.4 Sparse grids

Starting from the hierarchical representation of V_n by the subspaces $W_{\boldsymbol{l}}$, we can now select those subspaces that contribute most to the overall solution of the full-grid interpolation in eq. (5.10). If the function we want to approximate meets certain smoothness conditions—the mixed second derivatives have to be bounded—this can be done a priori as we can derive bounds for the contributions of the different subspaces. We then obtain the sparse grid space

$$V_n^{(1)} \equiv \bigoplus_{|\boldsymbol{l}|_1 \leq n+d-1} W_{\boldsymbol{l}}\,, \tag{5.11}$$

leaving out those subspaces from the full grid space V_n with many basis functions of small support. (The exact choice of subspaces depends on the norm in which we measure the error; the result above is optimal for both the L_2 norm and the maximum norm.) Note that in the one-dimensional case, the sparse grid space equals the full grid space.

Fig. 5.4 shows the selection of subspaces and the resulting sparse grid for $n = 3$, i.e. the sparse grid space $V_3^{(1)}$. Compared to the full grid for the same discretisation level n (the full grid space V_3 would also comprise the grey subspaces in Fig. 5.4), this reduces the number of grid points (and therefore function evaluations and unknowns) significantly from $\mathcal{O}(h_n^{-d}) = \mathcal{O}(2^{nd})$ to $\mathcal{O}(h_n^{-1}(\log h_n^{-1})^{d-1})$ – whereas the asymptotic accuracy deteriorates only slightly from $\mathcal{O}(h_n^2)$ to



$W_{(1,1)^T}$ $W_{(2,1)^T}$ $W_{(3,1)^T}$ …

$W_{(1,2)^T}$ $W_{(2,2)^T}$ $W_{(3,2)^T}$ …

$W_{(1,3)^T}$ $W_{(2,3)^T}$ $W_{(3,3)^T}$ …

$\vdots$ $\vdots$ $\vdots$ $\ddots$

Figure 5.3: Basis functions of the subspaces $W_{\boldsymbol{l}}$ for $|\boldsymbol{l}|_\infty \leq 3$ in two dimensions.

$O(h_n^2(\log h_n^{-1})^{d-1})$, see Bungartz & Griebel (2004) for detailed derivations. Fig. 5.5 shows sparse grids in two and three dimensions for level $n = 6$ each.

Functions which do not meet the smoothness requirements or which show significantly differing characteristics (comprising steep regions as well as flat ones, e.g.) can be tackled as well, if adaptive refinement is used. The sparse grid structure defined in eq. (5.11) defines an a priori selection of grid points which is optimal if certain smoothness conditions are met and no further knowledge about the function in question is known or used. An adaptive (a posteriori) refinement can additionally select which grid points in a sparse grid structure should be refined next, due to local error estimation, e.g. To refine a grid point, often all $2d$ children in the hierarchical structure are added to the current grid, if they haven't been created yet. Note that it usually has to be ensured that all missing parents have to be created, as algorithms working on sparse grids depend on traversals of the hierarchical tree of basis functions. If additional knowledge about the problem at hand is available, it can be used in the criterion for adaptive refinement, allowing to better adapt to problem specific characteristics.

5.2.5 Extension to functions that are non-zero on the boundary

Up unto now we have only considered functions that are zero on the domain's boundary $\delta\Omega$. To allow for non-zero values on the boundary, usually additional grid points located directly on $\delta\Omega$ are introduced. For example, the one-dimensional basis on level one, containing only $\varphi_{1,1}(x)$, is extended by two basis functions with level 0 and indices 0 and 1 restricted to Ω, namely $\varphi_{0,0}(x)$ and $\varphi_{0,1}(x)$. They are then extended to the d-dimensional case as before, with the exception that the new basis now contains basis functions on the modified level one with overlapping support.

Apparently, this approach results in many more grid points (and therefore expensive function

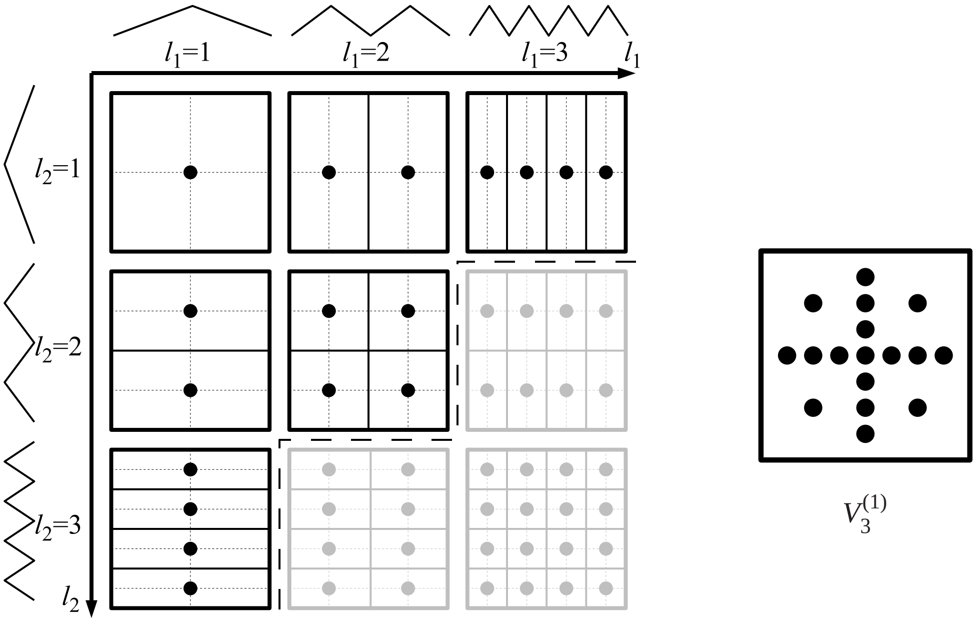


Figure 5.4: The two-dimensional subspaces $W_{\mathbf{l}}$ up to $l = 3$ ($h_3 = 1/8$) in each dimension. The optimal selection of subspaces (black) and the corresponding sparse grid on level $n = 3$ for the sparse grid space $V_3^{(1)}$. The corresponding full grid of level 3 corresponds to the direct sum of all subspaces that are shown.

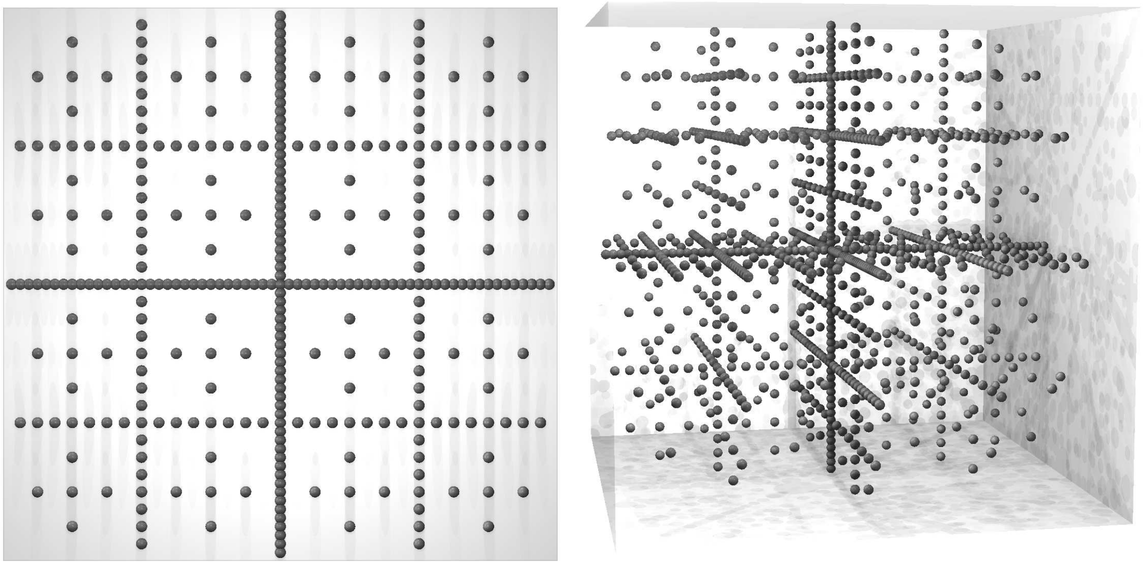


Figure 5.5: Sparse grids in two and three dimensions for level $n = 6$.

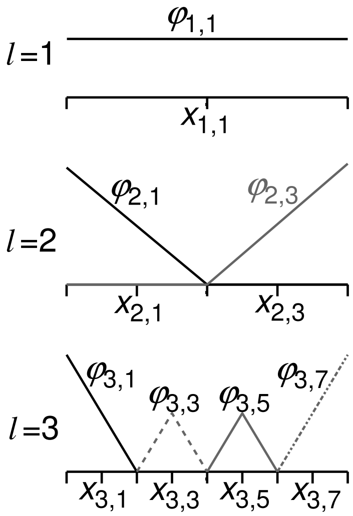


Figure 5.6: Modified one-dimensional basis functions $\varphi_{l,i}$: constant on level 1 and "folded up" if adjacent to the boundary on all other levels.

evaluations) than before. This shows quite nicely that it is not sufficient to just consider the asymptotic behaviour: asymptotically, nothing changes, but for very high dimensionalities we are not able to even start to interpolate any more. In d dimensions, the basis for the subspace $W_{\mathbf{1}}$ for example contains already 3^d basis functions, rather than a single one. Especially in settings where a very high accuracy close to the boundary is not required—which holds in our case—(or where an adaptive selection of grid points is used in any case), it can be advantageous to omit the grid points on the boundary, and instead modify the basis functions to extrapolate towards the boundary of the domain.

We modify the one-dimensional basis functions as follows: On level 1, we have only one degree of freedom; the best guess towards the boundary is to assume the same value, leading to a constant basis function. On all other levels, we extrapolate linearly towards the boundary, "folding up" the uttermost basis functions. All other basis functions remain unchanged, yielding

$$\varphi_{l,i}(x) \equiv \begin{cases} 1 & \text{if } l = 1 \wedge i = 1\,, \\ \left\{ \begin{array}{ll} 2 - 2^l \cdot x & \text{if } x \in \left[0, \frac{1}{2^{l-1}}\right] \\ 0 & \text{else} \end{array} \right\} & \text{if } l > 1 \wedge i = 1\,, \\ \left\{ \begin{array}{ll} 2^l \cdot x + 1 - i & \text{if } x \in \left[1 - \frac{1}{2^{l-1}}, 1\right] \\ 0 & \text{else} \end{array} \right\} & \text{if } l > 1 \wedge i = 2^l - 1\,, \\ \varphi\left(x \cdot 2^l - i\right) & \text{else} \end{cases} \tag{5.12}$$

in one dimension, see Fig. 5.6. The d-dimensional basis functions are obtained as before via a tensor product of the one-dimensional ones.

5.3 Interpolation of the WMAP likelihood surface

We now construct an interpolation of the logarithm of the WMAP likelihood, $\ln\mathcal{L}$, using sparse grids. In order to adapt the problem to our interpolation approach, we first use a 6-dimensional set of so-called normal parameters introduced in Sandvik et al. (2004), which are a transformation of the usual cosmological parameters such that the major axes of the Gaussian align with the coordinate axes. The logarithm of the likelihood is then well-approximated by a sum of one-dimensional parabolas in the different parameters, a fact that we will take advantage of by using the modified basis-functions described in eq. (5.12). For this set of normal parameters, we obtain an accurate interpolation already for a comparably low refinement level. This is shown for the 6-dimensional model as well as for a 7-dimensional extension, using the running of the spectral index as an additional parameter.

However, as a subsequent step we demonstrate that the parameter transformation is not essential for obtaining a good interpolation. By investing more grid points, we obtain an accurate interpolation as well when using directly the 6- and 7-dimensional standard parameter set, which is usually used in cosmological parameter sampling. This approach shows the advantage of sparse grids of being rather generic. Furthermore, we are not restricted to the parameter range in which the transformation to normal parameters can be inverted.

5.3.1 Choice of basis functions

We use the modified basis functions as introduced in eq. (5.12), which are well-suited for our problem. First, and as already indicated in section 5.2, the region close to the domain's boundary is less important in our setting than the centre of Ω: We will centre the domain of interest roughly at the maximum of the log-likelihood function $\ln\mathcal{L}$ and determine the boundary such that it includes the region with $(\ln\mathcal{L}_{\max} - \ln\mathcal{L}) \gtrsim 25$, which we will refer to as the 25 log-likelihood region (see section 5.3.3). Towards the boundaries of our intervals, the likelihood is then effectively zero and thus no great accuracy is needed in these regions. Therefore, we do not want to spend too much work on $\delta\Omega$. Using the modified boundary functions, we extrapolate (d-linearly) towards $\delta\Omega$, see the discussion of the modified basis functions above.

Second, the modifications are especially well-suited if the function to interpolate can be separated into a sum of one-dimensional functions[3]. Assume that the likelihood $\mathcal{L}$ was a perfect product of one-dimensional Gaussians,

$$\mathcal{L}(\boldsymbol{x}) = c \cdot \mathrm{e}^{-a_1(x_1-\mu_1)^2-\ldots-a_d(x_d-\mu_d)^2}\,, \tag{5.13}$$

centred at $(\mu_1, \ldots, \mu_d)^T$. Then the interpolation of the log-likelihood $\ln\mathcal{L}$ reduces to d one-dimensional problems,

$$\ln\mathcal{L}(\boldsymbol{x}) = \ln c + \sum_{k=1}^{d} f_k(x_k)\,, \qquad f_k(x_k) = -a_k(x_k-\mu_k)^2\,, \tag{5.14}$$

separating into a constant term plus a sum of functions that are constant in all directions but one.

Keeping in mind that the one-dimensional basis function on level 1, $\varphi_{1,1}(x)$, is constant (cf. Fig. 5.6), this simplifies the interpolation task. The d-dimensional basis function on level $\mathbf{1}$, $\varphi_{\mathbf{1},\mathbf{1}}(\boldsymbol{x})$, serves as an offset. (Only if $(\mu_1, \ldots, \mu_d)^T$ is the centre of Ω, $\alpha_{\mathbf{1},\mathbf{1}}\varphi_{\mathbf{1},\mathbf{1}}(\boldsymbol{x})$ exactly expresses $\ln c$.) Additionally, it is sufficient to spend only grid points on the main axes of the sparse grid (level 1 in all dimensions but one) to approximate the remaining one-dimensional contributions $f_k(x_k)$ arbitrarily well:

---

[3] In this case, the analysis of variance (ANOVA) would exhibit no terms coupling between the parameters.

$$u(\boldsymbol{x}) = \underbrace{\alpha_{\mathbf{1},\mathbf{1}}\varphi_{\mathbf{1},\mathbf{1}}(\boldsymbol{x})}_{\ln c} + \sum_{k=1}^{d} \underbrace{\left(\sum_{l_k,i_k} \alpha_{\boldsymbol{l},\boldsymbol{i}}\, \varphi_{l_k,i_k}(x_k) \prod_{1\le j\le d,\, j\neq k} \varphi_{1,1}(x_j) \right)}_{f_k(x_k)} . \tag{5.15}$$

Of course, $\mathcal{L}$ is not a perfect product of one-dimensional Gaussians; grid points that do not lie on the sparse grid's main axes account for the additional mixed (correlated) terms of $\ln \mathcal{L}$. Given that in sparse grids a large amount of points lie on the main axes, this mechanism works very well—the better, the less correlation between the different parameters exists.

