A Quadratic Estimator for Statistical Anisotropy in the Cosmic Microwave Background

Master of Science thesis by
Svein Rune Saxrud
Institute of Theoretical Astrophysics
University of Oslo

June, 2011

Abstract

The standard model of cosmology has proven to be a great success. Through measurements of the Cosmic Microwave Background radiation, cosmologists have been able to constrain parameters such as the age and geometry of the universe, or the amounts of different types of energy, to high accuracy. The inflationary theory of the model states that the anisotropies in the background radiation should be distributed so that they follow statistical isotropy and homogeneity. However, the CMB maps detected by the WMAP satellite have shown to reveal possible deviations from this assumption. There have been claims of a hemispherical power asymmetry between two parts of the sky, where one part exhibits stronger fluctuations and the other weaker. This feature has been investigated since the first-year maps were released, and a number of independent studies have found consistent results, in particular regarding the preferred direction of the asymmetry. However, the methods have traditionally been limited computationally to only including the largest scales in the analysis. In a recent paper, Hanson & Lewis (2009) developed a Quadratic Maximum Likelihood estimator for analyzing the dipolar asymmetry, and found consistent results with previous studies but noted that the effect diminished with including higher l in the analysis.

We present the Quadratic Maximum Likelihood estimator introduced by Hanson & Lewis, and re-implement it for our own analyses. We investigate the claims made by the WMAP team, in Bennett et al. (2011), that the reported dipolar asymmetry effect is a statistical fluke and that the previous claims found in the literature are insignificant. The WMAP team's claims are based on the QML estimator.

The results we obtain are found to be consistent with previous studies, H&L and with Bennett et al., and the significance of the asymmetry is found to vary with the assumed modulation multipole l_{mod}. However, we find the claims by the WMAP team to be premature as they have not taken into account the consistency of the estimated direction. Comparing with 1000 isotropic simulations we find that less than 1% of them exhibit a similar consistency with estimated direction across l_{mod}.

Acknowledgements

First and foremost I need to thank my supervisor, Hans Kristian Eriksen. Through his positivity and helpfulness he always manages to reassure me and make me feel I am on the right path. Such qualities are essential when, over the course of a year, morale and inspiration can go down from time to time. I would also like to thank all my fellow students at Stjernekjelleren, and with regards to the project, especially Dag Sverre Seljebotn who has been of an invaluable help.

Part of this work was performed on the Titan Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Research Computing Services group at USIT, the University of Oslo IT-department.
The computations rely on the use of HEALPix.

Svein Rune Saxrud
June 1, 2011

Contents

Abstract iii

Acknowledgements v

Contents vii

1 Introduction 1
1.1 Overview of Cosmology . . . 1
1.1.1 The standard model of cosmology (ΛCDM) . . . 2
1.1.2 Perturbations in the smooth universe . . . 6
1.2 The Cosmic Microwave Background and its properties . . . 7
1.2.1 The physics of the CMB . . . 8
1.2.2 The power spectrum . . . 11
1.2.3 Cosmological parameters and their current best-fit values . . . 12
1.3 The dipole model . . . 14
1.3.1 Possible physical explanations for the asymmetry . . . 17

2 Mathematical description of the CMB 19
2.1 Decomposing into spherical harmonics . . . 19
2.2 Observations into pixelized map . . . 20
2.2.1 Galaxy and foreground removal . . . 22
2.2.2 Dealing with the instrumental beam . . . 24
2.2.3 Representing the decomposed map . . . 24
2.2.4 Downgrading maps . . . 25
2.3 The signal covariance matrix . . . 26
2.3.1 Including the dipole model . . . 26

3 Statistical estimators for the dipole model 29
3.1 Quadratic estimator (Hanson & Lewis) . . . 29
3.1.1 Derivation of the estimator . . . 29
3.1.2 The estimator applied to the dipole anisotropy case . . . 32
3.1.3 The Fisher matrix for the dipole case . . . 36
3.1.4 Calculating the dipole amplitude and direction from the harmonic modulation . . . 37

3.2 Optimal maximum-likelihood (Eriksen, Hoftuft) 38

4 Implementation **41**
4.1 The conjugate gradients algorithm 41
4.1.1 The algorithm 43
4.1.2 Preconditioner 44
4.2 Computational expense 45
4.2.1 The estimator 45
4.2.2 The CG search 45

5 Analysis of simulations **49**
5.1 Creating simulated maps 49
5.2 Verification of unbiasedness 50
5.2.1 Simulations with a dipole modulation 51
5.2.2 Simulations without a modulation 51
5.2.3 Simulations with a sky-cut 56
5.3 Estimating confidence and statistical significance 60
5.4 Determining the full Fisher matrix 60

6 Analysis of WMAP7 data **61**
6.1 Template cleaning 61
6.2 Analyzing the full data set 62
6.2.1 Creating the noise rms maps 63
6.2.2 Analysis results 63
6.2.3 Using more than one iteration in the estimator 74

7 Conclusions / Final words **77**
7.1 Is there evidence of a dipolar asymmetry in the WMAP data? 77
7.2 Future projects 78

A Resource **81**
A.1 The HEALPix software 82
A.2 The 3-j symbols 84
A.3 Spherical harmonics 86
A.3.1 Real-valued version 86

Bibliography **89**

Chapter 1

Introduction

For thousands of years people have gazed into the heavens, trying to figure out what they see. In that sense, astronomy is one of the oldest sciences around. But it is also one of the youngest: Only recently have we had the technology and the physical insights available to allow us to make more than mere guesses, especially when it comes to the field of cosmology. It is probably safe to say that few scientific disciplines have made greater progress in the past 15-20 years than cosmology.

Much of the material in this chapter is based on Dodelson (2003) [1] and Liddle (2003) [2]. The former is highly recommended for a thorough mathematical description of cosmology and structure formation in the universe, while the latter serves as a more accessible introduction to the field.

1.1 Overview of Cosmology

Cosmology is the study of the entire universe. During the 20th century, the discipline went through quite a transition: At the onset, cosmology was much speculation, but with the advent of general relativity, quantum physics and more and more sophisticated observations, it has moved to becoming largely precision science. Today astronomers can proclaim to knowing the age of the universe to within 1% accuracy, as well as its composition and evolutionary history.

One of the first important observations was that of Edwin Hubble in 1929, who showed that galaxies were more and more redshifted the further away their distance from us. This was interpreted as that the universe was expanding, and laid the groundwork for an acceptance of the Big Bang theory over the static Steady-State universe, though that still had many followers. Of note is that before Hubble's discovery, Albert Einstein, the creator of the general theory of relativity, was a strong believer in the Steady-State-model, so much so that he modified his equations when they seemed to suggest an evolving universe. This modification was the addition of a constant Λ, named the *cosmological constant*, and by tweaking its value the models could support a stationary universe. After Hubble's discovery, however, he famously deemed the introduction of this constant the "greatest blunder" of his life. However, with the discovery

of *dark energy* and an accelerating universe in 1998, the cosmological constant again became relevant as a means of describing this new phenomenon. The constant is perfectly valid in the theory (it's really just a constant of integration), and there is no real reason it should just equal 0.

The discovery of the cosmic microwave background in 1965 strongly favoured the Big Bang theory. Not only was it the only theory that could readily explain this radiation; it had also predicted it. After this observational milestone, nearly all cosmologists accepted the Big Bang model as the basis for explaining the evolution of the universe.

The initial discovery of the microwave background was in many respects a fluke: Two radio astronomers, Penzias and Wilson [3], found that their telescope produced some noise they hadn't accounted for. After carefully removing dust and accounting for all the noise they could think of, this background noise still persisted, and they realized it had to be the proposed cosmic microwave background radiation. In the years since, several instruments have been built explicitly for measuring the CMB radiation, and particularly the anisotropies in the smooth background. The most prominent ones have been the COBE[1] satellite mission, which mapped the whole sky, the BOOMERanG[2] balloon experiment, which had higher resolution but only scanned part of the sky, and the WMAP[3] satellite, that mapped the whole sky to sub-degree resolution. The WMAP satellite has given us the greatest observations yet, but the new spacecraft Planck[4], launched in 2009, offers even higher resolution and will soon release its first data for scientific study. Results from the various observations have provided us with a large degree of information about the universe, and they have allowed us to form a *standard model of cosmology*. COBE showed the CMB to have a virtually perfect blackbody spectrum, and also gave the first detections of the anisotropies in the background radiation. BOOMERanG determined the geometry of the universe, and WMAP has improved on these earlier results in addition to providing numerous other constraints on the model. The Planck spacecraft will in particular give us much better data of the polarization of the CMB, which will further enhance our knowledge of the cosmos.

1.1.1 The standard model of cosmology (ΛCDM)

In the generally accepted, standard Big Bang model, there are two basic, underlying assumptions which need to be fulfilled (on large scales):

- The universe must be homogeneous
- The universe must be isotropic

These assumptions stem from the notion of the *Cosmological principle*, which states that the properties of the universe look the same for all observers, i.e. no observer occupies

[1] **CO**smic **B**ackground **E**xplorer, `http://lambda.gsfc.nasa.gov/product/cobe/`

[2] **B**alloon **O**bservations **O**f **M**illimetric **E**xtragalactic **R**adiation **a**nd **G**eophysics, `http://www.astro.caltech.edu/~lgg/boomerang/boomerang_front.htm`.

[3] **W**ilkinson **M**icrowave **A**nisotropy **P**robe, `http://wmap.gsfc.nasa.gov`.

[4] Named after German scientist Max Planck, `http://www.rssd.esa.int/index.php?project=Planck`.

a special place. This principle has been a cornerstone of cosmology for many years, and means that even though two observers placed at different points in the cosmos observe different samples of the universe, the statistical properties inferred should be the same.

The framework in which we describe the universe is the general theory of relativity, published by Einstein in 1916. With it came a fundamental shift in how we viewed gravity: No longer was it to be considered a force, but rather a consequence of curved spacetime. For an introduction to general relativity with emphasis on the underlying mathematics see, e.g., Grøn (2009) [4], which was an occasional reference for this chapter. The Einstein field equation, which relates the energy content to the geometry of the universe, is given in compact form by the tensor equation

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{1.1}$$

where we have set the constants $c = G = 1$. The $T_{\mu\nu}$ is the energy-momentum tensor, and $G_{\mu\nu}$ is the Einstein tensor, which represents the geometry. A slightly more expanded version, also including the cosmological constant, is

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{1.2}$$

where $R_{\mu\nu}$ is the Ricci tensor, R the Ricci scalar and $g_{\mu\nu}$ the *metric tensor*, which details the geometry of spacetime. By requiring that the universe is homogeneous and isotropic the metric tensor becomes the Friedmann-Robertson-Walker (FRW) metric,

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}, \tag{1.3}$$

where the first item on the diagonal is the time component and the rest are the spatial ones, with $a(t)$ being the scale factor, which details the expansion of the universe. One commonly sets $a(t_{\text{today}}) = 1$. By solving the Einstein equation for this metric, along with requiring the energy content to consist of so-called *perfect fluids* (having no viscosity or heat flow, with $T_{\mu\nu} = \text{diag}(-\rho, p, p, p)$), one obtains the Friedmann equations, from which we can calculate the evolution of the universe for a number of different cases. These equations can be written as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.4}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.5}$$

where ρ is the energy density[5] and p the pressure of the fluids, and k is the curvature constant. The dots denote derivation with respect to cosmic time. In the fluid description of the cosmological constant, its density is set to $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$, which is also constant.

[5] I use the term energy density loosely here; strictly speaking ρ is the matter density and ρc^2 the energy density. Often c is set to 1 and then they are equal.

An expression for the evolution of the densities can be calculated, and is given by

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0, \tag{1.6}$$

which describes the expansion of the universe to be *adiabatic*. The density and pressure for the cosmic fluids are commonly related through the simple equation of state $p = w\rho c^2$, where w is a unique constant for each type of fluid. For matter (both ordinary and dark) it is 0 and for radiation $\frac{1}{3}$, while for the dark energy fluid it is -1. From equation 1.6 it is thus quite easy to calculate the a-dependence of these different types of fluid. We find the relations

$$\begin{aligned} \rho_r &= \rho_{r0}\, a^{-4} \\ \rho_m &= \rho_{m0}\, a^{-3} \\ \rho_\Lambda &= \rho_{\Lambda 0}, \end{aligned} \tag{1.7}$$

where ρ_r is the density of radiation and ρ_m for matter, while the ρ_0's indicate the respective values today.