In order to take as much advantage as possible of the effects described above, we introduce a parameter transformation in the following section, for which the new parameters are less correlated. However, the fact that the interpolation using the standard parameters—which have much stronger correlations—works as well, spending just more grid points, will show that the sparse grid approach does not depend on this argumentation: Sparse grids can make use of such properties but do not rely on them.

5.3.2 Normal parameters

The set of cosmological parameters describing the ΛCDM model consists of the Hubble constant, $h \equiv \frac{H_0}{100\,\mathrm{km/(sMpc)}}$, the density parameter of vacuum energy, Ω_Λ, the ones of baryons, Ω_b, and of matter (baryonic + dark), Ω_m, the optical depth to the last scattering surface, τ, the scalar spectral index of the primordial power spectrum, n_s, and the scalar initial amplitude, A_s. We will refer to these parameters as cosmological parameters. For a more detailed description of the cosmological parameters, we refer to Coles & Lucchin (2002). In the literature, there have been several attempts to transforming these parameters into a set of parameters that mirror the various physical effects on the CMB power spectrum (Hu et al. 2001; Kosowsky et al. 2002). In Chu et al. (2003), a set of parameters is provided in which the likelihood-surface of the CMB is well approximated by a multivariate Gaussian with the major axes aligned with the coordinate axes. Here, we use the parameters given by Sandvik et al. (2004), where the parameter set of Chu et al. (2003) is combined with the other parameter sets mentioned, in order to bring the major axes of the likelihood surface even closer to the coordinate axes. The new parameters are then $\{\Theta_s, h_2, h_3, t, A_*, Z\}$, which we refer to as normal parameters. When working with the latter, the logarithm of the likelihood is well-approximated by a sum of one-dimensional parabolas in the different parameters. The basis functions introduced above are therefore ideally adapted to this problem. In the following, we repeat the definitions of the normal parameters for convenience.

The first parameter of our set is the angle subtended by the acoustic scale

$$\Theta_s \equiv \frac{r_s(a_{\mathrm{ls}})}{D_A(a_{\mathrm{ls}})} \frac{180}{\pi} , \tag{5.16}$$

where the index ls denotes the time of last scattering, $D_A(a_{\mathrm{ls}})$ stands for the comoving angular diameter distance to the surface of last scattering (which we will come back to later), and $r_s(a_{\mathrm{ls}})$ is the comoving sound horizon at last scattering,

$$r_s(a_{\rm ls}) \equiv \int_0^{t_{\rm ls}} \frac{c_s(t)}{a(t)} dt\,. \tag{5.17}$$

Here, $c_s(t)$ denotes the sound speed for the baryon-photon-fluid at time t, which is well approximated by

$$c_s(t)^2 \approx \frac{1}{3}(1+3\frac{\rho_b}{\rho_\gamma})^{-1}\,, \tag{5.18}$$

with the index b standing for baryons and the index γ for photons. Using the Friedmann equations and ignoring the vacuum energy at last scattering, $r_s(a_{\rm ls})$ can be shown to be (Sandvik et al. 2004; Kosowsky et al. 2002)

$$r_s(a_{\rm ls}) = \frac{2\sqrt{3}}{3H_0\sqrt{\Omega_m}}\sqrt{\frac{a_{\rm ls}}{R_{\rm ls}}}\ln\frac{\sqrt{1+R_{\rm ls}}+\sqrt{R_{\rm ls}+r_{\rm ls}R_{\rm ls}}}{1+\sqrt{r_{\rm ls}R_{\rm ls}}}\,, \tag{5.19}$$

where

$$R_{\rm ls} \equiv \frac{3\rho_b(a_{\rm ls})}{4\rho_\gamma(a_{\rm ls})} = 30\,w_b\left(\frac{z_{\rm ls}}{10^3}\right)^{-1}\,, \tag{5.20}$$

$$r_{\rm ls} \equiv \frac{\rho_r(a_{\rm ls})}{\rho_m(A_*)} = 0.042\,w_m^{-1}\left(\frac{z_{\rm ls}}{10^3}\right)\,. \tag{5.21}$$

The index r stands for radiation, i.e., ρ_r consists of the sum of photon and neutrino energy densities, and the index m is used for matter (baryons + dark matter). We define $w_m \equiv \Omega_m h^2$ in the same way as $w_b \equiv \Omega_b h^2$. The redshift at last scattering, $z_{\rm ls}$, is well approximated by (Hu et al. 2001)

$$z_{\rm ls} = 1048\,(1+0.00124\,w_b^{-0.738})(1+g_1 w_m^{g_2})\,, \tag{5.22}$$

$$g_1 \equiv 0.0783\,w_b^{-0.238}(1+39.5\,w_b^{0.763})^{-1}\,, \tag{5.23}$$

$$g_2 \equiv 0.560\,(1+21.1w_b^{1.81})^{-1}\,. \tag{5.24}$$

As already mentioned, $D_A(a_{\rm ls})$ in eq. (5.16) denotes the comoving angular diameter distance to the surface of last scattering and is given by

$$D_A(a_{\rm ls}) = \frac{c}{H_0}\int_{a_{\rm ls}}^1 \frac{1}{\sqrt{\Omega_\Lambda \tilde{a}^4+\Omega_m\,\tilde{a}+\Omega_r}}d\tilde{a}\,. \tag{5.25}$$

The second and third parameters in our set are the ratios of the second and the third peak to the first peak in the C_l^T spectrum of the CMB (Hu et al. 2001), where the tilt-dependence is factored out (Page et al. 2003),

$$h_2 \equiv 0.0264\,w_b^{-0.762}\exp\left(-0.476\,[\ln(25.5\,w_b+1.84\,w_m)]^2\right)\,, \tag{5.26}$$

$$h_3 \equiv 2.17\left(1+\left(\frac{w_b}{0.044}\right)^2\right)^{-1} w_m^{0.59}\left(1+1.63\left(1-\frac{w_b}{0.071}\right)w_m\right)^{-1}\,. \tag{5.27}$$

We use the tilt parameter given by Sandvik et al. (2004), which is a slightly modified version of the one in Chu et al. (2003) in order to minimise the correlation with w_b:

$$t \equiv \left(\frac{w_b}{0.024}\right)^{-0.5233} 2^{n_s-1} . \tag{5.28}$$

The amplitude parameter is

$$A_* \equiv \frac{\tilde{A}_s}{2.95 \times 10^{-9}} \mathrm{e}^{-2\tau} \left(\frac{k}{k_p}\right)^{n_s-1} w_m^{-0.568} , \tag{5.29}$$

where $k_p = 0.05\mathrm{Mpc}^{-1}$ denotes the pivot point. The normalisation factor of 2.95×10^{-9} comes in because we use the scalar amplitude $\tilde{A}_s$ of CMBEASY, which is defined as the primordial power of the curvature fluctuations evaluated at the pivot point, $\tilde{A}_s \equiv \Delta_R^2(k_p)$. It is related to the scalar amplitude A_s of CMBFAST, which is used in Sandvik et al. (2004), by $A_s = \frac{\tilde{A}_s}{2.95\times10^{-9}}$ (Verde et al. 2003a). Finally, we use the physical damping due to the optical depth to last scattering as our last parameter:

$$Z \equiv \mathrm{e}^{-\tau} . \tag{5.30}$$

In order to construct the interpolation of the likelihood surface, we need the transformation that maps the normal parameters back onto cosmological parameters. The reason for this is the way we construct the interpolation: Our sparse grid algorithm chooses the normal parameters where it wants to refine the grid, which we then need to transform into cosmological parameters to run CMBEASY and the WMAP-likelihood code. Our technique of inverting the parameter transformation is presented in appendix D.1.

5.3.3 Generation of test set and choice of interpolation range

For choosing the parameter range in which to construct the interpolation, we have run MCMCs containing about 50,000 points at a temperature of $T = 3$. That is, in the Metropolis algorithm we choose the transition probability $a(\boldsymbol{x}, \boldsymbol{y})$ from a point $\boldsymbol{x}$ in the chain to a new point $\boldsymbol{y}$ to be $a(\boldsymbol{x}, \boldsymbol{y}) \equiv \min\left\{(\mathcal{L}(\boldsymbol{y})/\mathcal{L}(\boldsymbol{x}))^{\frac{1}{T}}, 1\right\}$. Using this transition probability with $T = 1$ results in the usual Metropolis algorithm, whereas choosing $T = 3$ allows us to explore a larger parameter range than with the regular algorithm. These chains covered a region reaching out to about 25 log-likelihoods around the peak.

The optical depth to the last scattering surface, τ, which can be determined from the CMB polarization, is not well-constrained by the WMAP polarization data due to their low signal-to-noise ratio. Therefore, when running the MCMCs at $T = 3$, we had to restrict τ to the physically meaningful range $\tau \geq 0$. This restriction corresponds to $Z \leq 1$ for the normal parameters. In the case of the normal parameters in 7 dimensions, we had to additionally restrict the intervals to $h_2 \leq 0.52$ and $h_3 \geq 0.38$, which is the range in which the parameter transformation is invertible. Furthermore, we chose to restrict our set of points to be within the 25 log-likelihood region around the peak.

In order to roughly centre our intervals at the maximum of the log-likelihood function, we have determined the latter using a few runs of a simple simplex search.[4] The interval boundaries were then defined as the box centred at the maximum which contains all points of the above-described

---

[4] We were running several simplex searches and chose the result with the highest value of the log-likelihood. The runs did not all converge to exactly the same point, which we think was due to numerical issues (the log-likelihood was presumably not completely convex, which could be dues to the dips we will mention in section 5.3.5).

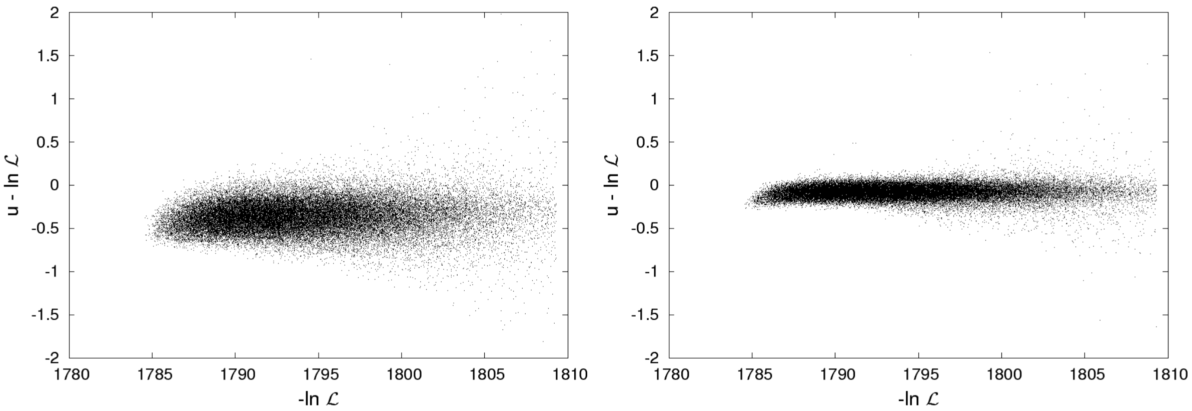


Figure 5.7: Absolute error of the interpolation with respect to the real log-likelihood in 6 dimensions for an interpolation with a sparse grid of level 5 (left panel) and of level 6 (right panel) for normal parameters.

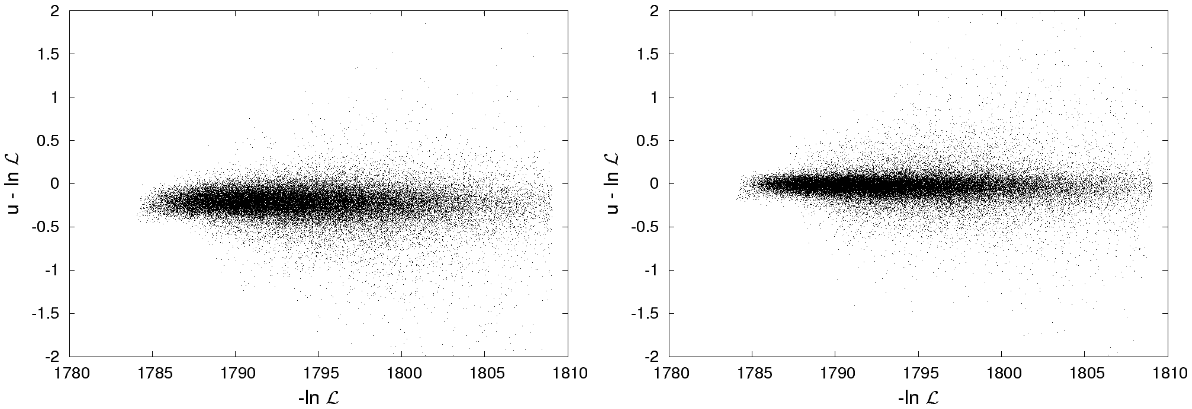


Figure 5.8: Absolute error of the interpolation with respect to the real log-likelihood in 7 dimensions for an interpolation with a sparse grid of level 6 (left panel) and of level 7 (right panel) for normal parameters.

chains. Note that it is not important for the accuracy of the interpolation that the intervals are well-centred at the maximum. Note further that we have also used this set of points as a test set for comparing our interpolation with the real log-likelihood.

5.3.4 Results

We have interpolated the log-likelihood of the WMAP 5 year data in the 6-dimensional normal parameter space described in section 5.3.2. The same has been done for a 7-dimensional model, in which we have chosen the running of the spectral index of the primordial power spectrum, α, as an additional parameter. Constructing the interpolation can be parallelised to an arbitrary degree, according to the available computational resources.

For the 6-dimensional model, we plot the absolute error of the log-likelihood, $(u - \ln \mathcal{L})$, against the negative WMAP log-likelihood, $(-\ln \mathcal{L})$, for the points in the test set in Fig. 5.7. We have used an interpolation on a sparse grid of level $n = 5$ (consisting of 2561 grid points) in the left panel, and of level $n = 6$ (consisting of 10625 grid points) in the right panel. One clearly sees the improvement in accuracy when increasing the grid level from $n = 5$ to $n = 6$. Fig. 5.8 shows the same plot for the 7-dimensional model, for grid level $n = 6$ (18943 grid points) in the left panel and $n = 7$

(78079 grid points) in the right panel. We again see the improvement in accuracy with increasing refinement level. However, the additional parameter α is quite strongly correlated with many of the other parameters, whereas the correlations between the normal parameters in 6 dimensions are reduced to a minimum. We therefore have to increase the grid level by one in 7 dimensions, in order to obtain results comparable to the 6-dimensional ones. In both figures, we note a systematic negative offset of the interpolation with respect to the real function, which becomes less severe for the higher refinement levels. This offset is due to the fact that we construct a d-linear interpolant of a convex function, which systematically lies below the function. This could be easily coped with by adding a small offset to $\alpha_{\mathbf{1},\mathbf{1}}$ after the interpolation, or, even better, by using piecewise polynomial instead of the piecewise linear basis functions. We leave the usage of piecewise polynomial basis functions, which promise to be well-adapted to the log-likelihood, for future work.

Note that we have restricted the plot range to [-2,2], because only 0.1% or less of the points lie outside this range.[5] Almost all of these points lie in the corners of Ω due to relatively strongly correlated parameters. These are the regions in parameter space where the 25 log-likelihood range around the peak extends to the interval boundaries. Due to the extrapolation we use close to the boundaries (cf. the end of section 5.2), we obtain relatively large uncertainties in those regions, which do not affect the one-dimensional projections of the likelihood function, though. The uncertainties can be further reduced, spending (adaptively) more grid points in those regions, see also the discussion about adaptivity in section 5.3.5.

For the 6-dimensional interpolation with a sparse grid of level 6, 2.5% of the test points have an absolute error > 0.25 in the log-likelihood, and 0.03% of the test points have an absolute error > 1. In 7 dimensions and for refinement level 7, the corresponding numbers are 9% and 0.5%, respectively. This is a higher level of accuracy as reached by Pico (Fendt & Wandelt 2007), for which about 90 per cent of the points in a region reaching out to 25 log-likelihoods around the peak have been calculated with an absolute error below 0.25. However, we note that these numbers for Pico are valid for a 9-dimensional parameter space, whereas we work in 6- and 7-dimensional spaces and leave the extension to higher-dimensional models for future work. But we also note that in all settings where a systematic offset in the interpolation error can be observed, it is sufficient to reduce the offset to improve our results significantly, in particular for interpolations on lower levels (see, e.g., the scatterplot for the 6-dimensional model and grid level 5, Fig. 5.7).

We have projected both the interpolation and the WMAP likelihood function using MCMCs of about 150,000 points, and compare the results for the 6-dimensional model for grid level $n = 5$ in Fig. 5.9. We reproduce the projected one-dimensional likelihood curves almost perfectly. The results for the 7-dimensional model for $n = 6$ are shown in Fig. 5.10. Again, the one-dimensional curves are reproduced with great accuracy. The visual comparison of our results with the projected one-dimensional likelihoods obtained by CosmoNet (Auld et al. 2007, 2008) shows that we reproduce the original curves of both the 6-dimensional and the 7-dimensional model with a higher accuracy than the latter. Note also that our interpolation is constructed in a rather wide region, encompassing about 25 log-likelihoods around the peak, whereas in Auld et al. (2008) the region in which $\ln \mathcal{L}$ was fitted covers only 4σ around the peak for the combined likelihood of CMB and LSS. This corresponds to a region of about 8 log-likelihoods around the peak for the combined likelihood, and even less when using only the CMB likelihood.

Consider now the interpolation of the WMAP likelihood surface using directly the standard

---

[5] In 6 dimensions, the number of points outside this range is 0.02% (0.003%) for $n = 5$ ($n = 6$); in 7 dimensions, it is about 0.1% for both grid levels.

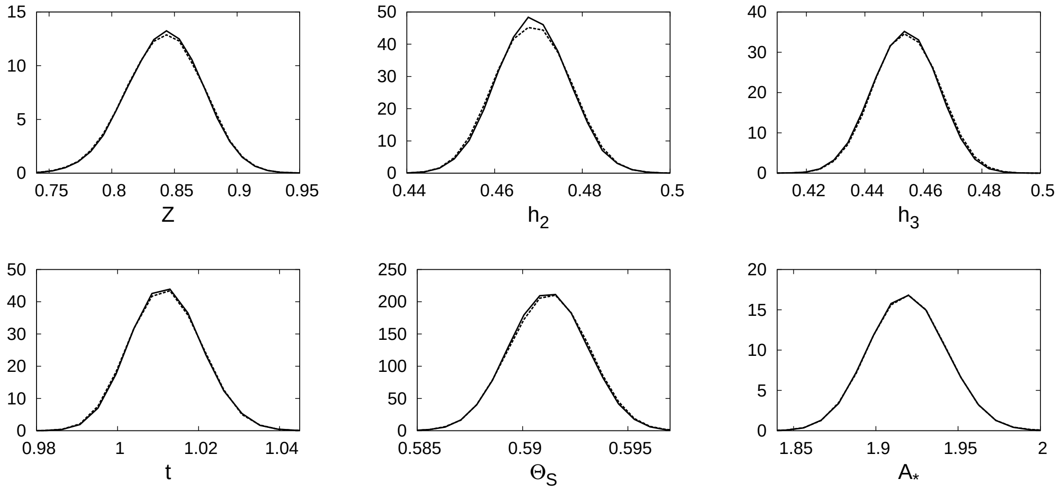


Figure 5.9: Comparison of the one-dimensional projections of the 6-dimensional WMAP 5 year likelihood (solid) and its interpolation (dashed) using a sparse grid of level $n = 5$ (consisting of 2561 grid points) for normal parameters. The curves match almost perfectly.

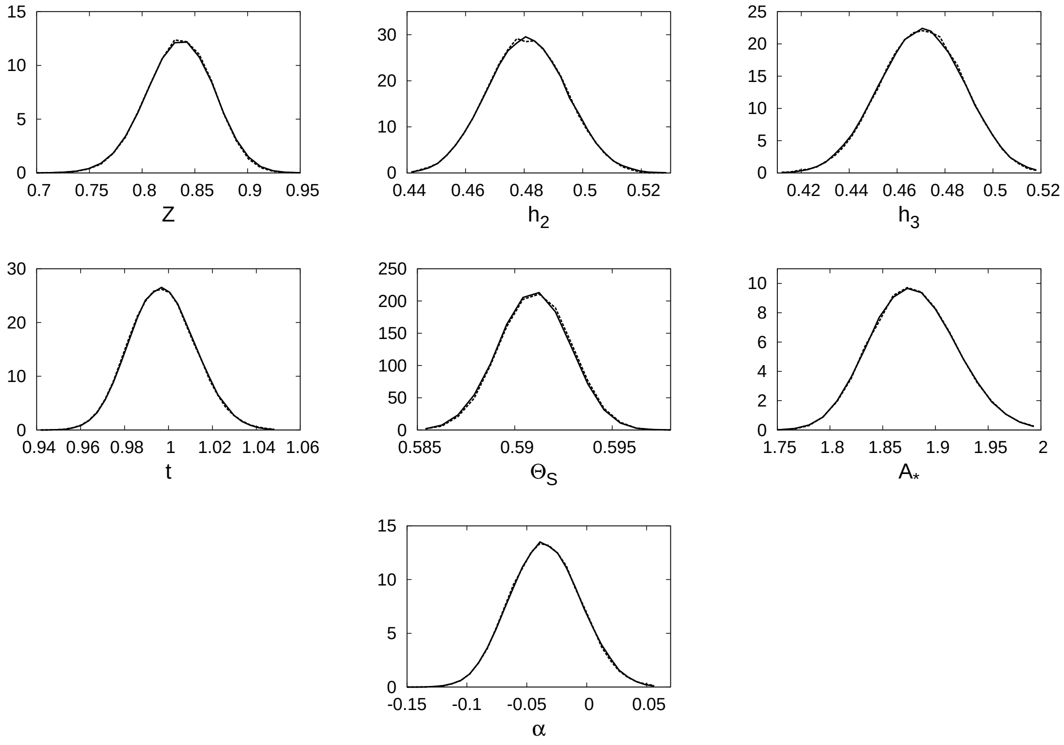


Figure 5.10: Comparison of the one-dimensional projections of the 7-dimensional WMAP 5 year likelihood (solid) and its interpolation (dashed) using a sparse grid of level $n = 6$ (consisting of 18943 grid points) for normal parameters.

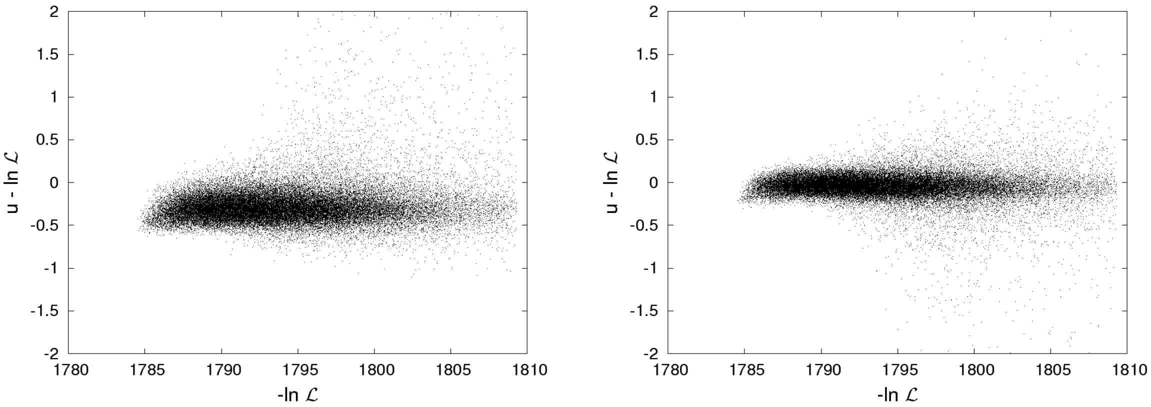


Figure 5.11: Absolute error of the interpolation with respect to the real log-likelihood in 6 dimensions for an interpolation with a sparse grid of level 6 (left panel) and of level 7 (right panel) for standard parameters.

parameters, which are used by default when doing cosmological parameter sampling with the MCMC driver from CMBEASY (Doran & Müller 2004): $\{w_m, w_b, h, \tau, n_s, \ln(10^{10}A_s) - 2\tau\}$, to which we again add α as an additional parameter in the 7-dimensional case. Working with these parameters has the advantage that we do not have to restrict ourselves to the parameter range in which the parameter transformation is invertible. However, the problem is now less adapted to our choice of basis functions, due to the stronger correlations between the different parameters. We therefore pay the price of having to increase the grid level by one in this case in order to reach an accuracy as good as before. We show the absolute error of the log-likelihood, $(u - \ln \mathcal{L})$, against the negative WMAP log-likelihood, $(-\ln \mathcal{L})$, for the 6-dimensional model for grid level $n = 6$ (10625 points) and $n = 7$ (40193 points) in Fig. 5.11,[6] and for the 7-dimensional one for $n = 7$ (87079 points) and $n = 8$ (297727 points) in Fig. 5.12,[7] For the 6-dimensional (7-dimensional) interpolation with a sparse grid of level $n = 7$ ($n = 8$), the fraction of test points with absolute error > 0.25 in the log-likelihood is 6% (20%), and 0.5% (2.5%) for an absolute error > 1. The one-dimensional projections for the 6-dimensional case for level $n = 6$ and for the 7-dimensional case for level $n = 7$ are presented in Figs 5.13 and 5.14, respectively.

We have tested the evaluation time of our interpolation by evaluating a sparse grid interpolant of level 6 in 6 dimensions for 2,000,000 points randomly chosen from within Ω. On a conventional desktop computer (Intel chipset, 2.8 GHz), this took about $92\,\mu s$ per point, including the random generation of the point. In 7 dimensions on the same level we have twice as many grid points and one dimension more, which doubles the evaluation time to $189\,\mu s$. For CosmoNet and Pico, the evaluation of a 6-dimensional model is specified to take about $10\,\mu s$ and $250\,\mu s$, respectively (Auld et al. 2008). Note that we do not know on which hardware the evaluation times of CosmoNet and Pico have been measured, which makes a comparison hardly possible. Note further that our code to evaluate a sparse grid function is not optimised for fast evaluation times and that there is still room for improvement. In any case, for all of these codes the bottleneck in cosmological parameter sampling is now the MCMC algorithm itself rather than the evaluation of the likelihood, at least with the MCMC driver used here (Doran & Müller 2004).

Note that in 7 dimensions, we need significantly more grid points than in 6 dimensions, since the

---

[6] Here, about 0.3% (0.2%) of the points in the test set lie outside the chosen plot-range for the grid of level $n = 6$ ($n = 7$).

[7] About 1% of the points in the test set lie outside the chosen plot-range for the grid of both level $n = 7$ and level $n = 8$.

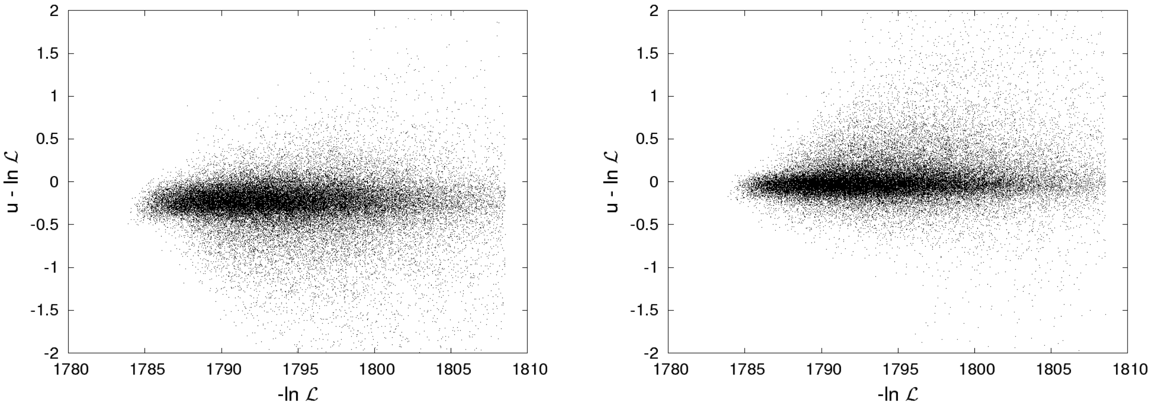


Figure 5.12: Absolute error of the interpolation with respect to the real log-likelihood in 7 dimensions for an interpolation with a sparse grid of level 7 (left panel) and of level 8 (right panel) for standard parameters.

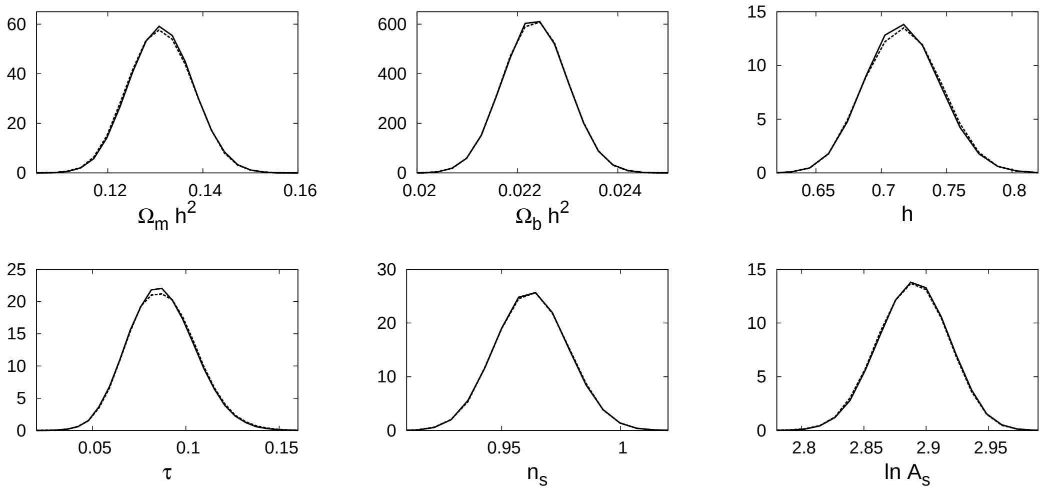


Figure 5.13: Comparison of the one-dimensional projections of the 6-dimensional WMAP 5 year likelihood (solid) and its interpolation (dashed) using a sparse grid of level $n = 6$ (consisting of 10625 grid points) for standard parameters.