Armed with all these equations, we can calculate the evolution of the scale factor and the evolution of the energy content in a variety of possible universes. The currently accepted model is known as ΛCDM, and its evolution is outlined below. As a reference it's useful to consider how the scale factor evolves with time in universes consisting of only *one* type of fluid, and with flat curvature ($k = 0$). This is found from equation 1.4, and the relations are

$$\begin{aligned} a_r(t) &\propto t^{1/2} \\ a_m(t) &\propto t^{2/3} \\ a_\Lambda(t) &\propto e^{\alpha t}, \end{aligned} \tag{1.8}$$

where the subscripts denote which fluid the universe consists of, and the α is a constant (equal to $\sqrt{\frac{\Lambda}{3}}c$).

The evolution of the universe

Observations today indicate that the universe is expanding, and extrapolating this backwards in time, we eventually come to a point where the universe is extremely hot and dense. The Big Bang models start at a time 0, but they don't really come into play until slightly after, at $t = 10^{-43}$s, corresponding to the Planck time. Before this time, known physics break down, so the model only concerns the evolution of the universe from this point, not how it came into being (to be able to explain what happened between $t = 0$ and the Planck time, we need to merge general relativity with quantum physics, which for instance the superstring theory tries to accomplish). The Big Bang is also not to be viewed as an explosion of matter, but rather as the universe expanding equally much at all locations (note also that there is no discrepancy between an infinitely large universe which is also expanding in this sense, since the expansion means that e.g. the space between two galaxy clusters is getting larger and larger). In any case, the universe was early on a very hot and dense plasma of free particles, with radiation being the dominant constituent. As the universe expanded, it also cooled, and

eventually became matter dominated. It continued to be matter dominated for most of its lifetime so far, but recently a type of dark energy has taken over. Dark energy was not "part of the equation", so to speak, before 1998, when it was discovered that the expansion was accelerating. Had the universe only consisted of radiation and matter, this should not be happening, as the gravitational attraction should cause the universe to decelerate its expansion. Instead, observations show it to have begun accelerating quite recently, indicating some other source of energy that exerts a repulsive gravitation. If this energy is of the form of a cosmological constant, the future of the universe is looking quite grim: The expansion will become exponential, and we may end up in a so-called "Big Rip", leaving the universe as a dark and empty place with only the scarce particle here and there.

Inflation

The picture outlined so far does not mention inflation. This is believed to be a physical process that serves to solve several problems regarding what we see from observations, among others the flatness problem and the horizon problem. In addition it also created the small perturbations from the smooth CMB background. Inflation is thought to have happened exceptionally early on in the lifetime of the universe, at about 10^{-35}s after the Big Bang. The process made the universe expand extremely rapidly, about a factor 10^{40}, over the course of just 10^{-34}s. The mechanism behind inflation may have been of the same sort that today drives the accelerated expansion.

The *flatness problem* is the notion that, without inflation, for the universe to appear flat today it must have been extremely fine-tuned towards being flat from the beginning (deviation from flatness increases with time in matter- or radiation-dominated universes, unless the universe is precisely flat, in which case it will remain flat forever). Physicists don't like their models to necessitate fine-tuned initial conditions; we rather want a pure physical basis for the observations we make. Inflation solves this problem since it would make the universe appear flat regardless of the initial geometry (much like a sphere would appear flat on smaller scales if it were blown up).

The *horizon problem* comes from the fact that observations show the CMB, and thus the universe at the time of its formation, to have close to uniform temperature across the whole sky. This would indicate that the matter on one part of the universe must have been in thermal contact and in thermal equilibrium with the other. However, when calculating the horizon (the size of the observable universe, i.e. the size on which information can have travelled) at the time of the formation of the CMB, we see that it was much smaller than the length from one side to the other. There would then be no reason why these two parts of the universe should have the same temperature. Inflation fixes this problem since small scales that were in thermal contact before, get blown up to much larger scales afterwards. Thus, two parts of the universe *outside* each others' horizon after inflation will still have the same temperature, since they used to be *inside* each others' horizon before inflation.

As mentioned there are also some other problems inflation helps to solve (such as the magnetic monopole problem), as well as it being responsible for preparing the universe for structure formation, through creating perturbations in the smooth background by

blowing up initial quantum fluctuations in the extremely dense early universe. These initial fluctuations would happen because of the statistical nature of quantum mechanics, where particles follow the Heisenberg uncertainty relation. The inflationary theory was first postulated by Guth (1981) [5].

The energy content of the universe
In the ΛCDM model, there are four main constituents to the energy density of the universe: Radiation (photons and neutrinos), baryons (a collective term for ordinary matter), dark matter (the CDM stands for Cold Dark Matter, which is thought to be the most abundant kind) and dark energy (represented by the Λ). The latter two make up about 95% of the energy density in this model, but unfortunately, as their names suggest, we don't have much of a clue what they actually *are*. Ongoing experiments at CERN[6] are hoped to give at least some answers with regards to dark matter. Nevertheless, this model matches well with several different types of observations, both CMB observations (including WMAP) and others. In section 1.2.3 the current best-fit values of the energy content are shown.

1.1.2 Perturbations in the smooth universe

The evolution of the universe described by the simple equations in the above section only concerns the large-scale behaviour of the smooth background universe. This is not a very good description of the universe as a whole, however, as we all can see each day that the universe is far from smooth. We therefore need to introduce perturbations in the smooth background, and see how these perturbations evolve with time to grow into larger structures. Doing this involves dealing with highly complicated equations, and is outside the scope of this text. The framework for calculating the evolution of the perturbations is the Einstein-Boltzmann equations, where one looks at the first-order perturbations (i.e. the perturbations are assumed *small*). We mentioned earlier that the origin of the anisotropies are quantum mechanical and come from the inflationary epoch. When calculating the evolution of the perturbation parameters we don't concern ourselves with that, but rather solve for initial perturbations set up after inflation.

In the simplest case where we only consider scalar perturbations[7] to the FRW metric the perturbed metric can be written as

$$g_{\mu\nu} = \begin{cases} -1 - 2\Psi(\mathbf{x}, t) & \text{if } \mu = \nu = 0 \\ \delta_{ij} a^2(t)(1 + 2\Phi(\mathbf{x}, t)) & \text{otherwise,} \end{cases} \tag{1.9}$$

where Ψ is the perturbation to the Newtonian gravitational potential and Φ the perturbation to the spatial curvature, and the vector $\mathbf{x}$ denotes the spatial coordinates. The E-B equations have to be solved for these perturbations, as well as perturbations in the densities of the different constituents and their velocities relative to the background. The Boltzmann equations relate the distribution of a particle to its interactions, or collisions, with other particles. From these we find the evolution of the perturbation

[6] CERN is the European Organization for Nuclear Research, `www.cern.ch`.
[7] There could also be vector and tensor perturbations. See [1] for details.

parameters of the particles, and to get the evolution of the gravitational perturbations Ψ and Φ we need to combine with the Einstein equations. Putting it together we get a dense set of coupled equations, which are transformed into the Fourier regime for easier calculations. For cold dark matter, for instance, the equations governing its perturbation parameters are

$$\begin{aligned} \dot{\delta} + ikv &= -3\dot{\Phi} \\ \dot{v} + \frac{\dot{a}}{a}v &= -ik\Psi, \end{aligned} \tag{1.10}$$

where δ is the density perturbation and v the velocity, and k is the Fourier mode, or wave number. The dots indicate derivation with respect to *conformal time* η, which is equal to the horizon apart from a division by c. The parameters as written here are then functions of k and η. The equations are solved independently for each k, which allows us to follow the evolution of structure formation on different size scales.

These early perturbations give rise to the anisotropies in the cosmic microwave background radiation, and studying them is crucial to our understanding of the universe. With the standard inflationary ΛCDM model these anisotropies follow the cosmological principle, and should therefore be statistically isotropic and homogeneous.

1.2 The Cosmic Microwave Background and its properties

The cosmic microwave background (the CMB) is a relic from when the universe was only about 380,000 years old. Before this time, the universe was *opaque*; that is to say, light could not escape. This was because of the extremely hot temperatures in the early universe, which made sure free particles did not form into atoms. Thomson scattering of photons off electrons, then, was the reason light could only go so far before being scattered off again. However, as the universe expanded, the temperatures dropped, eventually so that atoms *could* indeed form without being instantly ionized. The period where this happens is known as *recombination*, and is the epoch from where the CMB originates. After recombination, the universe was almost completely transparent, so the photons have been able to travel virtually without interference all the way to being picked up by our observational detectors today. In other words, the CMB radiation is the absolutely earliest photonic radiation[8] we can ever hope to observe, and looking at it we are effectively looking backwards in time, billions of years, to an age when the universe was very young.

The CMB is shown to have a virtually perfect black-body spectrum (figure 1.2), with a mean temperature of about 2.7K. Again, this agrees with the predictions of the Big Bang model, as the photon-baryon plasma was in perfect thermal equilibrium. The low temperature is due to the photons having lost energy as the universe has expanded (the wavelength is then "stretched", and the energy, and thus the temperature, is related to wavelength: Longer wavelength means lower energy).

[8]It is possible to observe other types of radiation from even further back, from neutrinos and gravitational waves.

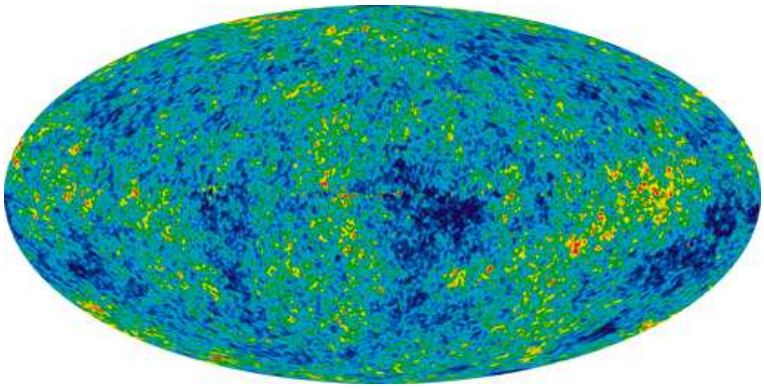


Figure 1.1: Full-sky map of the CMB produced using the *Internal Linear Combination* technique on 7-year WMAP data. Image courtesy of the WMAP Science Team.

Observations (from WMAP and other experiments) show the CMB to be remarkably uniform, with deviations from the mean temperature only on the order of 10^{-5}. These small deviations, however, are the reasons we exist today: They laid the groundwork for gravitational collapse of matter, which in turn created galaxies, stars, planets and, in the end, us. A full-sky map of the CMB fluctuations as measured by WMAP is shown in figure 1.1. The different colors correspond to different temperatures, where red is hotter and blue is colder. The difference between the hottest and coldest spots on the map are a couple hundred μK.

1.2.1 The physics of the CMB

In this section we will review shortly the different physical processes that made the CMB "what it is". They are a combination of processes occurring before the radiation was sent out towards us, and processes affecting the photons afterwards, on the way to being detected. In addition to Dodelson and Liddle, Tegmark (1996) [6] was a useful reference.

Primordial fluctuations set up by inflation
At the time before inflation is thought to have happened, the universe was so dense that it was governed by quantum mechanics. Random quantum mechanical fluctuations then appeared in the otherwise smooth primordial "soup". With the extremely rapid expansion of the inflationary era, these tiny fluctuations were blown up, thereby creating a basis for gravitational attractions to further enhance the fluctuations. These fluctuations are the main reason we see the anisotropies in the CMB, and inflation is also the reason we see the CMB to be as uniform as it is.