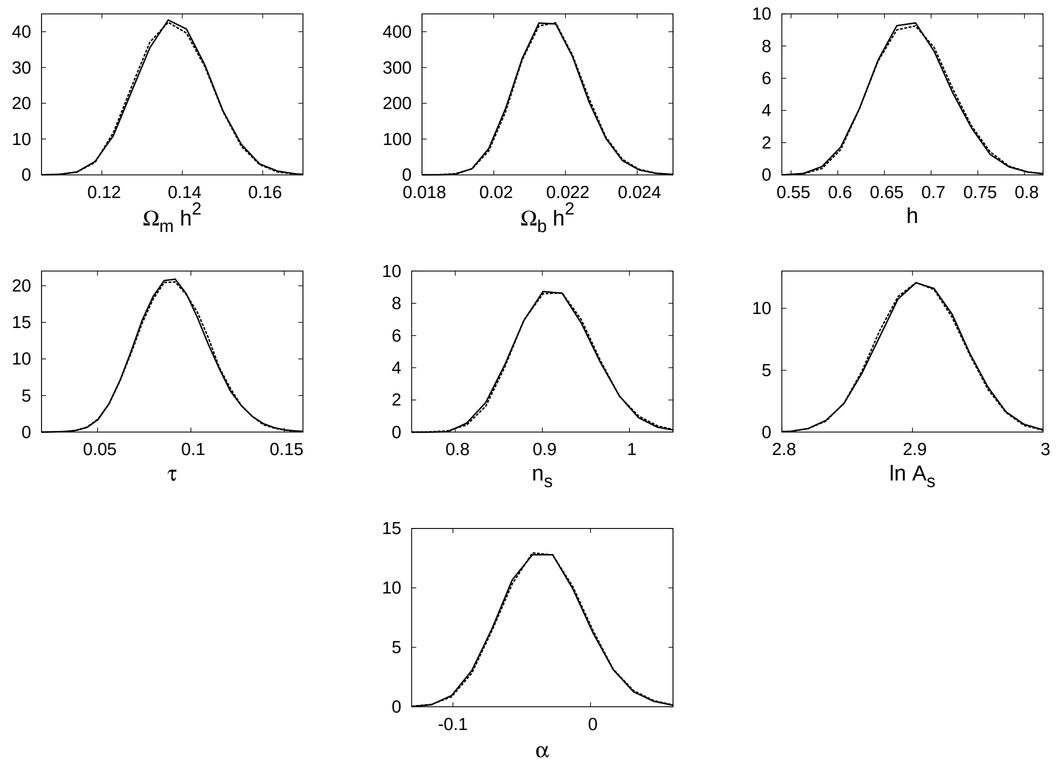


Figure 5.14: Comparison of the one-dimensional projections of the 7-dimensional WMAP 5 year likelihood (solid) and its interpolation (dashed) using a sparse grid of level $n = 7$ (consisting of 78079 grid points) for standard parameters.

additional parameter α strongly correlates with the other parameters, and we thus need to increase the grid level by one to obtain good results. As the storage requirements are rather low, this mainly increases the number of evaluations that are needed for constructing the sparse grid interpolation. As already stated before, though, the construction of the interpolation can be parallelised to an arbitrary degree, according to available computational resources, so that this point should not be an issue. To store the interpolant for a regular sparse grid in d dimensions for level l with N grid points, we would only need N doubles for the coefficients and two integers to remember both d and l, leading to $(N + 1)8$ Bytes. For adaptively refined sparse grids, we additionally have to store at least which grid points have been refined, requiring slightly more storage. For current hardware architectures, the size of the memory is therefore not a limiting factor for our application.

5.3.5 Improvements with adaptive sparse grids

As it has already been mentioned, the log-likelihood is not a perfect sum of one-dimensional functions. The different parameters contribute differently to $\mathcal{L}$ and correlate more or less with each other. It is therefore reasonable, especially when using the standard parameters which correlate more, to employ adaptivity, spending more grid points in critical regions and less grid points elsewhere. In this section, we demonstrate the utility of adaptivity by showing some first results as a proof of concept. As they can clearly be improved further, we leave a thorough study of adaptive sparse grids for the interpolation of $\ln \mathcal{L}$ for future work.

Employing adaptivity, one can attempt to either obtain better results fixing roughly the number of grid points used, or to achieve a similar accuracy using less grid points. In the following, we show results for the former, tackling the 7-dimensional example using the standard parameters on level 7 with 78079 grid points presented above. We start with a regular sparse grid of some low level and refine grid points, creating all $2d$ children in the hierarchical structure (if possible) each, until the grid size exceeds 78000 grid points. In settings where the contributions of the dimensionalities differ significantly, it can be useful to start with level 2 to allow dimensional adaptivity, neglecting unimportant dimensions; here, the grid points on low levels will be created in any case, so we can start with a sparse grid on level 5, e.g., to save on the number of adaptive steps.

Choosing a suitable refinement criterion, it can be determined whether to refine in a broad way (close to regular sparse grids) or in a more greedy way in the sparse grid's hierarchical tree structure. It is reasonable to take the surplusses of the grid points into account as they contain the local information about the functions, i.e., if the function has a high gradient locally. Furthermore, they decay quickly with increasing level-sum in the convergent phase. The mere surplus-based criterion, refining the grid points with the highest absolute value of the surplus first, is known to tend to minimise the L_2-norm of the error. As we do not spend grid points on the domain's boundary, and as we are extrapolating towards the boundary, the biggest surplusses per level can be found for the modified basis functions which are adjacent to the boundary. A mere surplus-based criterion will therefore only refine towards the boundary. This reduces the error especially for sampling points with a high error in the scatterplots, as they are located towards the boundary of the domain.

In the following, we theoretically derive a refinement criterion which is better suited to our problem than the purely surplus-based one. In order to maximise the information our interpolation contains about the real likelihood, we attempt to minimise the Kullback-Leibler distance d_{KL}

between the interpolation and the likelihood function,

$$\begin{aligned} d_{KL} &\equiv \int d^d x\, \mathcal{L}(\boldsymbol{x}) \ln \frac{\mathcal{L}(\boldsymbol{x})}{\exp(u(\boldsymbol{x}))} \\ &= \int d^d x\, \mathcal{L}(\boldsymbol{x}) \left(\ln \mathcal{L}(\boldsymbol{x}) - u(\boldsymbol{x})\right) , \end{aligned} \tag{5.31}$$

which is defined for two normalised probability distributions $\mathcal{L}$ and $\exp(u)$. Let us now derive the refinement criterion we obtain from minimising d_{KL}. Assume that we have already computed an interpolation $u(\boldsymbol{x})$ with N grid points, then the Kullback-Leibler distance d_{KL} when evaluating the function at an additional point $\boldsymbol{x}_{N+1}$ is changed by

$$\begin{aligned} \left|d_{KL}^{\text{new}} - d_{KL}^{\text{old}}\right| &= \left|\int d^d x\, \mathcal{L}(\boldsymbol{x}) \left[\ln \mathcal{L}(\boldsymbol{x}) - u^{\text{new}}(\boldsymbol{x}) - \ln \mathcal{L}(\boldsymbol{x}) + u^{\text{old}}(\boldsymbol{x})\right]\right| \\ &= \left|\int d^d x\, \mathcal{L}(\boldsymbol{x}) \left[u^{\text{old}}(\boldsymbol{x}) - u^{\text{new}}(\boldsymbol{x})\right]\right| \\ &= \left|\int d^d x\, \mathcal{L}(\boldsymbol{x}) \left[\sum_{i=1}^{N} \alpha_i \varphi_i(\boldsymbol{x}) - \sum_{i=1}^{N+1} \alpha_i \varphi_i(\boldsymbol{x})\right]\right| \\ &= \left|\int d^d x\, \mathcal{L}(\boldsymbol{x}) \left[\alpha_{N+1} \varphi_{N+1}(\boldsymbol{x})\right]\right| . \end{aligned} \tag{5.32}$$

If we refine the interpolation around the grid point that contributed most to the Kullback-Leibler distance, we can hope to converge towards the minimum of d_{KL} fastest. In order to obtain a suitable refinement criterion, we have to simplify the formula in eq. (5.32) considerably. We thus assume the likelihood $\mathcal{L}(\boldsymbol{x})$ as well as the basis function $\varphi_{N+1}(\boldsymbol{x})$ to be locally constant on φ_{N+1}'s support, obtaining

$$\left|d_{KL}^{\text{new}} - d_{KL}^{\text{old}}\right| \sim V_{N+1} \mathcal{L}(\boldsymbol{x}_{N+1}) \left|\alpha_{N+1}\right| , \tag{5.33}$$

where V_{N+1} is the volume covered by the basis function φ_{N+1} (i.e. its support), and we have used $\varphi_{N+1}(\boldsymbol{x}_{N+1}) = 1$.

With eq. (5.33), we have derived an estimation of the contribution of a basis function to d_{KL}, which is a reasonable refinement criterion in our setting. In addition to the surplus of the grid point, $|\alpha_{N+1}|$, it takes into account the value of the likelihood $\mathcal{L}(\boldsymbol{x}_{N+1})$ at the grid point, and the volume of the basis function V_{N+1}. The likelihood takes care of the fact that we would like to be more accurate where the likelihood is higher. The regions of very low likelihood are less interesting for us—the likelihood being already very close to zero beyond a difference of about 20 log-likelihoods. The volume factor, on the other hand, prevents the interpolation to refine too deeply (to very high grid levels) locally in the parameter space. However, since this usually only takes effect after several refinement steps, and as we have restricted the number of grid points, we choose not to include the volume factor but rather to refine several points at the same time, which addresses this issue in an alternative way, and which will be discussed later on. We further choose to introduce a temperature T again, which allows us to weight the likelihood with respect to the surplus and thus to influence how much to refine close to the maximum. The refinement criterion we used in this study is thus

$$\left(\frac{\mathcal{L}(\boldsymbol{x}_{\boldsymbol{l},\boldsymbol{i}})}{\mathcal{L}_{\max}}\right)^{1/T} \left|\alpha_{\boldsymbol{l},\boldsymbol{i}}\right| , \tag{5.34}$$

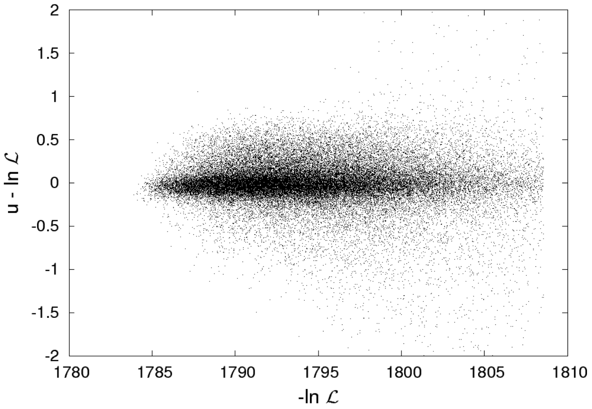


Figure 5.15: Absolute error of the interpolation with respect to the real log-likelihood in 7 dimensions for an interpolation with an adaptively refined sparse grid for standard parameters.

where we have divided the likelihood by its peak value, $\mathcal{L}_{\max}$, (which we have already obtained determining the interpolation domain) because the WMAP code returns the log-likelihood only up to a constant offset, so that we do not know the correct normalisation of $\mathcal{L}$. For $T = 1$, refinement takes place only very close to the maximum as $\mathcal{L}$ decays quickly; a temperature of $T = 6$ showed to provide good results within the whole domain of interest.

Refining only one grid point per refinement step often causes adaptivity to get stuck in a single, special characteristic of the function. Interpolating $\ln\mathcal{L}$ with our choice of basis functions, all grid points are likely to be created only in the direction where the log-likelihood decays fastest, or around one of the local dips we will address later on. Refining more than one grid point at the same time helps to circumvent such effects, resulting in a broader refinement scheme.

The Kullback-Leibler distance d_{KL} can also be used to measure the quality of our interpolation: The distance between the real likelihood and or interpolation should be as small as possible. However, as we have already mentioned above, we do not know the normalisation of the WMAP likelihood function. Therefore, d_{KL} is not necessarily positive and thus looses its property of being a useful measure of the ‘closeness’ of the two functions. We thus use a slightly modified version,

$$\widehat{d_{KL}} \equiv \int d^d x\, \mathcal{L}(\boldsymbol{x}) |\ln\mathcal{L}(\boldsymbol{x}) - u(\boldsymbol{x})|\,, \tag{5.35}$$

as a measure of the quality of our interpolation, instead of the actual Kullback-Leibler distance. It can be easily calculated from an MCMC with $T = 1$ obtained for $\mathcal{L}$, by simply averaging the sum of the absolute errors $|\ln\mathcal{L}(\boldsymbol{x}_i) - u(\boldsymbol{x}_i)|$ over all points. Furthermore, we quote this value averaged over a chain of $T = 3$ (exploiting the interpolation domain better), which corresponds to $\int d^d x\, \mathcal{L}(\boldsymbol{x})^{\frac{1}{T}} |\ln\mathcal{L}(\boldsymbol{x}) - u(\boldsymbol{x})|$.

Fig. 5.15 shows the scatterplot for an adaptively refined sparse grid in 7 dimensions. Starting from a regular grid of level 5, we refined 100 grid points each according to the refinement criterion in eq. (5.34) with $T = 6$. Needing only about as much grid points (78551) as for the regular sparse grid of level 7, Table 5.1 shows that we obtain results which are close to those of a regular sparse grid of level 8 with almost 4 times as many grid points. We provide the Mean Squared Error (MSE) as well as $\widehat{d_{KL}}$ for both $T = 1$ and $T = 3$ chains for regular sparse grids of level 7 and 8, and for the adaptively refined case. We also quote how many points exhibit an absolute error larger than 1 or

	err > 1, T=3	err > 0.25, T=3	MSE, T=1	MSE, T=3	$\widehat{d}_{KL}$, T=1	$\widehat{d}_{KL}$, T=3
level 7	4.2%	50.5%	0.087	0.532	0.256	0.354
level 8	2.3%	19.3%	0.017	0.210	0.091	0.193
adaptive	1.8%	23.6%	0.027	0.202	0.110	0.204

Table 5.1: Comparison of errors of regular sparse grids of level 7 and level 8, respectively, and an adaptively refined sparse grid using approximately as many grid points as contained in the regular grid of level 7. Shown are the number of points with an absolute error larger than 1 or 0.25 in the $T = 3$ chains, the MSE for chains of $T = 1$ and $T = 3$, and $\widehat{d}_{KL}$, which denotes the absolute value of the error averaged over chains of $T = 1$ and $T = 3$.

0.25 for the $T = 3$ chains. We do not show the histograms of the adaptively refined model, as the histograms for both the regular grid on level 8 and the adaptively refined one, are very close to the already very good results presented for level 7 above.

We would like to mention, that, due to numerical problems, the current version of CMBEASY produces local, unphysical and sometimes rather high dips. This problem is already known and will be corrected in the next release. For stochastic approaches, this is not a big problem, though: The dips are local and just cause some noisy evaluations. But it poses a problem for numerical approaches if a grid point hits a dip. Then it can happen, that spending more grid points can even deteriorate the results. For our regular grid in 6 dimensions using the standard parameters, e.g., increasing the level from 7 to 8 caused a higher overall error on the chain-data used for the histograms, as especially two new basis functions close to the peak caused an error of up to 12 of the log-likelihood for all evaluations affected by those basis functions.

Fortunately, dips can be detected automatically due to the hierarchical structure of the sparse grid and the smoothness of $\ln \mathcal{L}$, using a criterion that is once more based on the surplusses. Furthermore, it is not a severe problem when using adaptivity, as adaptivity localises the effects of the dips automatically. One just has to take care not to spend too much grid points to compensate for the dips.

The first adaptive results are promising, but there is still room for a lot of improvement. Even better refinement criteria than those used so far could be employed. Using not only piecewise linear functions, but rather piecewise polynomials, and applying adaptivity in both the mesh-width and the polynomial degree is very promising; especially the extrapolation properties towards the boundary would be improved, and less grid points would be needed to obtain the same accuracies.

5.4 Conclusions

In this chapter, we have explored the utility of interpolating the WMAP log-likelihood surface using sparse grids. We demonstrated that the results are excellent and competitive to other approaches regarding speed and accuracy, and we discussed advantages over fitting the likelihood surface with polynomials (Fendt & Wandelt 2007; Sandvik et al. 2004) or neural networks (Auld et al. 2008):

The interpolation based on sparse grids converges towards the exact function in the limit of the grid level going to infinity. We can therefore reach an arbitrary accuracy by simply increasing the amount of work we spend. In the case of a polynomial fit, this is not guaranteed since increasing the polynomial degree runs the risk of becoming unstable.

In order to construct the sparse grid interpolation, we do not need to sample a set of training points using MCMCs beforehand, since the sampling points are determined by the sparse grid structure which is given a priori. Once we have chosen the volume of interest, the time for constructing the interpolation is dominated by the evaluation of the likelihood function at the grid points. We do not need additional training time as for neural networks (Auld et al. 2008), for example. Constructing the interpolation can thus be done almost arbitrarily in parallel, only limited by the computational resources that are available.

The sparse grid technique is rather general and not restricted to certain classes of functions. In particular, the choice of sampling points and basis functions is not tailored to a single problem as for neural networks, where the properties of the network such as its topology and transfer functions have to be chosen problem-specifically (often in a heuristic way). The sparse grid interpolation technique as well as our extensions can therefore be readily applied to other problems in astrophysics and cosmology, and will be useful in further tasks, where an accurate interpolation of a function is needed.

The excellent performance of the sparse grid interpolation can be further improved, leaving future research to do: It can be applied to models with more than seven parameters by spending more computational effort. Further modification of the basis functions, for example allowing for a piecewise polynomial interpolation, promises better convergence rates and higher accuracies. Adaptive refinement schemes, which take into account the characteristics of the interpolated function, can be used to further increase the accuracy of the interpolation, as we have already demonstrated for a first example.

Chapter 6

Conclusions and outlook

In this PhD work, we have studied fundamental questions of modern cosmology by developing and applying data analysis techniques that combine the CMB temperature data with its polarization data and LSS measurements.

We have derived optimal methods for detecting the ISW effect, which serves as a probe of dark energy and has the power to yield constraints on the dark-energy related cosmological parameters. Our optimal methods for ISW detection extract more information about the ISW effect from the CMB and LSS data than existing methods. Thus, compared to the usual method, the detection significance of our optimal polarization method in an ideal scenario is 16 per cent higher for low redshift surveys such as the SDSS galaxy sample and 23 per cent for surveys going out to a redshift of about 2. With currently available polarization data from WMAP, our method yields approximately the same detection significance as existing methods, due to the high contamination of the polarization data by noise and the low volume coverage of currently available LSS surveys. However, with polarization data from *Planck*, our method will be significantly better than existing methods, where the exact improvement depends on how well Galactic foregrounds can be removed. A very crude estimate yields an improvement of the detection significance of at least 10 per cent. Note that the numbers quoted refer to the average detection significance reached with our method, where the average is taken over all possible realisations of the LSS. The actual detection significance reached with our methods depends on the specific realisation of the LSS in our Universe and can thus be higher (or lower) than the average. Since the standard method does not keep the LSS fixed in the analysis, but uses an average over all possible realisations, the specific realisation of the LSS causes a biasing of the detection significance, which gets stronger for more unlikely LSS realisations.

Our optimal method for ISW detection can be extended to the detection of other secondary anisotropies, such as the kinetic Sunyaev-Zel'dovich effect, the Rees-Sciama effect or gravitational lensing. This extension will require some additional work, since all secondary effects apart from the ISW originate on smaller scales, on which the LSS has already undergone non-linear structure growth and can thus no longer be described by a Gaussian distribution. The extension of our method to a non-Gaussian posterior distribution can be done using the field theoretical techniques described in section 2.3 and in Enßlin et al. (2009).

In the course of developing our optimal method for ISW detection, we have derived the correct form of the joint likelihood of CMB and LSS data for cosmological parameter estimation, consistently including the small coupling between the two data-sets introduced by the ISW effect. Previously, in parameter sampling studies, it has usually been assumed that the joint likelihood factorises into a product of the likelihoods of the respective data-sets. Once LSS surveys cover a large enough volume that we can create an ISW template containing most of the ISW signal, using our likelihood instead will provide better constraints on the dark-energy related cosmological parameters Ω_Λ and the equation of state of dark energy. The ideal survey for this will be EUCLID, which will measure the galaxy distribution for all of the accessible sky outside the Galaxy out to a redshift of $z \sim 2$. However, even for currently available LSS surveys, Ho et al. (2008) and Giannantonio et al. (2008) obtain an effect on cosmological parameter constraints when including the information contained in the ISW effect in the analysis.

The second part of this work was devoted to understanding the origin of the axis of evil, an unusual alignment between the preferred directions of the quadrupole and the octopole in the CMB temperature map. To this end, we have subtracted from the CMB polarization data the fraction of which is correlated with the temperature map, and checked for the axis of evil in the remaining polarization map. This uncorrelated polarization map serves as a statistically independent probe of the axis of evil and other anomalies present in the CMB temperature map, and can thus help to assess whether these anomalies are just due to chance fluctuations in the temperature, or if they have their origin in some preferred direction intrinsic to the geometry of the primordial Universe. We find that the preferred axis of the quadrupole aligns with the axis of evil, whereas that of the octopole does not. However, the contamination of the WMAP polarization data by detector noise and Galactic foregrounds leaves us with a large uncertainty in our results, such that we do not obtain any evidence for or against a preferred direction intrinsic to our Universe. Nevertheless, with the upcoming polarization data from the *Planck* satellite, we will have a powerful tool to probe the axis of evil of the CMB in polarization, thereby assessing the validity of the cosmological principle.

As we have already mentioned above, the method we proposed can be used to probe any anomalies present in the CMB temperature map. Among such anomalies, there is a significant power asymmetry between the northern and southern ecliptic hemisphere in the multipole range $l \approx 2-40$, a strong lack of power in the quadrupole of the temperature map, and some weaker anomalies in the low multipoles beyond the octopole. The analysis of the uncorrelated polarization map of *Planck* will shed more light on the origin on all of these anomalies.

Note that our studies of the axis of evil were phenomenological so far, since not all theoretical models creating anomalies in the temperature map predict the same features for the uncorrelated polarization map. In order to go beyond the phenomenological nature of these studies, it will be necessary to consider particular models of the primordial Universe creating anomalies in the CMB temperature maps, and compute the statistical properties of the uncorrelated polarization map for these by modifying a Boltzmann code such as CMBEASY or by numerical simulations. Using these signatures predicted for the uncorrelated polarization map, one can then compare the different models with each other and with an isotropic Universe via Bayesian model selection. This will permit us to truly go beyond the usual a posteriori analysis of CMB anomalies, and to fully exploit the power of complete CMB data to assess how well the assumption of isotropy holds for our Universe.

In both of the projects described above, the study of the ISW effect and the analysis of anomalies in the CMB, we have studied the imprints of specific phenomena in the CMB. However, the main power of the CMB lies in its ability to provide us with simultaneous constraints on the cosmological parameters. Such parameter estimation studies usually rely on sampling the parameter space using MCMC techniques. The main bottleneck in these studies is the evaluation of the likelihood of the cosmological parameters, a problem which becomes increasingly serious with the ever-growing amount of data we have to handle. In the third part of this work, we have therefore developed and implemented a sparse-grids based technique to interpolate the WMAP likelihood surface, which reduces the evaluation time of the likelihood to a fraction of a millisecond, thus significantly speeding up MCMC sampling. In speed and accuracy, our interpolation technique is competitive to previous attempts to fit the likelihood with polynomials or neural networks, while overcoming some of the drawbacks of the latter. These include, for instance, the danger of creating unphysical wiggles in the fit if the polynomial degree is chosen too high with respect to the number of available training points, or the comparably long training time required for neural networks. Using our technique to interpolate the likelihood surface of *Planck* will significantly simplify the parameter estimation process, especially when combining *Planck* data with other data-sets whose likelihood is less expensive to evaluate, as, for example, in the parameter sampling studies using the correct joint likelihood for CMB and LSS data proposed above. Furthermore, our interpolation algorithm is completely general and can be applied to any function which is sufficiently smooth. We can thus use it to speed up the likelihood evaluation of any large data-set whose likelihood is expensive to evaluate, and apply it to other problems where an accurate interpolation of a function is needed.

Closing words

With this thesis, we have made a tiny step forward on the ever-lasting path towards understanding the world in which we live. Personally, I do not think that the basic picture of cosmology that we have now will be the last word in the history of cosmology, and that the remaining tasks are now merely to better understand the ingredients of the latter. There have been times in the history of physics, when we thought that the big picture is all set, as, for example, in 1874, when Max Planck was advised not to study physics by the Munich physics professor Philipp von Jolly, with the statement that “in this field, almost everything is already discovered, and all that remains is to fill a few holes.” Instead of filling holes, Planck formulated the quantisation of energy emitted by a black body, which marked the beginning of quantum physics, thus revolutionising our view of the world. I expect there will be yet many revolutions like the one initiated by Planck, and we should be prepared to constant changes in our view of the world as we go along and try to understand the Cosmos in which we live.

Appendix A

A.1 Derivation of the Wiener filter

In this section, we will derive in detail the posterior distribution for the free information field theory, eq. (2.26). We start from the data model given in eq. (2.16),

$$d = Rs + n\,, \tag{A.1}$$

and assume the signal prior and the noise distribution to be Gaussian. Note that we now explicitely state the dependence of the probability distributions on the cosmological parameters p. Rewriting the data model as $n = d - Rs$ and inserting this into the noise distribution, we obtain

$$\begin{aligned} \mathcal{P}(d\,|\,s,p) &= \mathcal{P}(d - R\,s\,|\,s,p) \\ &= \mathcal{G}(d - R\,s, N)\,. \end{aligned} \tag{A.2}$$

The joint probability distribution of signal and data, $\mathcal{P}(d,s\,|\,p) = \mathcal{P}(d\,|\,s,p)\,\mathcal{P}(s\,|\,p)$, is thus

$$\begin{aligned} \mathcal{P}(d,s\,|\,p) &= \mathcal{G}(d - R\,s, N)\,\mathcal{G}(s,S) \\ &= \mathcal{G}(s - D\,j, D)\,\mathcal{G}(d, RS\,R^\dagger + N)\,, \end{aligned} \tag{A.3}$$

where we have used the definition of the propagator, eq. (2.20)

$$D \equiv \left(R^\dagger N^{-1} R + S^{-1}\right)^{-1}\,,$$

and of the information source, eq. (2.21)

$$j \equiv R^\dagger N^{-1} d\,.$$

We will prove the second step in eq. (A.3) in section A.1.1. The distribution in eq. (A.3) can be trivially integrated over s in order to obtain the evidence

$$\mathcal{P}(d\,|\,p) = \mathcal{G}(d, RS\,R^\dagger + N)\,. \tag{A.4}$$

Therefore the posterior distribution $\mathcal{P}(s\,|\,d,p) = \mathcal{P}(s,d\,|\,p)/\mathcal{P}(d\,|\,p)$ reads

$$\mathcal{P}(s\,|\,d,p) = \mathcal{G}(s - D\,j, D)\,, \tag{A.5}$$

where

$$s_{\rm rec} \equiv D j = \left(R^\dagger N^{-1} R + S^{-1}\right)^{-1} R^\dagger N^{-1} d\,, \tag{A.6}$$

is the Wiener reconstruction of the signal s.

A.1.1 Lemma 1

We now prove the expression for the joint probability distribution $P(d,\ s\,|\,p)$ given in eq. (A.3). To this end, we start with

$$\begin{aligned} P(d,s\,|\,p) &= \mathcal{G}(d - R\,s, N)\,\mathcal{G}(s, S) \\ &= \frac{1}{\sqrt{|2\pi N||2\pi S|}} \\ &\quad \times \exp\left(-\frac{1}{2}(d - R\,s)^\dagger N^{-1}(d - R\,s)\right) \\ &\quad \times \exp\left(-\frac{1}{2}\,s^\dagger\,S^{-1} s\right)\,. \end{aligned} \tag{A.7}$$

Let us first rewrite the exponent

$$\begin{aligned} & (d - R\,s)^\dagger N^{-1}(d - R\,s) + s^\dagger\,S^{-1} s \\ =\; & s^\dagger\,D^{-1} s - 2\,j^\dagger\,s + d^\dagger\,N^{-1}\,d \\ =\; & (s - Dj)^\dagger D^{-1}(s - Dj) - j^\dagger D\,j + d^\dagger\,N^{-1} d \\ =\; & (s - Dj)^\dagger D^{-1}(s - Dj) + d^\dagger (RS\,R^\dagger + N)^{-1} d\,, \end{aligned} \tag{A.8}$$

where we have used the definitions of D and j, eqs (2.20) and (2.21), in the first step, then completed the square in the second step, and we will separately prove the last step as Lemma 2 in the next subsection. After doing that we will prove that

$$|2\pi N||2\pi S| = |2\pi D||2\pi(RS\,R^\dagger + N)|\,, \tag{A.9}$$

which we name Lemma 3, allowing us to reformulate eq. (A.7) as

$$\begin{aligned} P(d,s\,|\,p) &= \frac{1}{\sqrt{|2\pi D||2\pi(RS\,R^\dagger + N)|}} \\ &\quad \times \exp\left(-\frac{1}{2}(s - Dj)^\dagger D^{-1}(s - Dj)\right) \\ &\quad \times \exp\left(-\frac{1}{2}\,d^\dagger\,(RS\,R^\dagger + N)^{-1} d\right)\,, \end{aligned} \tag{A.10}$$

which is what we claimed in eq. (A.3).