Acoustic oscillations
The primordial fluctuations/perturbations set up during inflation affected the photons as well as matter. Photons and matter were also tightly coupled through Thomson

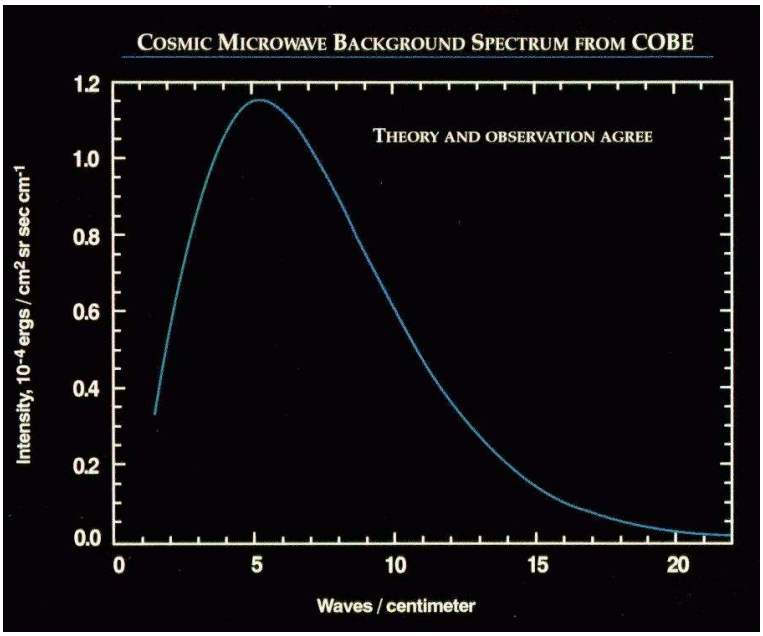


Figure 1.2: The blackbody spectrum of the CMB as measured by COBE. The theoretical curve obscures the error bars in the data points. Image courtesy of NASA.

scattering in the photon-baryon fluid. Due to the relativistic pressure of the photons, in opposition to the gravitational attraction of matter (baryons exert some pressure as well as setting up gravity, dark matter only interacts through gravity), oscillations happen. Since the photons have such a strong pressure, after they're decoupled from the matter they will travel unperturbed so the fluctuations at the time of the CMB are preserved when we observe them.

Different modes of oscillation entered the horizon at different times. For modes with wavelengths larger than the horizon at recombination, they have not been able to undergo any oscillations (since causal physics couldn't have affected the perturbations), so the perturbations observed at these modes will be close to those set up initially after inflation. Modes on somewhat smaller scales have entered the horizon some time before recombination, so the perturbations have begun to grow. This mode will then have stronger perturbations than the previous, super-horizon one. Even smaller modes have entered the horizon even earlier, so their perturbations could have experienced full oscillations, making the perturbations less powerful. This picture will continue, so the anisotropy spectrum will have peaks (where the perturbations have undergone oscillations so that they now are at a maximum) and troughs (where they have undergone oscillations so that they now are at minimum amplitude).

Recombination and decoupling

As touched upon already, after the universe had expanded and cooled sufficiently, the energy of the photons were low enough to no longer be able to prevent atoms from form-

ing. Electrons and photons first forming into atoms is called *recombination*. Even after recombination, the photons still interacted somewhat with the matter, but eventually the photons were completely decoupled from matter, and this is called *decoupling*. The two eras closely follow each other. The CMB radiation we observe is that from decoupling, or the *surface of last scattering*. Obviously this process was not instantaneous, but it was quite close to, so all the CMB photons we see originates from more or less the same surface (at a redshift of about $z = 1100$). However, there is a certain thickness to the last-scattering surface, and this serves to wash out the acoustic oscillations on scales corresponding to this size.

Curvature
Curvature would affect the CMB signal since light follows different paths in a curved than a flat universe. We call the universe *flat* if it follows euclidean curvature, *closed* if two parallel light trajectories converge (like on a spherical surface), and *open* if they diverge. In an open universe the structure we observe on the CMB would be projected onto a *smaller* angle than in a flat universe, and vice versa for the closed case. Thus, different geometries would shift the position of the peaks in the anisotropy spectrum (to the right, towards smaller scales in the open case, and to the left in the closed case). Observations tell us that the position of the first peak is nearly exactly where it would be in the flat case. However, it is highly unlikely for the universe to actually have exactly flat geometry, since that is just *one* particular possible case out of an infinite number of possibilities. But at least we can say the it is very, very close to being flat. Note also that flat and open universes would be infinitely large, while a closed universe would be finite in size.

The Integrated Sachs-Wolfe effect
This effect is not part of the true CMB signal, but the signal we receive has been affected by it. It comes into play recently, when the universe has become dominated by dark energy. On its journey from the origin of the CMB to us today, the photon has had to travel through gravitational potentials set up by matter. Going into this well the photon is blue-shifted, and going out again it gets redshifted. In the matter-dominated era, however, these effects cancel each other out since the gravitational potential remains constant. When the universe has become dominated by dark energy this causes a shift in the gravitational potentials the photons have to travel through to get to us, and the net effect makes the photons redshifted. There was also a Sachs-Wolfe effect early on, shortly after the surface of last scattering, when there were still enough radiation left to affect the photons gravitationally, so the potential wells were not constant. The late effect makes a good probe for measuring dark energy, since it affects the anisotropy spectrum on the largest scales in a distinct way.

Reionization
Between the origin of the CMB and today, it's possible that the photons pass through regions that have once again become ionized. Particularly a global reionization could strongly affect the spectrum. Then a fraction of the CMB photons would again be scattered off electrons, causing them to have a new last-scattering surface. This effect would also serve to wash out the smallest scales: The newly scattered photon coming

from a certain direction in space, could in actuality have come from a number of directions at the true CMB surface. Thus, the measured temperature in this direction becomes a weighted average of the temperatures at a certain part of the original CMB surface. An additional effect of strong reionization would be the emergence of new acoustic oscillations, serving to strengthening the anisotropies on the largest scales.

Contaminations from other sources of microwave radiation

The true cosmic microwave background radiation is not the only microwave radiation we receive from the universe. Our own galaxy, the Milky Way, emits a good amount of this radiation (through, for instance, dust, synchrotron and free-free emissions), and so do other galaxies and point-sources. This unwanted radiation is called *foregrounds*, since the sources of it are in front of, or closer to us, than the CMB radiation.

In addition, we also get contaminations from processes that are not due to any radiation: The movement of the satellite across the sky introduces a doppler effect, and each pixel has intrinsic random noise.

1.2.2 The power spectrum

The power spectrum, C_l, is defined as the expectation value of the square of the spherical harmonics coefficients from the CMB map (see section 2.1), or alternatively, the variance of these coefficients. It gives the power of the fluctuations on different angular scales l, where increasing l corresponds to smaller angular scales, and is given by

$$C_l \equiv \langle |a_{lm}|^2 \rangle = \langle a_{lm} a_{lm}^* \rangle = \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2, \tag{1.11}$$

where the a_{lm}'s are the spherical harmonic coefficients. This is precisely the anisotropy spectrum mentioned in the previous section, complete with the peaks and troughs of the acoustic oscillations.

Using the power spectrum we can constrain a great many parameters in the ΛCDM model, as well as ruling out several possibilities. As mentioned the position of the first peak determines the curvature of space (and we saw earlier that it points to the universe having euclidean geometry), and the height difference of the peaks points to how large the baryon density is. Figure 1.3 shows the striking agreement between the power spectrum as measured by WMAP and that predicted by the current ΛCDM model. Only six basic parameters are required in the model to fit the data.

Computing the power spectrum for a given cosmological model

The computation of a theoretical power spectrum involves solving the perturbation equations and taking into account the different physical processes described in the above section. Specifically we need to obtain the perturbations to the photon temperature, $\Theta_l(k)$, where we can follow how the perturbations change over different multipoles and different Fourier mode size scales. How such a computation can be performed with relative ease is excellently described in Callin (2006) [7]. Specifically, for a so-called

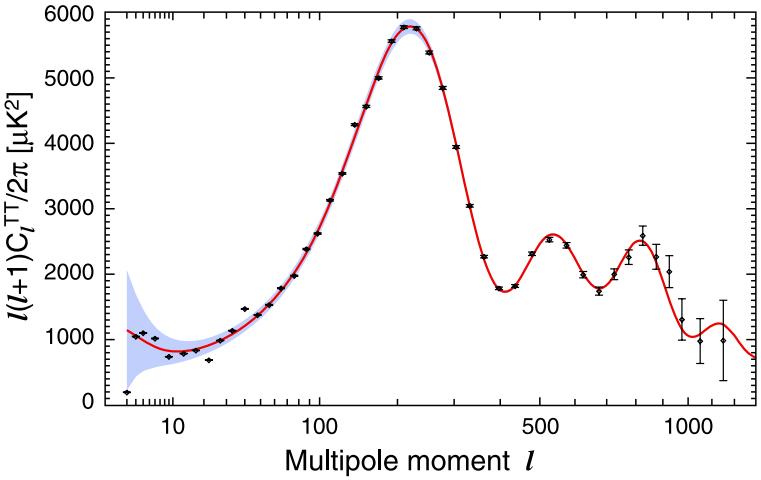


Figure 1.3: The CMB power spectrum as predicted by the ΛCDM model and observed by WMAP. The shaded band indicates *cosmic variance* (see section 2.1). Image courtesy of the WMAP Science Team.

Harrison-Zel'dovich spectrum (which is predicted by many inflationary models), we would get

$$C_l = \int_0^\infty \left(\frac{ck}{H_0}\right)^{n_s-1} \Theta_l^2(k)\frac{\mathrm{d}k}{k}, \tag{1.12}$$

where H_0 is the Hubble parameter and n_s the spectral index, both described in the next section.

1.2.3 Cosmological parameters and their current best-fit values

The newest WMAP analyses, from Larson et al. (2011) [8], give these values for some of the main parameters of the ΛCDM universe we are believed to inhabit (the $_0$ subscript indicate that these are the values *today*):

$$\begin{aligned}
\Omega_{b0} &= 0.0449 \pm 0.0028 \\
\Omega_{c0} &= 0.222 \pm 0.026 \\
\Omega_{\Lambda 0} &= 0.734 \pm 0.029 \\
t_0 &= 13.75 \pm 0.13 \text{ Gyr} \\
H_0 &= 71.0 \pm 2.5 \text{ km s}^{-1} \text{ Mpc}^{-1} \\
n_s &= 0.963 \pm 0.014 \\
\tau &= 0.088 \pm 0.015 \\
\Delta^2_{\mathcal{R}} &= (2.43 \pm 0.11) \cdot 10^{-9}
\end{aligned}$$

Here Ω_b is the baryon density, Ω_c the density of cold dark matter, Ω_Λ the dark energy density, t_0 the age of the universe, H_0 the Hubble parameter (which is a measure of the

expansion of the universe, $H \equiv \frac{\dot{a}}{a}$), n_s the spectral index, τ the reionization parameter and $\Delta_{\mathcal{R}}$ the curvature perturbation amplitude. The Ω's represent the energy density of the parameters given in terms of the *critical density*,

$$\Omega_b \equiv \frac{\rho_b}{\rho_{\text{crit}}}, \tag{1.13}$$

where ρ_{crit} is the density the universe *would have had* if it were completely flat. This density is given by

$$\rho_{\text{crit}} = \frac{3H^2}{8\pi G}. \tag{1.14}$$

Inserting the proper values, we find that the current value of the critical density is

$$\rho_{\text{crit0}} \approx 9.596 \cdot 10^{-27} \ \frac{\text{kg}}{\text{m}^3}.$$

This is an extraordinarily low value, equivalent to about 5 hydrogen atoms per cubic metre. Since the universe is measured to be very close to flat, we can interpret the values of the density parameters as being the percentage of that type of energy content in the universe.

It is also possible to construct a *curvature density parameter*, given by

$$\Omega_k = -\frac{kc^2}{H^2a^2}. \tag{1.15}$$

Using the density parameters we can write the first Friedmann equation (eq. 1.4) as

$$H^2 = H_0^2 \left(\Omega_{r0}a^{-4} + \Omega_{m0}a^{-3} + \Omega_{k0}a^{-2} + \Omega_{\Lambda 0}\right), \tag{1.16}$$

where $\Omega_{m0} = \Omega_{c0} + \Omega_{b0}$, and $\Omega_{k0} = -\frac{kc^2}{H_0^2}$. At all times there is a requirement that the sum of all density parameters, Ω, equals 1. This also explains the flatness problem: If $k = 0$, the sum of the ordinary density parameters should be 1, which would make $\Omega - 1 = 0$; however, if the universe only *appears* flat we would have (still considering only the ordinary densities in Ω)

$$|\Omega - 1| \propto \frac{1}{H^2a^2} \propto \begin{cases} t & \text{radiation domination} \\ t^{2/3} & \text{matter domination,} \end{cases} \tag{1.17}$$

making clear the increasing deviation from flatness with time. Under dark energy, or cosmological constant, domination the deviation $|\Omega - 1|$ instead goes as $e^{-2\alpha t}$, which gets smaller with time.