Note that from the second line in eq. (A.8), and adding the contribution of the determinants in eq. (A.7), we readily obtain the Hamiltonian of the free theory, eq. (2.19):

$$\begin{aligned} H_{\mathcal{G}}[s] &= -\log\left[\mathcal{P}(d|s)\,\mathcal{P}(s)\right] \\ &= \frac{1}{2} s^\dagger D^{-1} s - j^\dagger s + H_0^{\mathcal{G}}\,, \end{aligned} \tag{A.11}$$

with

$$H_0^{\mathcal{G}} = \frac{1}{2}\, d^\dagger\, N^{-1}\, d + \frac{1}{2}\, \log\left(|2\,\pi\, S\,|\,|2\,\pi\, N|\right)\ . \tag{A.12}$$

A.1.2 Lemma 2

In this subsection we prove that

$$j^\dagger D\, j - d^\dagger\, N^{-1} d = -d^\dagger\, (RS\,R^\dagger + N)^{-1} d\,. \tag{A.13}$$

In order to simplify the notation, let us introduce

$$M \equiv R^\dagger N^{-1} R\,. \tag{A.14}$$

It can be easily seen that eq. (A.13) is equivalent to

$$N^{-1} R\,(S^{-1} + M)^{-1} R^\dagger\, N^{-1} - N^{-1} = -(RS\,R^\dagger + N)^{-1} \tag{A.15}$$

by inserting the respective expressions for D and j. We start with eq. (A.15) and transform it into an equation which is true.

$$\begin{aligned}
& N^{-1} R\,(S^{-1} + M)^{-1} R^\dagger\, N^{-1} - N^{-1} = -(RS\,R^\dagger + N)^{-1} \\
\Longleftrightarrow\quad & RS\,R^\dagger N^{-1} R(S^{-1} + M)^{-1} R^\dagger N^{-1} + R(S^{-1} + M)^{-1} R^\dagger N^{-1} - RS\,R^\dagger N^{-1} - 1 = -1 \\
\Longleftrightarrow\quad & RS\,[M(S^{-1} + M)^{-1} + (1 + MS\,)^{-1} - 1] R^\dagger N^{-1} = 0 \\
\Longleftrightarrow\quad & RS\,[MS\,(1 + MS\,)^{-1} + (1 + MS\,)^{-1} - 1] R^\dagger N^{-1} = 0 \\
\Longleftrightarrow\quad & RS\,[(1 + MS\,)(1 + MS\,)^{-1} - 1] R^\dagger N^{-1} = 0\,.
\end{aligned} \tag{A.16}$$

This equation is true, QED

A.1.3 Lemma 3

In the following we prove that

$$|2\pi N||2\pi S\,| = |2\pi D||2\pi(RS\,R^\dagger + N)|\,, \tag{A.17}$$

which is equivalent to

$$|N||S\,| = |D||RS\,R^\dagger + N|\,, \tag{A.18}$$

for the factors of 2π cancel for matrices that operate on the same vector space. Let us write

$$\begin{aligned}
\frac{|N||S|}{|D|} &= |N||S||D^{-1}| \\
&= |N||S\,D^{-1}| \\
&= |N||S\;(S^{-1} + R^{\dagger} N^{-1} R)| \\
&= |N| \exp\left(\log|1 + S\,R^{\dagger} N^{-1} R|\right) \\
&= |N| \exp\left(\mathrm{Tr}\;\log(1 + \mathrm{SR}^{\dagger}\mathrm{N}^{-1}\mathrm{R})\right) \\
&= |N| \exp\left(\mathrm{Tr}\;\log(1 + \mathrm{RSR}^{\dagger}\mathrm{N}^{-1})\right) \\
&= |N| \exp\left(\log|1 + RS\,R^{\dagger} N^{-1}|\right) \\
&= |N||RS\,R^{\dagger} N^{-1} + 1| \\
&= |(RS\,R^{\dagger} N^{-1} + 1)N| \\
&= |RS\,R^{\dagger} + N|\,.
\end{aligned} \tag{A.19}$$

The crucial step here was to use the cyclic invariance of the trace Tr and to notice that this cyclic invariance still holds for the trace of a logarithm, which can be easily verified using the Taylor expansion of the logarithm.

Appendix B

B.1 Proof of the equivalence of the number of bins

We now outline the proof that if one uses the correct kernel, i.e. the ISW kernel rather than the kernel for the galaxy density contrast in the analysis, the estimated amplitude $\widehat{A}_{cc}$ and the variance σ_{cc}^2 are independent of the number of bins chosen, provided that all bins together cover the whole volume relevant for the ISW effect. The proof here is done only for the variance, but follows the same scheme for the estimated amplitude. The total variance σ_{cc}^2 one obtains when working with N bins is given by eq. (3.32), where we have substituted the index *gi* by ISW(i), following the argument of section 3.3.2:

$$\sigma_{cc}^2 = \left[\sum_l (2l+1) \sum_{i,j} C_l^{\mathrm{isw}(i),T} (C_l^{cc})^{-1}(i,j) C_l^{\mathrm{isw}(j),T}\right]^{-1} . \tag{B.1}$$

We then use the form of the covariance matrix given by eq. (3.28) and the following relations that only hold for the ISW kernel:

$$C_l^{\mathrm{isw}(i),T} = \sum_{j=1}^{N} C_l^{\mathrm{isw}(i),\mathrm{isw}(j)} \tag{B.2}$$

$$C_l^{\mathrm{isw}} = \sum_{j=1}^{N} C_l^{\mathrm{isw}(j),T} . \tag{B.3}$$

Now we choose a fixed but arbitrary number of bins N, invert the covariance matrix and by inserting the above relations we obtain

$$\sum_{i,j} C_l^{\mathrm{isw}(i),T} (C_l^{cc})^{-1}(i,j) C_l^{\mathrm{isw}(j),T} = \frac{C_l^{\mathrm{isw}}}{C_l^{\mathrm{isw}} + C_l^T} . \tag{B.4}$$

Inserting this into eq. (B.1), the resulting formula for σ_{cc}^2 is exactly what we obtain from one single bin covering the whole volume relevant for the ISW effect. We have checked this explicitly for $N = 2 - 5$ and it is straightforward, though timely, to also check it for any other number of bins.

B.2 Proof of the factorization of the likelihood

We now explicitely prove the factorization of the likelihood in eq. (3.67) into a reduced temperature part and a polarisation part, as given in eq. (3.74). We will do this for the more general case that we not only have a signal template T_τ for the temperature part, but also a non-zero template P_τ for the polarisation part. In this case, the covariance matrix is

$$\tilde{C} = \begin{pmatrix} C_{\Delta T} & C_{\Delta T,\Delta P} \\ C_{\Delta P,\Delta T} & C_{\Delta P} \end{pmatrix}, \tag{B.5}$$

instead of the simplified one given in eq. (3.68). Here, ΔP and ΔT are defined as $\Delta P \equiv P - P_\tau$ and $\Delta T \equiv T - T_\tau$, respectively. A block matrix

$$\tilde{C} \equiv \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{B.6}$$

with A and D being invertible square matrices, can be blockwise inverted as given by the following formula:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C\,(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C\,(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}. \tag{B.7}$$

Let us define a a generalised version of the reduced temperature map and covariance matrix, which we had introduced in eq. (3.72):

$$\begin{aligned} T^{\mathrm{rec}}_{\mathrm{uncorr}} &\equiv T - C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P \\ C_{\mathrm{uncorr}} &\equiv A - BD^{-1}C \equiv C_{\Delta T} - C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,C_{\Delta P,\Delta T}\,. \end{aligned} \tag{B.8}$$

Using these definitions and the blockwise matrix inversion, we first rewrite the exponent of $\mathcal{G}(d - \tau, \tilde{C})$ in eq. (3.67):

$$\begin{aligned} & \left(\Delta T^\dagger, \Delta P^\dagger\right)\tilde{C}^{-1}\begin{pmatrix} \Delta T \\ \Delta P \end{pmatrix} \\ = \;& \Delta T^\dagger C^{-1}_{\mathrm{uncorr}}\Delta T - \Delta T^\dagger C^{-1}_{\mathrm{uncorr}}C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P - \Delta P^\dagger C^{-1}_{\Delta P}C_{\Delta P,\Delta T}C^{-1}_{\mathrm{uncorr}}\Delta T \\ & + \Delta P^\dagger\left(C_{\Delta P} + C^{-1}_{\Delta P}C_{\Delta P,\Delta T}C^{-1}_{\mathrm{uncorr}}C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\right)\Delta P \\ = \;& \Delta T^\dagger C^{-1}_{\mathrm{uncorr}}\Delta T - \Delta T^\dagger C^{-1}_{\mathrm{uncorr}}\left(C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P\right) - \left(C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P\right)^\dagger C^{-1}_{\mathrm{uncorr}}\Delta T \\ & + \Delta P^\dagger C_{\Delta P}\,\Delta P + \left(C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P\right)^\dagger C^{-1}_{\mathrm{uncorr}}\left(C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P\right) \\ = \;& \left(\Delta T - C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P\right)^\dagger C^{-1}_{\mathrm{uncorr}}\left(\Delta T - C_{\Delta T,\Delta P}\,C^{-1}_{\Delta P}\,\Delta P\right) + \Delta P^\dagger C^{-1}_{\Delta P}\Delta P \\ = \;& (T^{\mathrm{rec}}_{\mathrm{uncorr}} - T_\tau)^\dagger\,C^{-1}_{\mathrm{uncorr}}\,(T^{\mathrm{rec}}_{\mathrm{uncorr}} - T_\tau) + \Delta P^\dagger C^{-1}_{\Delta P}\Delta P\,, \end{aligned} \tag{B.9}$$

where we have used $C^\dagger_{\Delta T,\Delta P} = C_{\Delta P,\Delta T}$, $C^\dagger_{\Delta T} = C_{\Delta T}$, and $C^\dagger_{\Delta P} = C_{\Delta P}$, and have completed the square in the second last step. Similarly, using the factorization of the determinant of a block matrix,

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = |D|\,|A - BD^{-1}C|\,, \tag{B.10}$$

we can rewrite the determinant of the covariance matrix $\tilde{C}$ as

$$|\tilde{C}| = |C_{\text{uncorr}}|\,|C_{\Delta P}|\,. \tag{B.11}$$

Inserting eqs (B.9) and (B.11) into $\mathcal{G}(d-\tau,\tilde{C})$, allows us to write

$$\mathcal{G}(d-\tau,\tilde{C}) = \mathcal{G}(T^{\text{rec}}_{\text{uncorr}} - T_\tau, C_{\text{uncorr}})\,\mathcal{G}(P - P_\tau, C_{\Delta P})\,. \tag{B.12}$$

In the case of the polarisation template P_τ being zero, this expression reduces to the one given in eq. (3.74).

Appendix C

C.1 Proof of vanishing correlation between $T^{\rm rec}_{\rm corr}$ and $T^{\rm rec}_{\rm uncorr}$

We now prove that the two maps $T^{\rm rec}_{\rm corr}$ and $T^{\rm rec}_{\rm uncorr}$, into which we split the temperature map T in section 4.2, are indeed uncorrelated. Let us first look at the following covariance matrices, which we will need in the derivation:

$$C_P \equiv \langle P P^\dagger \rangle_{\mathcal{P}(P\,|\,p)} = W S_P W^\dagger + N_P \,, \tag{C.1}$$

where we have assumed that $P_{\rm cmb}$ is uncorrelated with $P_{\rm det}$ and $P_{\rm fg}$, and we have inserted the definition of N_P, eq. (4.5).

Since we neglect the detector noise and residual foregrounds in the temperature data, we obtain for the covariance between temperature and polarization data

$$\begin{aligned} C_{T,P} \equiv \langle T P^\dagger \rangle_{\mathcal{P}(T,P\,|\,p)} &= \langle T P_{\rm cmb}{}^\dagger \rangle_{\mathcal{P}(T,P_{\rm cmb}\,|\,p)} W^\dagger \\ &\equiv S_{T,P} W^\dagger \,, \end{aligned} \tag{C.2}$$

where we have assumed that detector noise and residual foregrounds in the polarization map are uncorrelated with the CMB temperature map.

We now write

$$\begin{aligned} T^{\rm rec}_{\rm corr} &= S_{T,P}\, S_P^{-1} P^{\rm rec}_{\rm cmb} \\ &= S_{T,P} S_P^{-1} (S_P^{-1} + W^\dagger N_P^{-1} W)^{-1} W^\dagger N_P^{-1} P \\ &= S_{T,P} (1 + W^\dagger N_P^{-1} W S_P)^{-1} W^\dagger N_P^{-1} P \\ &= S_{T,P} W^\dagger (1 + N_P^{-1} W S_P W^\dagger)^{-1} N_P^{-1} P \\ &= S_{T,P} W^\dagger (N_P + W S_P W^\dagger)^{-1} P \\ &= C_{T,P} C_P^{-1} P \,, \end{aligned} \tag{C.3}$$

where we have inserted $P^{\rm rec}_{\rm cmb}$ from eq. (4.7) in the first step. The third step can be easily verified by using the geometric series for $\left(1 + W^\dagger N_P^{-1} W S_P\right)^{-1} W^\dagger$, which has a convergence radius of 1, and is thus valid for $|W^\dagger N_P^{-1} W S_P| < 1$. In our case, this holds because our polarization data are noise-dominated.[1] Note that we have just proven that the Wiener reconstruction of $P_{\rm cmb}$ translated into a temperature map, eq. (4.9), is equivalent to $T^{\rm rec}_{\rm corr}$ as given in eq. (3.73). We have proven it here for a zero signal template, $T_\tau = 0$, but the proof for a non-zero signal template can be done in the same way.

---

[1] By adding a small ϵ-term to the response W, and thus making it invertible, the third step also holds generally.

Let us now look at

$$\begin{aligned}
\langle T^{\rm rec}_{\rm uncorr} T^{\rm rec\,\dagger}_{\rm corr}\rangle_{\mathcal{P}(T,P\,|\,p)} &= \langle T T^{\rm rec\,\dagger}_{\rm corr}\rangle_{\mathcal{P}(T,P\,|\,p)} - \langle T^{\rm rec}_{\rm corr} T^{\rm rec\,\dagger}_{\rm corr}\rangle_{\mathcal{P}(T,P\,|\,p)} \\
&= \langle T P^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} (N_P + W S_P W^\dagger)^{-1} W S_{P,T} \\
&\quad - S_{T,P} W^\dagger (N_P + W S_P W^\dagger)^{-1} \langle P P^\dagger\rangle_{\mathcal{P}(P\,|\,p)} \\
&\qquad (N_P + W S_P W^\dagger)^{-1} W S_{P,T} \\
&= S_{T,P} W^\dagger (N_P + W S_P W^\dagger)^{-1} W S_{P,T} \\
&\quad - S_{T,P} W^\dagger (N_P + W S_P W^\dagger)^{-1} (N_P + W S_P W^\dagger) \\
&\qquad (N_P + W S_P W^\dagger)^{-1} W S_{P,T} \\
&= S_{T,P} W^\dagger (N_P + W S_P W^\dagger)^{-1} W S_{P,T} \\
&\quad - S_{T,P} W^\dagger (N_P + W S_P W^\dagger)^{-1} W S_{P,T} \\
&= 0\,, \qquad\qquad (\text{C.4})
\end{aligned}$$

where we have inserted eqs (C.3), (C.1), and (C.2). This proves that $T^{\rm rec}_{\rm uncorr}$ and $T^{\rm rec}_{\rm corr}$ are uncorrelated. QED

C.2 Proof of vanishing correlation between $P_{\rm corr}$ and $P^{\rm rec}_{\rm uncorr}$

For the splitting of the polarization map, we first prove that the unfiltered uncorrelated map defined in eq. (4.15), $P^{\rm raw}_{\rm uncorr}$, is uncorrelated with $P_{\rm corr}$:

$$\begin{aligned}
\langle P^{\rm raw}_{\rm uncorr} P_{\rm corr}{}^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} &= \langle P P_{\rm corr}{}^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} - W \langle P_{\rm corr} P_{\rm corr}{}^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} \\
&= \langle P T^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} S_T^{-1} S_{T,P} \\
&\quad - W S_{P,T} S_T^{-1} \langle T T^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} S_T^{-1} S_{T,P} \\
&= W S_{P,T} S_T^{-1} S_{T,P} - W S_{P,T} S_T^{-1} S_T S_T^{-1} S_{T,P} \\
&= W S_{P,T} S_T^{-1} S_{T,P} - W S_{P,T} S_T^{-1} S_{T,P} \\
&= 0\,. \qquad\qquad (\text{C.5})
\end{aligned}$$

From the above, we readily obtain that also the Wiener filtered uncorrelated map,

$$\begin{aligned}
P^{\rm rec}_{\rm uncorr} &= [(S_P - S_{P,T} S_T^{-1} S_{T,P})^{-1} + W^\dagger N_P^{-1} W]^{-1} \\
&\quad W^\dagger N_P^{-1} P^{\rm raw}_{\rm uncorr}\,,
\end{aligned}$$

is uncorrelated with $P_{\rm corr}$:

$$\begin{aligned}
\langle P^{\rm rec}_{\rm uncorr} P_{\rm corr}{}^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} &= [(S_P - S_{P,T} S_T^{-1} S_{T,P})^{-1} + W^\dagger N_P^{-1} W]^{-1} \\
&\quad W^\dagger N_P^{-1} \langle P^{\rm raw}_{\rm uncorr} P_{\rm corr}{}^\dagger\rangle_{\mathcal{P}(T,P\,|\,p)} \\
&= 0\,. \qquad\qquad (\text{C.6})
\end{aligned}$$

QED

C.3 Probability for chance alignment in an isotropic universe

We would like to assess whether the rough alignment of the axis of the quadrupole in $P^{\mathrm{rec}}_{\mathrm{uncorr}}$ actually provides us with some information about the axis of evil. We therefore compute the probability for at least one of the axes of $P^{\mathrm{rec}}_{\mathrm{uncorr}}$ aligning with the axis of the temperature map in an isotropic universe. To this end, let us take the preferred axis in the temperature map T as given, and assume that the axes of $P^{\mathrm{rec}}_{\mathrm{uncorr}}$ are distibuted isotropically over the sky and independently from each other. We then work out the probability for at least one of the axes of $P^{\mathrm{rec}}_{\mathrm{uncorr}}$ being such that the axis of the temperature map is included in the 1σ region around it.