The spectral index of density perturbations, n_s, is a measure of the so-called *tilt* of the power spectrum, which serves to decrease the spectrum on the small scales and increase it on the largest scales, when n_s is less than 1. A spectral index equal to one details a scale-invariant spectrum of primordial perturbations, while inflationary models generally predict slight deviations from scale-invariance. The curvature perturbation amplitude $\Delta_{\mathcal{R}}$ quantifies the magnitude of the primordial perturbations.

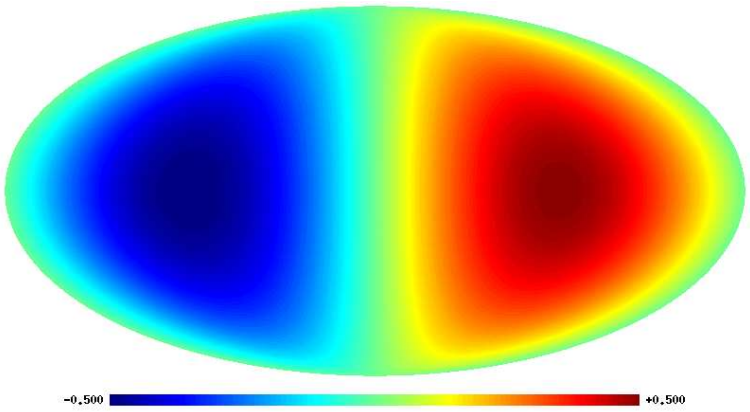


Figure 1.4: An example of a dipole with amplitude 0.5, in the direction $(\theta, \phi) = (90°, 270°)$.

The reionization parameter τ measures how many photons where scattered again after the surface of last scattering, due to passing through a reionized region. The fraction of photons escaping through the ionized region is given by $e^{-\tau}$, and also a fraction $1 - e^{-\tau}$ is scattered towards us from another direction. The τ itself is the *optical depth* of the reionized region.

1.3 The dipole model

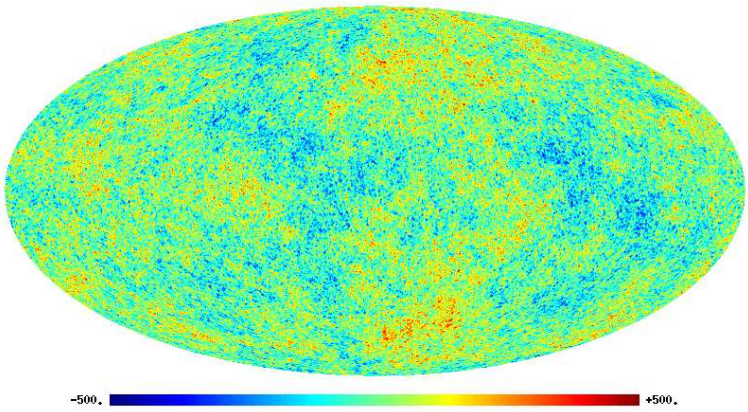
As we have seen in section 1.1.1, standard Big Bang cosmology requires the universe to be homogeneous and isotropic. However, when subjected to statistical analyses, the WMAP data have shown to reveal a power anisotropy between two parts of the sky. This anisotropy can be approximated as a simple dipole modulation of an inherent, statistically isotropic Gaussian signal:

$$\mathbf{d}(\hat{\mathbf{n}}) = (1 + A(\hat{\mathbf{n}} \cdot \hat{\mathbf{p}}))\mathbf{s}(\hat{\mathbf{n}}) + \mathbf{n}(\hat{\mathbf{n}}). \tag{1.18}$$

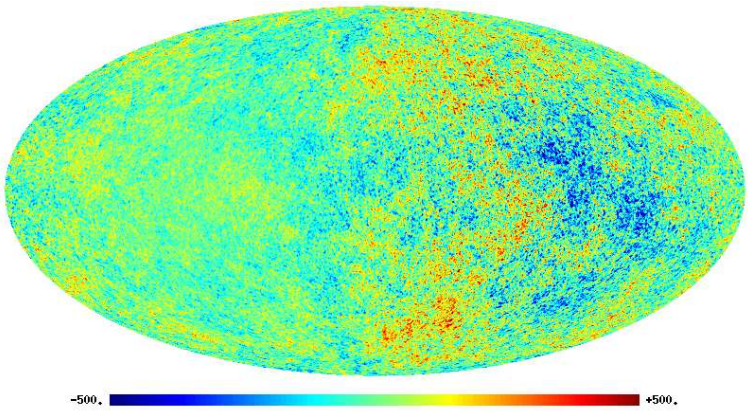
Here, $\mathbf{s}$ is the inherent, isotropic signal, A is the dipole amplitude, $\mathbf{n}$ is instrumental noise and $\mathbf{d}$ is the observed data. The direction on the sky is given by the unit vector $\hat{\mathbf{n}}$, and $\hat{\mathbf{p}}$ is the unit vector of the dipole asymmetry direction. This model doesn't really have a physical motivation, it's just a means of describing the asymmetry.

In figure 1.4 there is shown an example of a dipole field. The effect this modulation has on a CMB map is shown in figure 1.5. We see that it suppresses the fluctuations on one half of the map, while strengthening them on the other. Note that this particular modulation is strongly exaggerated, in reality the modulation we see has an amplitude of less than 0.1.

The first to make claims of a dipolar asymmetry were Eriksen et al. (2004) [9], who, by computing the power spectrum locally for different patches of the sky, showed signs



(a) Simulated, statistically isotropic CMB map.



(b) The same map as above, modulated with a dipole of 0.5 amplitude.

Figure 1.5: Isotropic and modulated maps.

of one hemisphere having significantly higher and the other lower power than what was expected from isotropic simulations, looking at the first-year WMAP data. They also found a similar asymmetry in the COBE-DMR data. The analysis reported the asymmetry up to an $l = 40$. Hansen et al. (2009) [10] did a similar analysis for much larger multipole ranges using the 5-year WMAP data, and found that the asymmetry extended to $l = 600$. The reported asymmetry direction was $(\theta = 107^\circ \pm 11^\circ, \phi = 216^\circ \pm 10^\circ)$.

Eriksen et al. (2007) [11] used a Bayesian maximum-likelihood analysis of the dipole modulation model of equation 1.18 on downgraded, low-resolution maps of the 3-year WMAP data, and found amplitudes of about $A \sim 0.12$, at a direction consistent with those reported by the other methods. This analysis was later repeated and expanded by Hoftuft et al. (2009) [12] with consistent results, arguing that the significance of the detection increased with including more scales in the analysis. This analysis was done up to a maximum modulation multipole $l_{\mathrm{mod}} = 80$. An amplitude $A \sim 0.07$ was stated to be consistent with all the different l_{mod} and datasets in their analyses. Specifically, for the ILC 5-year map and $l_{\mathrm{mod}} = 64$ they give an amplitude of $A = 0.072 \pm 0.022$ and direction $(\theta, \phi) = (112^\circ, 224^\circ) \pm 24^\circ$.

In addition to the above papers, several other authors have reported a similar power anisotropy using a variety of methods, see for instance [13], [14], [15] and [16].

The WMAP team, in Bennett et al. (2011) [17], argue that there is no significant evidence of a dipolar asymmetry in the data, and that the purported claims turn out to be just statistical flukes. They base this on results obtained with the QML estimator of Hanson & Lewis (2009) [18]. In fact, they go as far as to compare these claims with the notion that, if you search for something odd, you will inevitably find it. This estimator is approximate, while that used by Eriksen et al. (2007) and Hoftuft et al. (2009) is an optimal maximum-likelihood estimator. However, the QML approach is much more computationally effective so the analysis can be performed on much smaller scales, and there is no need to downgrade the maps. Our goal in this thesis is to re-implement the quadratic estimator and repeat the analyses done by Hanson & Lewis and Bennett et al., and see if the latter's claim of there being no significant evidence of a dipolar modulation is valid.

If there indeed turns out to be a dipole anomaly present, and it is not the cause of some hitherto unknown systematics, it will have profound consequences for the current cosmological models. They would need to be modified to be able to convincingly account for such a feature.

In addition to the dipolar asymmetry described above, several other anomalies have been reported in the WMAP data. These include the existence of a large cold spot (the strong blue region seen in the center of figure 1.1), the alignment of the quadrupole/octopole and the lack in quadrupolar power. A quick overview of these possibly anomalous effects can be found in [17].

1.3.1 Possible physical explanations for the asymmetry

The model for estimating the dipole modulation is, as said, strictly phenomenological, but with the increased attention to the asymmetry in recent years people have tried to come up with possible physical explanations. Going in-depth into such physical models and discussing their viability is not the focus of this work, but we will briefly present some published possibilities below.

Anisotropic inflation
Standard inflationary theory give rise to Gaussian, statistically isotropic fluctuations. Ackerman et al. (2007) [19] proposes a possible small violation of rotational invariance during the inflationary era, which could cause a quadrupolar power asymmetry in a specific direction. This is thus not directly related to the dipole case, but is an example of a physical model yielding an anisotropic directional effect in the CMB. The effect is also shown, for a simple model, to be approximately scale-invariant.

Erickcek et al. (2008) [20] proposes an alternative inflationary theory where a superhorizon perturbation to the curvaton field could generate a hemispherical power asymmetry, where the simplest model again predicts a scale-invariant modulation. In a more recent paper, Erickcek et al. (2009) [21] presents a modified version of this theory which will produce a scale-dependent asymmetry where the amplitude decays with l.

Anisotropic dark energy
Koivisto and Mota (2008) [22] suggest an anisotropic equation of state for dark energy, which would cause a direction-dependent expansion of the universe at late times. This could be a cause of large-angle anomalies in the observed CMB.

Chapter 2

Mathematical description of the CMB

The CMB is, in essence, just a temperature field. On each point in the sky, observations give a certain temperature. What we are usually interested in when studying the CMB, are not the actual temperatures though, but rather deviations from the mean. Let's call this field $\Delta T(\theta, \phi)$. Measuring these deviations allow us to quantify the anisotropies to test cosmological models. In order to make such tests, we need to be able to describe the observed field mathematically.

2.1 Decomposing into spherical harmonics

Observations of the CMB are taken in real space on the sphere surrounding us. When doing analyses, however, it is often more convenient to look at the field in a Fourier-type space. On the sphere this space is that of the spherical harmonics[1], and the temperature field is there decomposed as follows:

$$\Delta T(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta, \phi). \tag{2.1}$$

Here, Y_{lm} are of course the spherical harmonics, and

$$a_{lm} = \int_{\Omega} \Delta T(\theta, \phi) Y^*_{lm}(\theta, \phi) \, \mathrm{d}\Omega \tag{2.2}$$

are the components of the transform. They are complex numbers, with $a_{l-m} = (-1)^m a^*_{lm}$ since the temperature field is real. Having done this transformation, all the information present in the temperature field is also present in the a_{lm}'s. Just as the different modes k of an ordinary Fourier decomposition correspond to different size scales, the different modes l of the spherical harmonics decomposition correspond to

[1] See the appendix, section A.3.

different angular scales. A rough estimate of the angle is $\theta \sim 180^\circ/l$. The first few modes are known as the monopole ($l = 0$), the dipole ($l = 1$), the quadrupole ($l = 2$) and the octopole ($l = 3$). This dipole *mode* must not be confused with the dipole *modulation* mentioned earlier, however, as that is a modulation extending to all scales.

From the decomposition we can calculate the *power spectrum*, which gives the power of the CMB fluctuations on the different angular scales. This spectrum is one of the most important quantities in modern cosmology. From analysing it, we can constrain many of the parameters in the ΛCDM model, such as the dark matter density Ω_c and the spectral index n_s, because the different parameters have different effects on the shape of the spectrum. The power spectrum is given by, as shown in equation (1.11):

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2.$$

There is a fundamental uncertainty in how good a knowledge we can have about the C_l's, called *cosmic variance*[2]. This stems from the fact that the a_{lm}'s are essentially drawn from a distribution with variance C_l. So each a_{lm} has the same variance C_l for any given l, regardless of the value of m. There are $2l + 1$ a_{lm}'s for each l, so for higher l, we get much more information on the underlying variance than for low l. This uncertainty on the information is the cosmic variance, and it is given by

$$\left(\frac{\Delta C_l}{C_l}\right)_{\text{cosmic variance}} = \sqrt{\frac{2}{2l+1}}. \tag{2.3}$$

In figure 1.3, of the power spectrum as measured by WMAP, we see the cosmic variance represented by the shaded band around the curve.

To get a feel for how the map is decomposed into different multipoles, I have extracted the first 6 multipoles from a simulation and shown how these combine into a map in figure 2.1. Comparing the outcome with the full simulated map of figure 2.2(a) we see that already some of the main structures are present using only these low l.

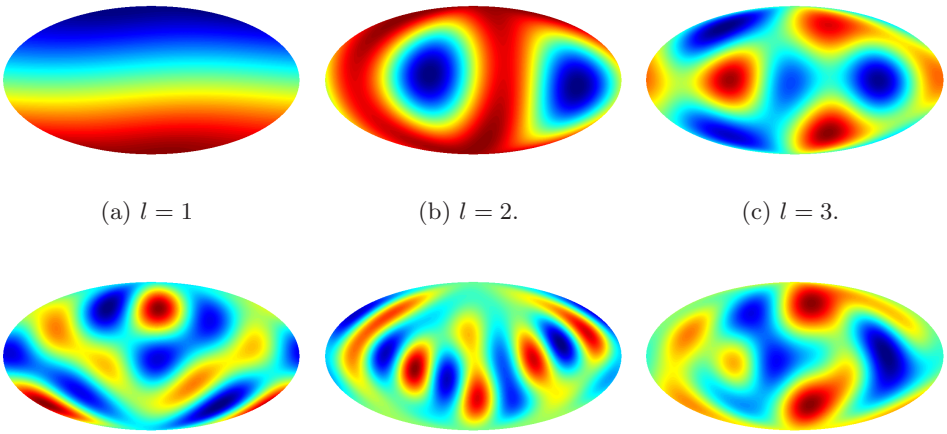
2.2 Observations into pixelized map

The data taken by observations can in the simplest form be written in the following way (see equation 1.18)

$$\mathbf{d}(\hat{\mathbf{n}}) = \mathbf{s}(\hat{\mathbf{n}}) + \mathbf{n}(\hat{\mathbf{n}}). \tag{2.4}$$

Here, $\mathbf{s}$ is the actual CMB signal and $\mathbf{n}$ is instrumental noise, while $\mathbf{d}$ is the data output. The unit vector $\hat{\mathbf{n}}$ denotes direction on the sky. The picture isn't quite as simple, however, as there are plenty of noise sources that aren't merely instrumental. These include the doppler-induced dipole caused by the satellite's movement around the sky, the signal contamination by our own galaxy, as well as that by other galaxies and foreground emissions. These contaminations need to be painstakingly removed from

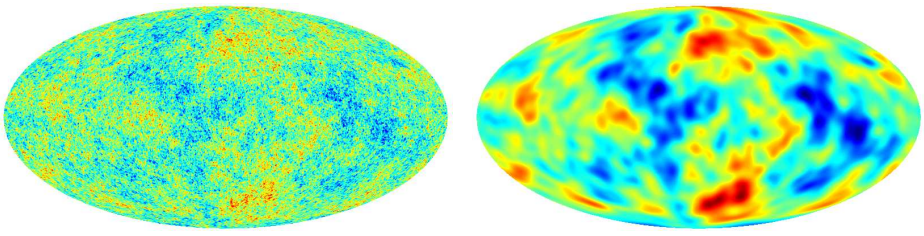
[2]See Dodelson [1].



(a) $l = 1$ (b) $l = 2$. (c) $l = 3$.

(d) $l = 4$. (e) $l = 5$. (f) Combined map of multipoles from $l = 0$ to $l = 5$.

Figure 2.1: Different multipoles extracted from a simulated map and their combined map. I have left out the monopole since it is just a uniform field.



(a) Smoothed with a 0.35 degree beam, equivalent to WMAP.

(b) Smoothed with a 7.0 degree beam, equivalent to COBE.

Figure 2.2: Simulated maps with different beam smoothing widths.

the data before we start analyzing, in order not to get false results. Also, the satellite itself does not have infinite resolution, so the signal gets smoothed by the instrumental beam. Scales smaller than what the beam is able to pick up will not be discerned in the data output (see figure 2.2). A more thorough picture of the data is therefore

$$\mathbf{d}(\hat{\mathbf{n}}) = \mathbf{B}\big(\mathbf{s}(\hat{\mathbf{n}}) + \mathbf{f}(\hat{\mathbf{n}})\big) + \mathbf{n}(\hat{\mathbf{n}}), \tag{2.5}$$

where $\mathbf{B}$ represents the smoothing by the instrumental beam, and $\mathbf{f}$ the foreground, galaxy and doppler dipole noise contaminations. In pixel space the beam smoothing is a convolution operation, while in harmonic space it is a simple multiplication. This is one of the reasons why we often wish to work in harmonic space; it makes life easier!

2.2.1 Galaxy and foreground removal

By measuring the signal at several different frequencies, and looking at other observations of galaxies, one is able to distinguish the actual CMB signal from the foregrounds. The frequency region where we measure the CMB is also one where the foreground emissions are quite low, see figure 2.3. Note that the figure only shows the level of the CMB *anisotropies*; the full signal is much stronger than any foregrounds. However, the anisotropies are what we are interested in. The WMAP satellite is also designed so as to only measure the temperature *difference* between two parts of the sky, thereby minimizing concerns over the 2.7K averaged full signal (monopole) of the CMB. Details on the design of the instrument are given in Bennett et al. (2003) [23]. For more information on the foreground contaminations in the WMAP observations see Gold et al. (2011) [24].

In the analysis we will use the foreground-reduced maps provided by the WMAP team, and the galaxy, point sources and other contaminations will be dealt with by applying a mask/sky-cut, effectively setting the signal at the pixels in question to 0. The kinematic dipole has been removed also from the raw maps. There are two commonly used masks that are provided, named KQ75 and KQ85 (because they retain approximately 75% and 85% of the sky, respectively), of which we will mostly use the latter. It is shown in figure 2.4. Note the thick band in the center, which masks out the galaxy. Each pixel to be masked out is set to 0, and the rest are set to 1. The way we apply this mask in our analysis is to multiply it with our inverse noise model, so that pixels that are masked out are effectively said to have infinite noise. Additionaly, we also project out template maps corresponding to the monopole and dipole modes, details are given in chapter 6.

The noise model for the intrinsic instrumental pixel noise is calculated from information given in the sky-maps, where the number of observations per pixel is given. The different number of observations makes the noise rms differ with each pixel, and thus across the sky. More information on this is given in section 6.2.1. The noise is considered to be uncorrelated between pixels, thus making the noise covariance diagonal in pixel space.

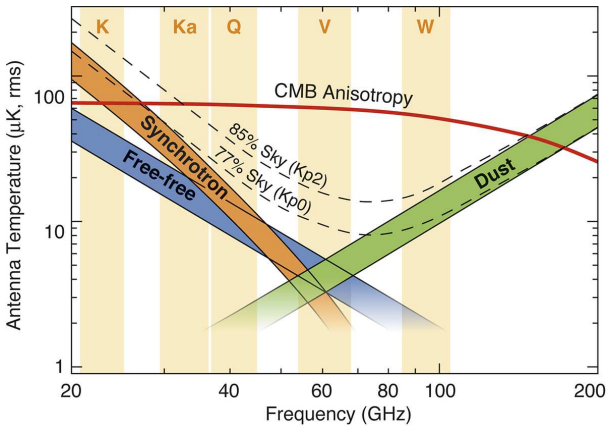


Figure 2.3: This picture compares the radiation from the CMB with those of three sources of foreground emission: Synchrotron, free-free and thermal dust emission. Also shown is the galactic emission using two common sky-cuts. Image courtesy of the WMAP Science Team.

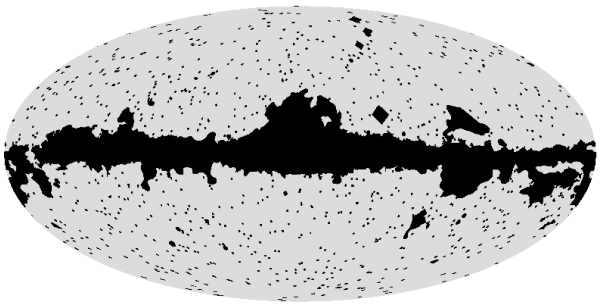


Figure 2.4: The KQ85 mask, retaining 78.3% of the sky [24].

2.2.2 Dealing with the instrumental beam

In harmonic space the beam smoothing function b_l is given by

$$b_l = \exp\left(-\frac{1}{2}l(l+1)\sigma^2\right), \text{ where } \sigma^2 = \left(\text{FWHM} \cdot \frac{\pi}{180} \cdot \frac{1}{\sqrt{8\ln(2)}}\right)^2. \tag{2.6}$$

This expression is valid for a Gaussian beam, which is a good approximation to the beams of the detectors on WMAP and other experiments. FWHM is the size (full width at half maximum) of the beam used, for WMAP it equals 0.51 degrees in the Q-band, 0.35 degrees in the V-band and 0.22 in the W-band[3]. We see here that this function falls off rapidly with increasing l. What happens when the instrument maps the sky is, in effect, that the "true" (disregarding contaminations) a_{lm}'s get multiplied by this beam function in equation 2.1. So at large l (small angular scales), it's obvious that the signal quickly gets diminished. Example graphs are shown in figure 2.5.

In addition we also get a smoothing effect by virtue of the map being stored as a set of pixels. Each pixel has a certain size, which again serves to smooth out the signal. This *pixel window* function, p_l, is provided in the HEALPix[4] package for different map resolutions. We deal with both of these features simultaneously, simply by multiplying the two functions together.

How the beam function is incorporated into the estimation procedure will be shown in coming chapters. Specifically, simulated maps smoothed with the same beam as the data are used to normalize the estimated parameters.

2.2.3 Representing the decomposed map

As mentioned we like to work in harmonic space since it simplifies many of our calculations. However the map we receive is the pixelized version. When we wish to decompose this into harmonic components a_{lm}, how to we choose how many multipoles to include in order to sufficiently represent the map? Choosing too few multipoles $l_{\max}$ will be akin to smoothing the map further, which we do not want, and choosing too many multipoles will cause other problems due to oversampling, such as aliasing. There is at least no reason to choose an $l_{\max}$ corresponding to scales smaller than what can be discerned from the pixel window. HEALPix stores the maps using $N_{\text{pix}} = 12 \cdot N_{\text{side}}^2$ number of pixels, where N_{side} is a measure of the resolution. The resolution of each pixel in terms of solid angles is $\Omega_{\text{pix}} = \frac{4\pi}{N_{\text{pix}}}$. The default and recommended size of $l_{\max}$ is $2 \cdot N_{\text{side}}$, and this is what we utilize throughout this work. We can convince ourselves that this value is reasonable if we observe that the angular resolution θ_{pix} of one pixel is $\theta_{\text{pix}} \equiv \sqrt{\Omega_{\text{pix}}} = \frac{2\sqrt{\pi}}{\sqrt{N_{\text{pix}}}}$, and remember that the angular scale represented

[3]The FWHM values are taken from the WMAP: Seven-Year Explanatory Supplement [25]. See chapter 6 for more on the different bands.