For simpliticy, we assume that the the 1σ regions are symmetric circles around the axes, with radius $\sigma \approx 45^\circ$ for both the quadrupole and the octopole. The solid angle A spanned by such a 1σ region is well approximated by $A \approx \pi\sigma^2$.[2] The probability of at least one of the 1σ regions hitting the axis of evil is just the solid angle spanned by the two 1σ regions divided by the solid angle of the hemisphere, 2π. However, the solid angle spanned by the two 1σ regions depends on the overlap B between them, it is $2A - B$ to avoid double counting of the overlapping area. Given the angular separation α between the axes of the quadrupole and the octopole, the overlap can be computed as follows:

$$B(\alpha) = 2\left[\sigma^2 \arccos\left(\frac{\alpha}{2\sigma}\right) - \frac{\alpha}{2}\sqrt{\sigma^2 - \frac{\alpha^2}{4}}\right], \tag{C.7}$$

which can be derived from the geometry of the problem in flat-sky approximation. We marginalize the hitting probability over the overlap $B(\alpha)$, using the fact that α is distributed as $\mathcal{P}(\alpha) = \sin(\alpha)$ (de Oliveira-Costa et al. 2004):

$$\begin{aligned} \mathcal{P}(\mathrm{hit}) &= \int_{\alpha=0}^{\pi/2} \mathcal{P}(\mathrm{hit}\,|\,B(\alpha))\mathcal{P}(\alpha)\,d\alpha \\ &= \int_{\alpha=0}^{\pi/2} \frac{2A - B(\alpha)}{2\pi}\sin(\alpha)\,d\alpha \approx 50\,\%\,. \end{aligned} \tag{C.8}$$

---

[2] This flat-sky approximation differs from the actual value of the solid angle by 6 per cent.

Appendix D

D.1 Inversion of the parameter transformation

In the following, we present a technique of inverting the parameter transformation of Sec. 5.3.2 to compute the cosmological parameters given the normal parameters. The normal parameter h_2 in terms of cosmological parameters is given by

$$h_2(w_m, w_b) = 0.0264\, w_b^{-0.762} \exp\left(-0.476\, [\ln(25.5\, w_b + 1.84\, w_m)]^2\right) . \tag{D.1}$$

We solve this equation for w_m as a first step:

$$w_m(h_2, w_b) \quad = \quad \left(\exp\left\{\pm\left[-\frac{1}{0.476}\ln\left(\frac{h_2}{0.0264}\, w_b^{0.762}\right)\right]^{1/2}\right\} - 25.5\, w_b\right)\frac{1}{1.84} . \tag{D.2}$$

Inconveniently, there exist two different solutions for $w_m(h_2, w_b)$, which complicates the inversion. We now substitute w_m in $h_3(w_m, w_b)$ (5.27) for (D.2) and thus obtain $h_3(h_2, w_b)$, which, of course, has two solutions as well. An example of the two branches of $h_3(h_2, w_b)$ for $h_2 = 0.45$ is depicted in Fig. D.1. We can calculate the critical point where only one solution exists using the condition

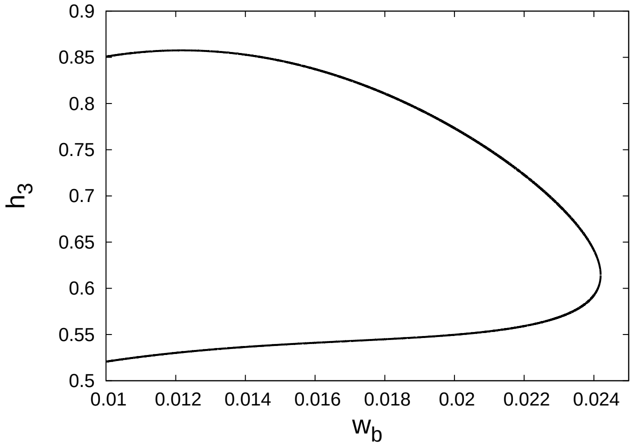


Figure D.1: The two branches of h_3 versus w_b for $h_2 = 0.45$.

$$- \frac{1}{0.476} \ln \left(\frac{h_2}{0.0264} \, w_b^{0.762} \right) = 0 \,, \tag{D.3}$$

as can be seen from (D.2). This condition gives us the following formulae for the parameter values at the critical point:

$$\begin{aligned}
w_{b,\mathrm{crit}}(h_2) &= \left(\frac{0.0264}{h_2} \right)^{1/0.762} \,, && \text{(D.4)} \\
w_{m,\mathrm{crit}}(h_2) &= (1 - 25.5 \, w_{b,\mathrm{crit}}) \, \frac{1}{1.84} \,, && \text{(D.5)} \\
h_{3,\mathrm{crit}}(h_2) &= 2.17 \left(1 + \left(\frac{w_{b,\mathrm{crit}}}{0.044} \right)^2 \right)^{-1} w_{m,\mathrm{crit}}^{0.59} \left(1 + 1.63 \left(1 - \frac{w_{b,\mathrm{crit}}}{0.071} \right) w_{m,\mathrm{crit}} \right)^{-1} \,. && \text{(D.6)}
\end{aligned}$$

The two parameters h_2 and h_3 can now be inverted to w_m and w_b. For a given h_2, we express h_3 in terms of h_2 and w_b, as described above. We then use $h_{3,\mathrm{crit}}(h_2)$ to choose the upper branch of $h_3(h_2, w_b)$ if our given h_3 is bigger than $h_{3,\mathrm{crit}}(h_2)$, and the lower branch if it is smaller. Using the respective branch of $h_3(h_2, w_b)$, we search numerically in w_b until $h_3(h_2, w_b)$ matches the given h_3. Substituting that value of w_b into equation (D.2), we readily obtain the value for w_m.

Now it is straightforward to compute the values for n_s and A_s from t and A_*. To obtain h from Θ_s, we follow the procedure suggested by Kosowsky et al. (2002), expressing Θ_s in terms of h in terms of h and then searching in h numerically until $\Theta_s(h)$ matches the given value of Θ_s.

Appendix E

Acknowledgements

First of all, I would like to thank my supervisor, Torsten Enßlin, for being a great supervisor, for having almost infinite resources of patience, new ideas, good advice, and for always being there when you need him. Thank you very much, it was great to work with you!

I would also like to thank Simon White for having hosted me and given me the opportunity to do this PhD at MPA. I am very happy that I had this great chance. I am also indebted to Simon for contributing to my work with very good ideas and critical questions, which significantly improved my work.

I wish to thank Dirk Pflüger for a great and fruitful collaboration, and for pushing our work on the sparse grids to its current stage. I would also like to thank Dirk's colleagues from the computer science department, Prof. Hans-Joachim Bungartz and Dr. Stefan Zimmer, for their support on the subject.

I am deeply indebted to Martin Reinecke, who is the most powerful master of programming I have ever met, and who solved all of my computing problems within at most five minutes each. I thank Thomas Riller for a great collaboration and for many enlightening discussions about statistics and mathematics, and Wolfgang Hovest for enlightening discussions about life in general. Furthermore, I would like to thank the rest of the Planck group at MPA for helping out with random programming questions, and especially the Planck-newcomer Georg Robbers for giving me a premium support contract for CMBEASY.

I am grateful to Carlos Hernández-Monteagudo, Anthony Banday, André Waelkens, Mariapaola Bottino, Jens Jasche, Francisco Kitaura, Zhen Hou, Cheng Li, Valeria Pettorino, Iain Brown, Stuart Sim, Mike Boylan-Kolchin, Henrik Junklewitz, Cornelius Weig, Martin Obergaulinger, and Eiichiro Komatsu for useful discussions and help on various subjects.

I thank the amochette and all my friends at MPA and MPE, I had a great time with you and I am very glad to have met you all. Special thanks to: André Waelkens for being André and for the Muh, Caroline d'Angelo for being a cool person and for her help with all sorts of English questions, and the two of them for the numerous trips to the machine of power. Shy Genel for being a great friend and for lots of enlightening discussions about life, Mariapaola Bottino for her help with difficult decisions, Silvia Bonoli for being mamma oca and for lots of fun, and Margarita Petkova and Payel Das for the good times. I thank my office mates, Federico Stasyszyn, Qi Guo, and Francesco de Gasperin for the nice atmosphere, Umberto Maio for the extensive Italian lessons, Daniele Pierini for teaching me everything that Umberto refused to say, and all the other Italians for teaching me this gorgeous language.

I would like to thank Johann Sebastian Bach and Franco Battiato, without whom I would have

probably gone crazy at certain stages of this PhD.

My very special thanks goes to my parents, Irene and Siegfried Frommert, for being great parents, and to my boyfriend, Boris Schulze, for being who he is and for his support in the last three years, which have not always been so easy.

Bibliography

Abramo, L. R., Bernui, A., Ferreira, I. S., Villela, T., & Wuensche, C. A. 2006, Phys. Rev. D, 74, 063506

Ackerman, L., Carroll, S. M., & Wise, M. B. 2007, Phys. Rev. D, 75, 083502

Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297

Afshordi, N. 2004, Phys. Rev. D, 70, 083536

Afshordi, N., Loh, Y.-S., & Strauss, M. A. 2004, Phys. Rev. D, 69, 083524

Alpher, R. A. & Herman, R. 1948, Nature, 162, 774

Armendariz-Picon, C., Mukhanov, V., & Steinhardt, P. J. 2001, Phys. Rev. D, 63, 103510

Auld, T., Bridges, M., & Hobson, M. P. 2008, MNRAS, 387, 1575

Auld, T., Bridges, M., Hobson, M. P., & Gull, S. F. 2007, MNRAS, 376, L11

Ayaita, Y., Weber, M., & Wetterich, C. 2009, arXiv:0905.3324

Bardeen, J. M. 1980, Phys. Rev. D, 22, 1882

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15

Basak, S., Hajian, A., & Souradeep, T. 2006, Phys. Rev. D, 74, 021301

Baumann, D., Cooray, A., Dodelson, S., et al. 2008, arXiv:0811.3911

Bayes, T. 1763, Phil. Trans. Roy. Soc., 53, 370

Bekenstein, J. D. 2004, Phys. Rev. D, 70, 083509

Bellman, R. 1961, Adaptive Control Processes: A Guided Tour (Princeton University Press)

Bernui, A. 2008, Phys. Rev. D, 78, 063531

Bernui, A., Villela, T., Wuensche, C. A., Leonardi, R., & Ferreira, I. 2006, A&A, 454, 409

Binney, J., Dowrick, N., Fisher, A., & Newman, M. 1992, The theory of critical phenomena (Oxford University Press, Oxford, UK: ISBN0-19-851394-1)

Boldt, E. 1987, Phys. Rep., 146, 215

Bolstad, W. 2004, Introduction to Bayesian Statistics (John Wiley & Sons, Hoboken, NJ)

Bondi, H. & Gold, T. 1948, MNRAS, 108, 252

Boughn, S. & Crittenden, R. 2004, Nature, 427, 45

Boughn, S. P. & Crittenden, R. G. 2005, New Astronomy Review, 49, 75

Boughn, S. P., Crittenden, R. G., & Turok, N. G. 1998, New Astronomy, 3, 275

Buchert, T. 2008, General Relativity and Gravitation, 40, 467

Bungartz, H.-J. & Griebel, M. 2004, Acta Numerica, 13, 147

Bungartz, H.-J., Pflüger, D., & Zimmer, S. 2008, in Modelling, Simulation and Optimization of Complex Processes, Proceedings of the High Performance Scientific Computing 2006, Hanoi, Vietnam, ed. H. Bock, E. Kostina, X. Hoang, & R. Rannacher (Springer), 121–130

Cabré, A., Fosalba, P., Gaztañaga, E., & Manera, M. 2007, MNRAS, 381, 1347

Caldwell, R. R. 2002, Physics Letters B, 545, 23

Caldwell, R. R. & Stebbins, A. 2008, Physical Review Letters, 100, 191302

Castaño-Martínez, A. & López-Blázquez, F. 2007, Test, 14, 397

Célérier, M.-N., Bolejko, K., Krasiński, A., & Hellaby, C. 2009, arXiv:0906.0905

Chib, S. & Greenberg, E. 1995, The American Statistician, 49, 327

Chu, M., Kaplinghat, M., & Knox, L. 2003, ApJ, 596, 725

Cole, S. et al. 2005, MNRAS, 362, 505

Coles, P. & Lucchin, F. 2002, Cosmology: The Origin and Evolution of Cosmic Structure, Second Edition (John Wiley & Sons)

Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693

Cooray, A. 2002a, Phys. Rev. D, 65, 103510

Cooray, A. 2002b, Phys. Rev. D, 65, 083518

Cooray, A. & Melchiorri, A. 2006, Journal of Cosmology and Astro-Particle Physics, 1, 18

Copi, C. J., Huterer, D., Schwarz, D. J., & Starkman, G. D. 2006, MNRAS, 367, 79

Copi, C. J., Huterer, D., Schwarz, D. J., & Starkman, G. D. 2007, Phys. Rev. D, 75, 023507

Copi, C. J., Huterer, D., & Starkman, G. D. 2004, Phys. Rev. D, 70, 043515

Crittenden, R. 2006, `http://www-astro-theory.fnal.gov/Conferences/ECcmbC/PresentationFiles/RobertCrittenden.ppt`, `http://www-astro-theory.fnal.gov/Conferences/ECcmbC/ECcmbCagenda.html`

Crittenden, R. G. & Turok, N. 1996, Physical Review Letters, 76, 575

Cruz, M., Tucci, M., Martínez-González, E., & Vielva, P. 2006, MNRAS, 369, 57

de Oliveira-Costa, A., Tegmark, M., Zaldarriaga, M., & Hamilton, A. 2004, Phys. Rev. D, 69, 063516

Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, ApJ, 142, 414

Dodelson, S. 2003, Modern cosmology (Academic Press, San Diego)

Dolag, K., Hansen, F. K., Roncarelli, M., & Moscardini, L. 2005, MNRAS, 363, 29

Doran, M. 2005, Journal of Cosmology and Astro-Particle Physics, 10, 11

Doran, M. & Müller, C. M. 2004, Journal of Cosmology and Astro-Particle Physics, 9, 3

Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, ApJS, 180, 306

Durrer, R. 2001, Journal of Physical Studies, 5, 177

Dvorkin, C., Peiris, H. V., & Hu, W. 2008, Phys. Rev. D, 77, 063008

Dyson, F. W., Eddington, A. S., & Davidson, C. 1920, Royal Society of London Philosophical Transactions Series A, 220, 291