[4]See appendix A.1

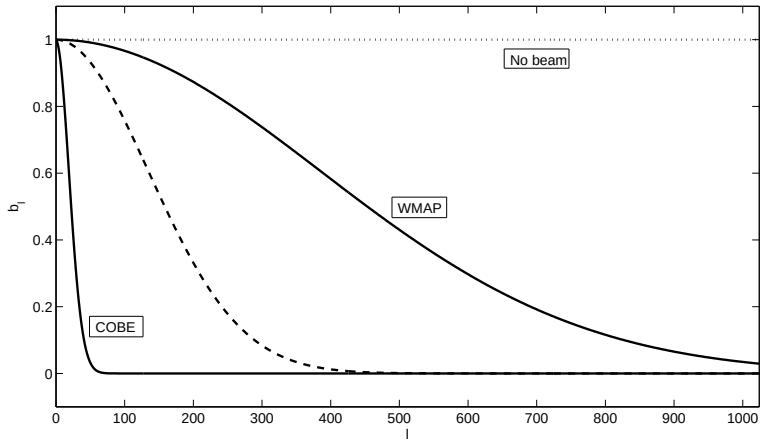


Figure 2.5: The beam smoothing function b_l for different beam widths. The line labeled WMAP corresponds to a beam of 0.35°, the one labeled COBE to 7.0°, and the dashed line to a beam of 1.0°. The dotted line represents no beam.

by l is roughly $\frac{\pi}{l}$. Calculating l for our smallest scale in terms of N_{side} gives

$$\frac{l}{\pi} = \frac{\sqrt{N_{\text{pix}}}}{2\sqrt{\pi}} = \frac{2\sqrt{3}N_{\text{side}}}{2\sqrt{\pi}} \approx N_{\text{side}}$$
$$\Rightarrow l \approx \pi N_{\text{side}}.$$

2.2.4 Downgrading maps

When we downgrade higher-resolution maps to a lower resolution we must smooth, or convolve, the new map with a beam corresponding to the new resolution. How to we choose the size of this beam, its FWHM? We have the beam function given by equation 2.6, and from this we can get an expression for the FWHM of a beam corresponding to a certain l_{max}:

$$\text{FWHM} = \frac{180}{\pi} \cdot \sqrt{-\frac{16\ln(2)\ln(b_{l_{\text{max}}})}{l_{\text{max}}(l_{\text{max}}+1)}}. \tag{2.7}$$

If we set $b_{l_{\text{max}}} = 0.1$, for instance, we get a FWHM of 4.5° for an l_{max} of 64. When downgrading we also need to apply the pixel window of the new resolution. However, before applying the new beams, we must deconvolve with the original beam and pixel window, and this is just a division in harmonic space.

When applying a sky-cut to downgraded maps we need to downgrade the mask as well. One way of doing this is to go through the original-size mask, and for each pixel in the new resolution set this to 1 if more than half of the corresponding pixels in the original mask is 1. Downgrading is not particularly important in this work, since the

quadratic estimator performs quickly also on full-size maps extending to more than 1000 multipoles. However, since previous analyses have often relied heavily on downgrading it is in our interest to be aware of how the process is performed.

2.3 The signal covariance matrix

For a vector of random variables $\mathbf{X} = [X_0\ X_1\ \ldots\ X_n]^T$, the *covariance matrix* $\mathbf{C}$ has components given by [26]

$$C_{ij} = \mathrm{Cov}(X_i, X_j) = \big\langle (X_i - \mu_i)(X_j - \mu_j) \big\rangle, \tag{2.8}$$

where $\mu_i = \langle X_i \rangle$. An alternative form of the equation is

$$C_{ij} = \langle X_i X_j \rangle - \mu_i \mu_j. \tag{2.9}$$

For our data model, equation 2.5, we get

$$C_{ij} = \langle \mathbf{d}(\hat{\mathbf{n}}_i)\mathbf{d}(\hat{\mathbf{n}}_j) \rangle, \tag{2.10}$$

or alternatively,

$$\mathbf{C} = \langle \mathbf{d}\mathbf{d}^T \rangle. \tag{2.11}$$

Written out in full this becomes (disregarding the beam)

$$\begin{aligned} \mathbf{C} &= \big\langle (\mathbf{s} + \mathbf{f} + \mathbf{n})(\mathbf{s} + \mathbf{f} + \mathbf{n})^T \big\rangle && (2.12)\\ &= \big\langle \mathbf{s}\mathbf{s}^T + \mathbf{f}\mathbf{f}^T + \mathbf{n}\mathbf{n}^T \big\rangle && (2.13)\\ &= \mathbf{S} + \mathbf{F} + \mathbf{N}, && (2.14) \end{aligned}$$

since the cross-correlation terms vanish. The covariances for each contribution is shown as $\mathbf{S} = \langle \mathbf{s}\mathbf{s}^T \rangle$ (the signal covariance), $\mathbf{F} = \langle \mathbf{f}\mathbf{f}^T \rangle$ (the covariance of the foregrounds) and $\mathbf{N} = \langle \mathbf{n}\mathbf{n}^T \rangle$ (the noise covariance).

Including the beam this would only affect the terms with the signal and the foregrounds. For the signal term we would get

$$\big\langle \mathbf{B}\mathbf{s}(\mathbf{B}\mathbf{s})^T \big\rangle = \big\langle \mathbf{B}\mathbf{s}\mathbf{s}^T\mathbf{B}^T \big\rangle = \mathbf{B} \big\langle \mathbf{s}\mathbf{s}^T \big\rangle \mathbf{B}^T = \mathbf{B}\mathbf{S}\mathbf{B}^T, \tag{2.15}$$

since the beam function is not affected by the averaging operation.

2.3.1 Including the dipole model

When including the dipole model the part of the data related to the signal gets transformed, from simply $\mathbf{s}(\hat{\mathbf{n}})$ to $(1 + A(\hat{\mathbf{n}} \cdot \hat{\mathbf{p}}))\mathbf{s}(\hat{\mathbf{n}})$, where A here is the dipole modulation amplitude and $\hat{\mathbf{p}}$ the direction. The covariance for the signal thus becomes, in directional components

$$\begin{aligned} \mathbf{S}(\hat{\mathbf{n}}, \hat{\mathbf{m}}) &= \langle \mathbf{s}(\hat{\mathbf{n}})\mathbf{s}(\hat{\mathbf{m}}) \rangle = \big\langle [(1 + A(\hat{\mathbf{n}} \cdot \hat{\mathbf{p}}))\mathbf{s}(\hat{\mathbf{n}})]\,[(1 + A(\hat{\mathbf{m}} \cdot \hat{\mathbf{p}}))\mathbf{s}(\hat{\mathbf{m}})] \big\rangle && (2.16)\\ &= \big\langle (1 + A(\hat{\mathbf{n}} \cdot \hat{\mathbf{p}}))\mathbf{s}(\hat{\mathbf{n}})\mathbf{s}(\hat{\mathbf{m}})(1 + A(\hat{\mathbf{m}} \cdot \hat{\mathbf{p}})) \big\rangle && (2.17)\\ &= (1 + A(\hat{\mathbf{n}} \cdot \hat{\mathbf{p}}))\mathbf{S}_{\mathrm{iso}}(\hat{\mathbf{n}}, \hat{\mathbf{m}})(1 + A(\hat{\mathbf{m}} \cdot \hat{\mathbf{p}})), && (2.18) \end{aligned}$$

where $\mathbf{S}_{\rm iso}$ is the covariance in the non-modulated case, equal to C_l in harmonic space. In pixel space it can be calculated through (the field is real so taking the complex conjugate has no effect, but it makes the calculations somewhat easier)

$$\mathbf{S}_{\rm iso}(\hat{\mathbf{n}},\hat{\mathbf{m}}) = \left\langle \mathbf{s}(\hat{\mathbf{n}})\mathbf{s}(\hat{\mathbf{m}})^* \right\rangle = \left\langle \left(\sum_{lm} a_{lm} Y_{lm}(\hat{\mathbf{n}})\right)\left(\sum_{l'm'} a^*_{l'm'} Y^*_{l'm'}(\hat{\mathbf{m}})\right)\right\rangle \tag{2.19}$$

$$= \left\langle \sum_{lm}\sum_{l'm'} a_{lm} a^*_{l'm'} Y_{lm}(\hat{\mathbf{n}}) Y^*_{l'm'}(\hat{\mathbf{m}})\right\rangle \tag{2.20}$$

$$= \sum_{lm}\sum_{l'm'} \left\langle a_{lm} a^*_{l'm'}\right\rangle Y_{lm}(\hat{\mathbf{n}}) Y^*_{l'm'}(\hat{\mathbf{m}}) \tag{2.21}$$

$$= \sum_{lm}\sum_{l'm'} C_l \delta_{ll'}\delta_{mm'} Y_{lm}(\hat{\mathbf{n}}) Y^*_{l'm'}(\hat{\mathbf{m}}), \tag{2.22}$$

since the averaging operation does not affect the spherical harmonics. Continuing, we get

$$\mathbf{S}_{\rm iso}(\hat{\mathbf{n}},\hat{\mathbf{m}}) = \sum_l C_l \sum_{m=-l}^{l} Y_{lm}(\hat{\mathbf{n}}) Y^*_{lm}(\hat{\mathbf{m}}). \tag{2.23}$$

The sum over m invokes a relationship between the spherical harmonics and the *Legendre polynomials* P_l, where

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

To close, then, we get

$$\mathbf{S}_{\rm iso}(\hat{\mathbf{n}},\hat{\mathbf{m}}) = \frac{1}{4\pi}\sum_l (2l+1) C_l P_l(\hat{\mathbf{n}}\cdot\hat{\mathbf{m}}). \tag{2.24}$$

Chapter 3

Statistical estimators for the dipole model

The dipole modulation model has been tested in the literature using various estimators. We will in detail present and derive the optimal quadratic estimator of [18], which is a fast, approximate method that allows analysis of the full-resolution data sets. We also give a short overview of the exact maximum-likelihood method used in [12].

3.1 Quadratic estimator (Hanson & Lewis)

The basic concept behind the quadratic estimator is to approximate the likelihood (strictly speaking the log-likelihood) as a quadratic function near its maximum. This is often a reasonable approximation, especially if the underlying likelihood is close to Gaussian (for a truly Gaussian likelihood it would be exact). This approach will allow us to estimate the parameters much more rapidly than if one were to explore the whole parameter space as in an exact, maximum-likelihood approach. What we do is to Taylor-expand (the derivative of) the log-likelihood with respect to our parameter vector, and then use the Newton-Raphson method to approximate the parameters.

In this thesis the estimator is occasionally abbreviated as QML (Quadratic Maximum Likelihood).

3.1.1 Derivation of the estimator

Follows Dodelson (2003), pp. 362, and Hanson & Lewis (2009). We begin by constructing a general quadratic estimator for an arbitrary parameter λ, and then apply it to the dipole model.

The CMB likelihood function is given by[1]

$$\mathcal{L} = \frac{1}{(2\pi)^{N_{\rm pix}/2}|C|^{1/2}} \exp\left[-\frac{1}{2}\Theta^\dagger C^{-1}\Theta\right], \tag{3.1}$$

[1] See Dodelson (2003), equation 11.20 (page 341).

where Θ is the observed CMB and C is the covariance. What we want is to *maximize* the likelihood, i.e. find the point where the derivative of the likelihood with respect to the parameter(s) is 0:

$$\left.\frac{\partial \mathcal{L}}{\partial \lambda}\right|_{\lambda=\bar{\lambda}} = 0. \tag{3.2}$$

Here, $\bar{\lambda}$ is the desired value of the parameter we are interested in, and thus the value we wish to estimate. What we do next is to Taylor expand the derivative around some point λ_0. In the following we will use the simplified notation of derivatives, denoted by commas:

$$\mathcal{L}_{,\lambda} \equiv \frac{\partial \mathcal{L}}{\partial \lambda}\,, \quad \mathcal{L}_{,\lambda\lambda} \equiv \frac{\partial^2 \mathcal{L}}{\partial \lambda^2}. \tag{3.3}$$

The Taylor expansion is then:

$$\mathcal{L}_{,\lambda}(\bar{\lambda}) = \mathcal{L}_{,\lambda}(\lambda_0) + \mathcal{L}_{,\lambda\lambda}(\lambda_0)(\bar{\lambda} - \lambda_0) + \ldots \tag{3.4}$$

The left-hand side equals 0, and discarding the higher-order terms we get

$$\bar{\lambda} \simeq \lambda_0 - \frac{\mathcal{L}_{,\lambda}(\lambda_0)}{\mathcal{L}_{,\lambda\lambda}(\lambda_0)}. \tag{3.5}$$

In order for this solution to be viable, the approximation of discarding higher-order terms must be a good one. That is the case for functions that are close to quadratic. However, the likelihood function $\mathcal{L}$ is *not* close to being quadratic, in fact, it is usually approximated as a Gaussian. That is what we will do also, and exploit the fact that the logarithm of a Gaussian function is quadratic. Repeating the derivation for $\ln \mathcal{L}$ in place of $\mathcal{L}$ above, we get (the derivation is exactly similar, since the logarithm is maximized at the same place as the actual function)

$$\bar{\lambda} \simeq \lambda_0 - \frac{(\ln \mathcal{L})_{,\lambda}(\lambda_0)}{(\ln \mathcal{L})_{,\lambda\lambda}(\lambda_0)}. \tag{3.6}$$

This, then, becomes our estimator $\hat{\lambda}$ for the value of $\bar{\lambda}$. It is really an iterative scheme, the Newton-Raphson method, where the n'th iteration would yield

$$\hat{\lambda}_{n+1} = \hat{\lambda}_n - \frac{(\ln \mathcal{L})_{,\lambda}(\hat{\lambda}_n)}{(\ln \mathcal{L})_{,\lambda\lambda}(\hat{\lambda}_n)}. \tag{3.7}$$

If the starting trial value $\hat{\lambda}_0$ is chosen sufficiently close to the true value, however, the estimator should yield good results with only one iteration[2].

The next step is to calculate the derivatives in equation 3.6. We have our likelihood function given by equation 3.1, and its logarithm is

$$\ln \mathcal{L} = -\frac{1}{2}\Theta^\dagger C^{-1}\Theta - \frac{N_{\text{pix}}}{2}\ln(2\pi) - \frac{1}{2}\ln|C|. \tag{3.8}$$

[2]This is what Hanson & Lewis (and also us) choose to do, assuming that the dipole anisotropy is weak and using the trial value of no anisotropy. For a discussion of using more iterations, see section 6.2.3.

Taking the derivative the constant term vanishes:

$$(\ln \mathcal{L})_{,\lambda} = \frac{\partial}{\partial \lambda}\left(-\frac{1}{2}\Theta^\dagger C^{-1}\Theta - \frac{1}{2}\ln|C|\right). \tag{3.9}$$

Only the covariance C depends on λ, so we get

$$(\ln \mathcal{L})_{,\lambda} = -\frac{1}{2}\Theta^\dagger C^{-1}_{,\lambda}\Theta - \frac{1}{2}\left(\ln|C|\right)_{,\lambda}. \tag{3.10}$$

Utilizing the following two identities,

$$\ln|C| = \mathrm{Tr}(\ln C) \quad \text{and} \quad C^{-1}_{,\lambda} = -C^{-1}C_{,\lambda}C^{-1}, \tag{3.11}$$

we get:

$$(\ln \mathcal{L})_{,\lambda} = \frac{1}{2}\Theta^\dagger C^{-1}C_{,\lambda}C^{-1}\Theta - \frac{1}{2}\mathrm{Tr}\left(C^{-1}C_{,\lambda}\right). \tag{3.12}$$

This is similar to equation 3 in Hanson & Lewis, except that they write the log-likelihood as $\mathcal{L}$. There are also some other differences (disregarding the difference in notation). First off, they take the derivative with respect to a vector of parameters $\boldsymbol{\lambda}$. This is just a generalization, and our equation would become

$$(\ln \mathcal{L})_{,\boldsymbol{\lambda}} = \frac{1}{2}\Theta^\dagger C^{-1}C_{,\boldsymbol{\lambda}}C^{-1}\Theta - \frac{1}{2}\mathrm{Tr}\left(C^{-1}C_{,\boldsymbol{\lambda}}\right). \tag{3.13}$$

But they also have some complex transposes and different signs. Their equation reads

$$(\ln \mathcal{L})_{,\boldsymbol{\lambda}^\dagger} = -\frac{1}{2}\Theta^\dagger C^{-1}C_{,\boldsymbol{\lambda}^\dagger}C^{-1}\Theta + \frac{1}{2}\mathrm{Tr}\left(C^{-1}C_{,\boldsymbol{\lambda}^\dagger}\right). \tag{3.14}$$

This does not mean there is a discrepancy: The complex transpose in the derivative is just the difference between considering column- or row-vectors, and sign difference suggests a different convention. We can look at their $\ln \mathcal{L}$ as actually being $-\ln \mathcal{L}$. In that case the maximization becomes a minimization, but the value at this point remains the same. In the following we will use the expression of H&L for easy comparison.

Hanson & Lewis now use the identity $\mathrm{Tr}(A) = \langle \mathbf{x}^\dagger A C^{-1}\mathbf{x}\rangle$, where $\mathbf{x}$ is a vector of Gaussian random variables with covariance C (this is precisely what Θ in the above equation is), to write equation 3.14 as

$$(\ln \mathcal{L})_{,\boldsymbol{\lambda}^\dagger} = \langle \mathcal{H}\rangle - \mathcal{H}, \tag{3.15}$$

where

$$\mathcal{H} = \frac{1}{2}(C^{-1}\Theta)^\dagger C_{,\boldsymbol{\lambda}^\dagger}(C^{-1}\Theta). \tag{3.16}$$

The second derivative $(\ln \mathcal{L})_{,\lambda\lambda}$ is related to the Fisher information, which can be written as

$$\mathcal{F} = \left\langle -\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2}\right\rangle. \tag{3.17}$$

From these facts and equation 3.7 (replacing the second derivative with its average) we get the iterative estimator

$$\hat{\boldsymbol{\lambda}}_{n+1} = \hat{\boldsymbol{\lambda}}_n + \mathcal{F}_n^{-1}[\mathcal{H}_n - \langle\mathcal{H}\rangle_n], \tag{3.18}$$

where the + sign occurs because we're really working with $-\ln\mathcal{L}$. We suppose a small modulation so we only use one iteration, with $\hat{\boldsymbol{\lambda}}_0 = \mathbf{0}$. The zeroth part of the quadratic estimator, $\mathcal{H}_0$, can then be written as

$$\mathcal{H}_0 = \frac{1}{2}\bar{\Theta}^\dagger \frac{\partial C}{\partial \boldsymbol{\lambda}^\dagger}\bar{\Theta}, \tag{3.19}$$

where $\bar{\Theta} = C^{-1}|_{\boldsymbol{\lambda}=0}\Theta$ is the inverse-variance weighted signal[3]. Component-wise the equation can be written as the following, where lm denotes rows and $l'm'$ columns:

$$\mathcal{H}_0 = \frac{1}{2}\sum_{lm,l'm'}\left[\frac{\partial C_{lm,l'm'}}{\partial \boldsymbol{\lambda}^\dagger}\right]\bar{\Theta}^*_{lm}\bar{\Theta}_{l'm'}. \tag{3.20}$$

This is equation 9 in Hanson & Lewis. Our estimator, then, becomes

$$\hat{\boldsymbol{\lambda}}_1 = \mathcal{F}^{-1}\left[\mathcal{H}_0 - \langle\mathcal{H}_0\rangle\right]. \tag{3.21}$$

3.1.2 The estimator applied to the dipole anisotropy case

Here we will follow the notation in Hanson & Lewis (2009). The dipole modulation described in section 1.3 can be written as

$$\Theta_f(\hat{\mathbf{n}}) = [1 + f(\hat{\mathbf{n}})]\,\Theta^i_f(\hat{\mathbf{n}}), \tag{3.22}$$

where Θ_f is the observed CMB temperature field, Θ^i_f is the intrinsic statistically isotropic CMB, and f is the dipole modulation field. The subscript $_f$ denotes restriction to a range of angular scales $l \leq l_{\max}$. Recall that the modulation field is of the form $f(\hat{\mathbf{n}}) = A(\hat{\mathbf{p}}\cdot\hat{\mathbf{n}})$, where A is the amplitude and $\hat{\mathbf{p}}$ the dipole direction. There are thus three parameters to determine the dipole: The amplitude and the direction (θ,ϕ). Alternatively, the dipole can be given as a harmonic field, with three parameters having $l = 1$ and $m = -1, 0, 1$. We expand each side in the equation in terms of spherical harmonics[4]:

$$\sum_{lm}\Theta_{lm}Y_{lm} = \left(1 + \sum_{l''m''} f_{l''m''}Y_{l''m''}\right)\sum_{l'm'}\Theta^i_{l'm'}Y_{l'm'} \tag{3.23}$$

$$= \sum_{l'm'}\Theta^i_{l'm'}Y_{l'm'} + \sum_{l''m''}\sum_{l'm'} f_{l''m''}\Theta^i_{l'm'}Y_{l'm'}Y_{l''m''}. \tag{3.24}$$

[3]The inverse-variance weighted signal is computed using the *conjugate gradients* algorithm, see section 4.1

[4]See equation A.11 in the appendix, section A.3.

Rewriting the dummy indices in the sum on the left-hand side from $lm \rightarrow l'm'$ and multiplying each side with Y^*_{lm} we get:

$$\sum_{l'm'} \Theta_{l'm'} Y_{l'm'} Y^*_{lm} = \sum_{l'm'} \Theta^i_{l'm'} Y_{l'm'} Y^*_{lm} + \sum_{l'm'} \sum_{l''m''} \Theta^i_{l'm'} f_{l''m''} Y_{l'm'} Y_{l''m''} Y^*_{lm}. \tag{3.25}$$

Next we integrate each term over all angles Ω (remember that the spherical harmonics coefficients Θ_{lm} and f_{lm} are just numbers, with no angular dependence):

$$\begin{aligned}\sum_{l'm'} \Theta_{l'm'} \int_\Omega Y_{l'm'} Y^*_{lm} \, \mathrm{d}\Omega = &\sum_{l'm'} \Theta^i_{l'm'} \int_\Omega Y_{l'm'} Y^*_{lm} \, \mathrm{d}\Omega \\ &+ \sum_{l'm'} \sum_{l''m''} \Theta^i_{l'm'} f_{l''m''} \int_\Omega Y_{l'm'} Y_{l''m''} Y^*_{lm} \, \mathrm{d}\Omega.\end{aligned} \tag{3.26}$$

Now we use the orthonormality relation of the spherical harmonics (equation A.10) to get

$$\sum_{l'm'} \Theta_{l'm'} \delta_{ll'} \delta_{mm'} = \sum_{l'm'} \Theta^i_{l'm'} \delta_{ll'} \delta_{mm'} + \sum_{l'm'} \sum_{l''m''} \Theta^i_{l'm'} f_{l''m''} \int_\Omega Y_{l'm'} Y_{l''m''} Y^*_{lm} \, \mathrm{d}\Omega. \tag{3.27}$$

This gives

$$\Theta_{lm} = \Theta^i_{lm} + \sum_{l'm',l''m''} \Theta^i_{l'm'} f_{l''m''} \int_\Omega Y^*_{lm} Y_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega. \tag{3.28}$$

From this expression we can calculate the covariance $C_{l'm',l''m''} = \langle \Theta_{l'm'} \Theta^*_{l''m''} \rangle$. We do this to first order in f_{lm}, since we assume the modulation to be small. The covariance is then given by

$$\left\langle \left[\Theta^i_{l'm'} + \sum_{lm,l'''m'''} \Theta^i_{lm} f_{l'''m'''} \int_\Omega \ldots \mathrm{d}\Omega \right] \left[\Theta^{i*}_{l''m''} + \sum_{lm,l'''m'''} \Theta^{i*}_{lm} f^*_{l'''m'''} \int_\Omega \ldots \mathrm{d}\Omega \right] \right\rangle. \tag{3.29}$$

Writing this out in full we get

$$\begin{aligned}C_{l'm',l''m''} \simeq \Bigg\langle &\Theta^i_{l'm'} \Theta^{i*}_{l''m''} + \Theta^i_{l'm'} \sum_{lm,l'''m'''} \Theta^{i*}_{lm} f^*_{l'''m'''} \int_\Omega Y_{l''m''} Y^*_{lm} Y^*_{l'''m'''} \, \mathrm{d}\Omega \\ &+ \left[\sum_{lm,l'''m'''} \Theta^i_{lm} f_{l'''m'''} \int_\Omega Y^*_{l'm'} Y_{lm} Y_{l'''m'''} \, \mathrm{d}\Omega \right] \Theta^{i*}_{l''m''} \Bigg\rangle\end{aligned} \tag{3.30}$$

$$\begin{aligned}= \Bigg\langle &\Theta^i_{l'm'} \Theta^{i*}_{l''m''} + \sum_{lm,l'''m'''} \Theta^i_{l'm'} \Theta^{i*}_{lm} f^*_{l'''m'''} \int_\Omega Y_{l''m''} Y^*_{lm} Y^*_{l'''m'''} \, \mathrm{d}\Omega \\ &+ \sum_{lm,l'''m'''} \Theta^i_{lm} \Theta^{i*}_{l''m''} f_{l'''m'''} \int_\Omega Y^*_{l'm'} Y_{lm} Y_{l'''m'''} \, \mathrm{d}\Omega \Bigg\rangle.\end{aligned} \tag{3.31}$$