Einstein, A. 1916, Annalen der Physik, 354, 769

Einstein, A. 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 142

Enqvist, K. 2008, General Relativity and Gravitation, 40, 451

Enßlin, T. A., Frommert, M., & Kitaura, F. S. 2009, accepted by Phys. Rev. D, arXiv:0806.3474

Erdoğdu, P., Lahav, O., Zaroubi, S., et al. 2004, MNRAS, 352, 939

Eriksen, H. K., Banday, A. J., Górski, K. M., Hansen, F. K., & Lilje, P. B. 2007, ApJL, 660, L81

Feix, M., Fedeli, C., & Bartelmann, M. 2008, A&A, 480, 313

Fendt, W. A. & Wandelt, B. D. 2007, arXiv:0712.0194

Fisher, K. B., Lahav, O., Hoffman, Y., Lynden-Bell, D., & Zaroubi, S. 1995, MNRAS, 272, 885

Fosalba, P. & Gaztañaga, E. 2004, MNRAS, 350, L37

Fosalba, P., Gaztañaga, E., & Castander, F. J. 2003, ApJL, 597, L89

Frommert, M. & Enßlin, T. A. 2009a, MNRAS, 395, 1837

Frommert, M. & Enßlin, T. A. 2009b, arXiv:0908.0453

Frommert, M., Enßlin, T. A., & Kitaura, F. S. 2008, MNRAS, 391, 1315

Gamerman, D. 1997, Markov Chain Monte Carlo (Chapman & Hall, London)

Gamow, G. 1948, Nature, 162, 680

Ganapathysubramanian, B. & Zabaras, N. 2007, J. Comput. Phys., 225, 652

Gao, X. 2009, arXiv:0903.1412

Garcke, J. 2006, in ICML ’06: Proceedings of the 23rd international conference on Machine learning (New York, NY, USA: ACM Press), 321–328

Garcke, J., Griebel, M., & Thess, M. 2001, Computing, 67, 225

Garcke, J. & Hegland, M. 2009, Computing, 84, 1

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. 2004, Bayesian data analysis (Boca Raton, Florida: Chapman & Hall/CRC)

Giannantonio, T., Scranton, R., Crittenden, R. G., et al. 2008, Phys. Rev. D, 77, 123520

Gold, B., Bennett, C. L., Hill, R. S., et al. 2009, ApJS, 180, 265

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Groeneboom, N. E. & Eriksen, H. K. 2009, ApJ, 690, 1807

Gruppuso, A. & Burigana, C. 2009, Journal of Cosmology and Astro-Particle Physics, 8, 4

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Guth, A. H. & Weinberg, E. J. 1983, Nuclear Physics B, 212, 321

Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., & Williams, B. 2007, Phys. Rev. D, 76, 083503

Hansen, F. K., Banday, A. J., & Górski, K. M. 2004, MNRAS, 354, 641

Hansen, F. K., Banday, A. J., Górski, K. M., Eriksen, H. K., & Lilje, P. B. 2009, ApJ, 704, 1448

Hernández-Monteagudo, C. 2008, A&A, 490, 15

Hinshaw, G. et al. 2007, ApJS, 170, 288

Hinshaw, G. et al. 2009, ApJS, 180, 225

Ho, S., Hirata, C., Padmanabhan, N., Seljak, U., & Bahcall, N. 2008, Phys. Rev. D, 78, 043519

Hoftuft, J., Eriksen, H. K., Banday, A. J., et al. 2009, ApJ, 699, 985

Holtz, M. 2008, Dissertation, Institut für Numerische Simulation, Universität Bonn

Hu, W., Fukugita, M., Zaldarriaga, M., & Tegmark, M. 2001, ApJ, 549, 669

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Hwang, J. 1990a, Classical and Quantum Gravity, 7, 1613

Hwang, J.-C. 1990b, Phys. Rev. D, 42, 2601

Inoue, K. T. & Silk, J. 2007, ApJ, 664, 650

Jackson, J. D. 1975, Classical electrodynamics (Wiley, New York)

Jaffe, A. H. 2003, New Astronomy Review, 47, 1001

Jaffe, T. R., Banday, A. J., Eriksen, H. K., Górski, K. M., & Hansen, F. K. 2005, ApJL, 629, L1

Jaffe, T. R., Banday, A. J., Eriksen, H. K., Górski, K. M., & Hansen, F. K. 2006, A&A, 460, 393

Jarrett, T. H., Chester, T., Cutri, R., et al. 2000, AJ, 119, 2498

Jimenez, R., Verde, L., Peiris, H., & Kosowsky, A. 2004, Phys. Rev. D, 70, 023005

Kaplinghat, M., Knox, L., & Skordis, C. 2002, ApJ, 578, 665

Kasai, M., Asada, H., & Futamase, T. 2006, Progress of Theoretical Physics, 115, 827

Kitaura, F. S. & Enßlin, T. A. 2008, MNRAS, 389, 497

Kitaura, F. S., Jasche, J., Li, C., et al. 2009, arXiv:0906.3978

Klimke, A., Nunes, R., & Wohlmuth, B. 2006, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems, 14, 561

Klypin, A. & Prada, F. 2009, ApJ, 690, 1488

Kodama, H. & Sasaki, M. 1984, Progress of Theoretical Physics Supplement, 78, 1

Kolb, E. W., Marra, V., & Matarrese, S. 2008, Phys. Rev. D, 78, 103002

Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330

Kosowsky, A., Milosavljevic, M., & Jimenez, R. 2002, Phys. Rev. D, 66, 063007

Land, K. & Magueijo, J. 2005, Physical Review Letters, 95, 071301

Land, K. & Magueijo, J. 2007, MNRAS, 378, 153

Lemaître, G. 1931, Nature, 127, 706

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Linde, A. D. 1982, Physics Letters B, 108, 389

Liu, X. & Zhang, S.-N. 2006, ApJL, 636, L1

Loverde, M., Hui, L., & Gaztañaga, E. 2007, Phys. Rev. D, 75, 043519

Lue, A., Scoccimarro, R., & Starkman, G. 2004, Phys. Rev. D, 69, 044005

Magueijo, J. & Sorkin, R. D. 2007, MNRAS, 377, L39

Martineau, P. & Brandenberger, R. H. 2005, Phys. Rev. D, 72, 023507

Martínez-González, E., Cruz, M., Cayón, L., & Vielva, P. 2006, New Astronomy Review, 50, 875

Masi, S., Ade, P. A. R., Bock, J. J., et al. 2007, New Astronomy Review, 51, 236

Maturi, M., Dolag, K., Waelkens, A., Springel, V., & Enßlin, T. 2007, A&A, 476, 83

Mavromatos, N. E., Sakellariadou, M., & Yusaf, M. F. 2009, Phys. Rev. D, 79, 081301

McEwen, J. D., Vielva, P., Hobson, M. P., Martínez-González, E., & Lasenby, A. N. 2007, MNRAS, 376, 1211

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087

Milgrom, M. 1983, ApJ, 270, 365

Misner, C. W., Thorne, K. S., & Wheeler, J. A. 1973, Gravitation (Freeman and Company, San Francisco)

Morales, J. A. & Sáez, D. 2008, ApJ, 678, 583

Mukhanov, V. 2005, Physical foundations of cosmology (Physical foundations of cosmology, by V. Mukhanov. Cambridge, UK: Cambridge University Press, 2005)

Naselsky, P. D., Christensen, P. R., Coles, P., et al. 2007, arXiv:0712.1118

Neal, R. M. 1993, in Technical Report CRG-TR-93-1 (University of Toronto: Dept. of Computer Science)

Nolta, M. R., Dunkley, J., Hill, R. S., et al. 2009, ApJS, 180, 296

Page, L., Hinshaw, G., Komatsu, E., et al. 2007, ApJS, 170, 335

Page, L., Nolta, M. R., Barnes, C., et al. 2003, ApJS, 148, 233

Peacock, J. A. 1999, Cosmological Physics (Cambridge University Press, Cambridge)

Peebles, P. J. E. & Ratra, B. 1988, ApJL, 325, L17

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419

Percival, W. J. et al. 2004, MNRAS, 353, 1201

Pereira, T. S. & Abramo, L. R. 2009, Phys. Rev. D, 80, 063525

Perlmutter, S. et al. 1999, ApJ, 517, 565

Pietrobon, D., Balbi, A., & Marinucci, D. 2006, Phys. Rev. D, 74, 043524

Pogosian, L., Corasaniti, P. S., Stephan-Otto, C., Crittenden, R., & Nichol, R. 2005, Phys. Rev. D, 72, 103519

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in C. The art of scientific computing (Cambridge: University Press, —c1992, 2nd ed.)

Prunet, S., Uzan, J., Bernardeau, F., & Brunier, T. 2005, Phys. Rev. D, 71, 083508

Raccanelli, A., Bonaldi, A., Negrello, M., et al. 2008, MNRAS, 386, 2161

Rakić, A. & Schwarz, D. J. 2007, Phys. Rev. D, 75, 103002

Räsänen, S. 2004, Journal of Cosmology and Astro-Particle Physics, 2, 3

Rassat, A., Land, K., Lahav, O., & Abdalla, F. B. 2007, MNRAS, 377, 1085

Reichardt, C. L., Ade, P. A. R., Bock, J. J., et al. 2009, ApJ, 694, 1200

Reisinger, C. & Wittum, G. 2007, SIAM J. Scientific Computing, 29, 440

Riess, A. G. et al. 1998, AJ, 116, 1009

Robert, C. P. 2001, The Bayesian choice (New York: Springer-Verlag)

Sachs, R. K. & Wolfe, A. M. 1967, ApJ, 147, 73

Samal, P. K., Saha, R., Jain, P., & Ralston, J. P. 2008, MNRAS, 385, 1718

Samal, P. K., Saha, R., Jain, P., & Ralston, J. P. 2009, MNRAS, 396, 511

Sandvik, H. B., Tegmark, M., Wang, X., & Zaldarriaga, M. 2004, Phys. Rev. D, 69, 063005

Schwarz, D. J., Starkman, G. D., Huterer, D., & Copi, C. J. 2004, Physical Review Letters, 93, 221301

Seljak, U. & Zaldarriaga, M. 1996, ApJ, 469, 437

Sievers, J. L., Mason, B. S., Weintraub, L., et al. 2009, arXiv:0901.4540

Souradeep, T., Hajian, A., & Basak, S. 2006, New Astronomy Review, 50, 889

Spergel, D. N. & et al. 2007, ApJS, 170, 377

Sunyaev, R. A. & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Tauber, J. A. 2000, Astrophysical Letters Communications, 37, 145

Tegmark, M. et al. 2004, Phys. Rev. D, 69, 103501

Temlyakov, V. N. 1993, Approximation of Periodic Functions (Nova Science Publishers)

Tomita, K. 2001, MNRAS, 326, 287

Verde, L., Peiris, H. V., Spergel, D. N., et al. 2003a, ApJS, 148, 195

Verde, L. et al. 2003b, ApJS, 148, 195

Vielva, P., Martínez-González, E., Barreiro, R. B., Sanz, J. L., & Cayón, L. 2004, ApJ, 609, 22

Vielva, P., Martínez-González, E., & Tucci, M. 2006, MNRAS, 365, 891

Vielva, P., Wiaux, Y., Martínez-González, E., & Vandergheynst, P. 2007, MNRAS, 381, 932

von Petersdorff, T. & Schwab, C. 2006, Appl. Math., 51, 145

Wetterich, C. 1988, Nuclear Physics B, 302, 668

Wiaux, Y., Vielva, P., Martínez-González, E., & Vandergheynst, P. 2006, Physical Review Letters, 96, 151303

Widmer, G., Hiptmair, R., & Schwab, C. 2008, Journal of Computational Physics, 227, 6071

Wiener, N. 1950, Math. Rev., 11, 118

Wright, E. L., Meyer, S. S., Bennett, C. L., et al. 1992, ApJL, 396, L13

Zaldarriaga, M. 1997, Phys. Rev. D, 55, 1822

Zaldarriaga, M. & Seljak, U. 1997, Phys. Rev. D, 55, 1830

Zaroubi, S. 1995, in Clustering in the Universe, ed. S. Maurogordato, C. Balkowski, C. Tao, & J. Tran Thanh van, 135–+

Zaroubi, S., Hoffman, Y., Fisher, K. B., & Lahav, O. 1995, ApJ, 449, 446

Zenger, C. 1991, in Notes on Numerical Fluid Mechanics, Vol. 31, Parallel Algorithms for Partial Differential Equations, ed. W. Hackbusch (Vieweg), 241–251

Zhang, P. 2006a, ApJ, 647, 55

Zhang, P. 2006b, Phys. Rev. D, 73, 123504

Mona Frommert

Born on the 08 March 1980 in Böblingen, Germany

Research Interests

- The integrated Sachs-Wolfe effect as a probe of dark energy
- Anomalies in the CMB
- Cosmological parameter sampling
- Non-linear data analysis

Research Experience and Education

2006-2009	**PhD with Dr. Torsten Enßlin and Prof. Simon D. M. White,** Max-Planck-Institut für Astrophysik (MPA) *Temperature and Polarization studies of the cosmic microwave background*
2005	**Diplom with Prof. Christof Wetterich,** Institut für Theoretische Physik, Universität Heidelberg *Coupled quintessence*
2002-2003	**Exchange student,** University of Victoria, Canada
2003	**Directed studies with Prof. Werner Israel,** University of Victoria, Canada *Cosmology and growth of structure*

Scholarships and Awards

Student fellowship, **Studienstiftung des deutschen Volkes**, 2002-2005

Student fellowship, **German Academic Exchange Service (DAAD)** 2002-2003

Student grant, **E-Fellows**, since 2005

Award for the best performance in chemistry, **Chemical Industry Fund**, 1999

Awarded to the best student in chemistry of every German high-school.

Teaching and Other Relevant Experience

Tutor, Theoretical Electrodynamics, 2nd year undergraduate physics, 2005

Student representative on the scientific council, MPA, 2006-2008

Contributions to the MPA-exhibition at the Deutsches Muesum: "CMB simulator" and timeline of the history of our Universe, 2009

Girl's Day, oral presentation and discussions with high-school students, Apr 2007-2009

Member of the first-aid team at MPA, since 2007

Internship at Siemens, programming, Dec 2005 - Mar 2006

Summer student, Max-Planck-Institut für Metallforschung, Stuttgart, 1999

Languages

German (native language), English (fluent), Italian (advanced proficiency), French (intermediate proficiency)

Refereed publications

- *Optimal integrated Sachs-Wolfe detection and joint likelihood for cosmological parameter estimation*
  Frommert M., Enßlin T. A., Kitaura F.-S.
  2008, MNRAS, 391, 1315

- *Ironing out primordial temperature fluctuations with polarization - optimal detection of cosmic structure imprints*
  Frommert M., Enßlin T. A.
  2009, MNRAS, 395, 1837

- *Information field theory for cosmological perturbation reconstruction and signal analysis*
  Enßlin T. A., **Frommert M.**, Kitaura F.-S.
  accepted for publication in Phys Rev D, arXiv:0806.3474

- *The axis of evil* - *a polarization perspective*
  Frommert M., Enßlin T. A.
  submitted to MNRAS, arXiv:0908.0453

- *Efficient cosmological parameter sampling using sparse grids*
  Frommert M., Pflüger D., Riller T., Reinecke M., Bungartz H.-J., Enßlin T. A.
  to be submitted to MNRAS