Since $\langle \Theta^i_{lm} \Theta^{i*}_{l'm'} \rangle = \delta_{ll'}\delta_{mm'} C_l$ (where C_l is the isotropic covariance), and the integrals over the spherical harmonics as well as the f_{lm}'s are not affected by the averaging, we get

$$\begin{aligned} C_{l'm',l''m''} = \delta_{l'l''}\delta_{m'm''} C_{l'} + \sum_{lm,l'''m'''} \delta_{ll'}\delta_{mm'} C_{l'} f^*_{l'''m'''} \int_\Omega Y_{l''m''} Y^*_{lm} Y^*_{l'''m'''} \, \mathrm{d}\Omega \\ + \sum_{lm,l'''m'''} \delta_{ll''}\delta_{mm''} C_l f_{l'''m'''} \int_\Omega Y^*_{l'm'} Y_{lm} Y_{l'''m'''} \, \mathrm{d}\Omega. \end{aligned} \tag{3.32}$$

Truncating the sums give

$$\begin{aligned} C_{l'm',l''m''} = \delta_{l'l''}\delta_{m'm''} C_{l'} + \sum_{l'''m'''} C_{l'} f^*_{l'''m'''} \int_\Omega Y_{l''m''} Y^*_{l'm'} Y^*_{l'''m'''} \, \mathrm{d}\Omega \\ + \sum_{l'''m'''} C_{l''} f_{l'''m'''} \int_\Omega Y^*_{l'm'} Y_{l''m''} Y_{l'''m'''} \, \mathrm{d}\Omega. \end{aligned} \tag{3.33}$$

Now we change the dummy variable in the sums from $l'''m''' \to lm$, and observe that the sums run over the same indices:

$$\begin{aligned} C_{l'm',l''m''} = \delta_{l'l''}\delta_{m'm''} C_{l'} + \sum_{lm} \Bigg(C_{l'} f^*_{lm} \int_\Omega Y_{l''m''} Y^*_{l'm'} Y^*_{lm} \, \mathrm{d}\Omega \\ + C_{l''} f_{lm} \int_\Omega Y^*_{l'm'} Y_{l''m''} Y_{lm} \, \mathrm{d}\Omega \Bigg). \end{aligned} \tag{3.34}$$

Using that $f^*_{lm} = (-1)^m f_{l-m}$ and $Y^*_{lm} = (-1)^m Y_{l-m}$ applied to the first term in the sum we obtain (since m is both positive and negative the two sums $\sum_{m=-l}^{l} f_{lm}$ and $\sum_{m=-l}^{l} f_{l-m}$ are equal)

$$C_{l'm',l''m''} = \delta_{l'l''}\delta_{m'm''} C_{l'} + \sum_{lm} f_{lm} \big[C_{l'} + C_{l''} \big] \int_\Omega Y_{lm} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega. \tag{3.35}$$

This expression for the anisotropy covariance should next be combined with the quadratic estimator, with the quadratic part $\tilde{\mathbf{h}}$ given by

$$\tilde{\mathbf{h}} = \frac{1}{2} \sum_{lm,l'm'} \left[\frac{\partial C_{lm,l'm'}}{\partial \mathbf{h}^\dagger} \right] \bar{\Theta}^*_{lm} \bar{\Theta}_{l'm'}, \tag{3.36}$$

where $\bar{\Theta}$ is the inverse-variance filtered sky signal. For the dipole, $\mathbf{h}$ is the vector with f_{lm} as components ($[f_{1-1}, f_{10}, f_{11}]$). Component-wise we thus get for $\tilde{\mathbf{h}}$:

$$\tilde{h}_{lm} = \frac{1}{2} \sum_{l'm',l''m''} \left[\frac{\partial C_{l'm',l''m''}}{\partial f^*_{lm}} \right] \bar{\Theta}^*_{l'm'} \bar{\Theta}_{l''m''}. \tag{3.37}$$

The derivative becomes (remember that C_l is the isotropic covariance and so does not depend on the anisotropy)

$$\frac{\partial C_{l'm',l''m''}}{\partial f^*_{lm}} = \sum_{l'''m'''} \frac{\partial \left(f_{l'''m'''}[C_{l'} + C_{l''}] \int Y_{l'''m'''} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega \right)}{\partial f^*_{lm}} \tag{3.38}$$

$$= [C_{l'} + C_{l''}] \sum_{l'''m'''} \frac{\partial f_{l'''m'''}}{\partial f^*_{lm}} \int Y_{l'''m'''} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega. \tag{3.39}$$

Again using that $f^*_{lm} = (-1)^m f_{l-m}$ we get

$$\begin{aligned}
\frac{\partial C_{l'm',l''m''}}{\partial f^*_{lm}} &= (-1)^m [C_{l'} + C_{l''}] \sum_{l'''m'''} \frac{\partial f_{l'''m'''}}{\partial f_{l-m}} \int Y_{l'''m'''} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega \\
&= (-1)^m [C_{l'} + C_{l''}] \sum_{l'''m'''} \delta_{ll'''} \delta_{-mm'''} \int Y_{l'''m'''} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega \\
&= (-1)^m [C_{l'} + C_{l''}] \int Y_{l-m} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega \\
&= [C_{l'} + C_{l''}] \int Y^*_{lm} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega.
\end{aligned}$$

Inserting the final expression into equation 3.37, we get

$$\tilde{h}_{lm} = \frac{1}{2} \sum_{l'm',l''m''} \left([C_{l'} + C_{l''}] \int_\Omega Y^*_{lm} Y^*_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega \right) \bar{\Theta}^*_{l'm'} \bar{\Theta}_{l''m''}. \tag{3.40}$$

Taking the integral outside the sum we obtain

$$\tilde{h}_{lm} = \frac{1}{2} \int_\Omega Y^*_{lm} \left(\sum_{l'm'} \sum_{l''m''} \left[C_{l'} + C_{l''} \right] \bar{\Theta}^*_{l'm'} Y^*_{l'm'} \bar{\Theta}_{l''m''} Y_{l''m''} \right) \mathrm{d}\Omega. \tag{3.41}$$

Using that $\bar{\Theta}^*_{l'm'} = (-1)^{m'} \bar{\Theta}_{l'-m'}$, $\quad Y^*_{l'm'} = (-1)^{m'} Y^*_{l'-m'}$ and noting that $\sum_{m'} \bar{\Theta}_{l'-m'} Y_{l'-m'} = \sum_{m'} \bar{\Theta}_{l'm'} Y_{l'm'}$, the equation can be written as

$$\tilde{h}_{lm} = \frac{1}{2} \int_\Omega Y^*_{lm} \left(\sum_{l'm'} \sum_{l''m''} \left[C_{l'} + C_{l''} \right] \bar{\Theta}_{l'm'} Y_{l'm'} \bar{\Theta}_{l''m''} Y_{l''m''} \right) \mathrm{d}\Omega. \tag{3.42}$$

We note that this differs from the expression found in Hanson & Lewis (equation 17); first and foremost they are missing the $C_{l'}$. The integral

$$\int_\Omega Y^*_{lm} Y_{l'm'} Y_{l''m''} \, \mathrm{d}\Omega \tag{3.43}$$

is a form of the *Gaunt integral*, and the solution is

$$(-1)^m \sqrt{\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi}} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix}, \tag{3.44}$$

where the latter two symbols are the so-called Wigner 3-j symbols[5]. Putting everything together we get for $\tilde{h}_{lm}$:

$$\begin{aligned}\tilde{h}_{lm} = (-1)^m \frac{1}{2} \sum_{l'm'l''m''} \sqrt{\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi}} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \\ \times \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix} \left[C_{l'} + C_{l''}\right] \bar{\Theta}_{l'm'} \bar{\Theta}_{l''m''}.\end{aligned} \tag{3.45}$$

This is seemingly different from the expression found in H&L, but if we play around a bit with the indices we will see that they are, in fact, equal. To see this, observe the following: Changing $\bar{\Theta}_{l''m''} \to (-1)^{m''} \bar{\Theta}^*_{l''-m''}$, we also get

$$(-1)^m (-1)^{m''} = (-1)^m (-1)^{m-m'} = (-1)^{m'},$$

using a selection rule of the second 3-j symbol. In the sum over m'' we now have

$$\begin{aligned}\sum_{m''=-l''}^{l''} \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix} \bar{\Theta}^*_{l''-m''} &= \sum_{m''=-l''}^{l''} \begin{pmatrix} l & l' & l'' \\ -m & m' & -m'' \end{pmatrix} \bar{\Theta}^*_{l''m''} \\ &= \sum_{m''=-l''}^{l''} \begin{pmatrix} l & l' & l'' \\ m & -m' & m'' \end{pmatrix} \bar{\Theta}^*_{l''m''},\end{aligned}$$

where the final equality holds because of symmetry relations[6]: Changing the signs of all the m's should introduce a phase factor $(-1)^{l+l'+l''}$, but because we also have the 3-j symbol $\left(\begin{smallmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{smallmatrix}\right)$ present, we know that $l + l' + l''$ must be even. All of this makes our equation 3.45 equal to the one found in H&L, which is given by

$$\begin{aligned}\tilde{h}_{lm} = \frac{1}{2} \sum_{l'm'l''m''} (-1)^{m'} \sqrt{\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi}} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \\ \times \begin{pmatrix} l & l' & l'' \\ m & -m' & m'' \end{pmatrix} \left[C_{l'} + C_{l''}\right] \bar{\Theta}_{l'm'} \bar{\Theta}^*_{l''m''}.\end{aligned} \tag{3.46}$$

This is the expression for the quadratic estimator we used in the computer code. Since we look at the dipole l must be 1; the sums over l' and l'' go up to some l_{mod}, which must be smaller than, or equal to, the entire range of multipoles l_{max} in the data set (usually chosen as $2 \cdot N_{\text{side}}$). The choice of l_{mod} corresponds to considering how many multipoles the dipole modulation extends to.

3.1.3 The Fisher matrix for the dipole case

The Fisher matrix can be thought of as the normalization of the estimated values $\tilde{h}_{lm}$. It gives the covariance of the values, and component-wise is given by

$$\mathcal{F}_{lm,l'm'} = \text{Cov}(\tilde{h}_{lm}, \tilde{h}_{l'm'}). \tag{3.47}$$

[5] See the appendix, section A.2

[6] See the appendix, equation A.7 combined with equation A.8.

Our estimator for $\hat{h}_{lm}$ becomes

$$\hat{h}_{lm} = \sum_{l'm'} \mathcal{F}^{-1}_{lm,l'm'} \big(\tilde{h}_{l'm'} - \langle \tilde{h}_{l'm'} \rangle\big). \tag{3.48}$$

Since we have three parameters $\tilde{h}_{1-1}, \tilde{h}_{10}$ and $\tilde{h}_{11}$ our covariance matrix becomes a 3×3 matrix. A covariance matrix for real variables is symmetric, but since we operate with complex variables the matrix $\mathcal{F}$ will instead be *hermitian*, i.e. $\mathcal{F} = \mathcal{F}^\dagger$. Still, it will only be necessary to calculate the elements on one half since the other is equal except for the conjugation.

The averaging $\langle \tilde{h}_{1m} \rangle$ and the computation of the Fisher matrix is done over isotropic realizations of the CMB for an ensemble of Monte Carlo simulations with the same noise model and beam as the data to be analyzed. Some more details on the estimation of these quantities is shown in section 5.4.

Hanson & Lewis show an analytical expression for $\mathcal{F}$ in the isotropic case, given by

$$[\mathcal{F}_{\rm iso}]_{lm,l'm'} = \delta_{ll'}\delta_{mm'} \sum_{l_1 l_2} \frac{(2l_1+1)(2l_2+1)}{8\pi} \begin{pmatrix} l & l_1 & l_2 \\ 0 & 0 & 0 \end{pmatrix}^2 \frac{(C_{l_1}+C_{l_2})^2}{C^{\rm tot}_{l_1} C^{\rm tot}_{l_2}}, \tag{3.49}$$

where $C^{\rm tot}_l = C_l + N_l$. For $C_l \to b_l^2 C_l$ this agrees well with simulations.

3.1.4 Calculating the dipole amplitude and direction from the harmonic modulation

From the estimator we get the dipole modulation in terms of a harmonic field $\hat{h}_{lm}$, with three values $\hat{h}_{1-1}, \hat{h}_{10}$ and $\hat{h}_{11}$. How do we calculate the amplitude and the direction from these parameters? The easiest way is perhaps to simply construct a map from the $\hat{h}_{lm}$'s using HEALPix, and then find the point where it is the strongest. This is somewhat time-consuming, so we rather want to calculate it directly from our estimated harmonic modulation coefficients. We have the following relations between a spherical and a cartesian coordinate system:

$$\begin{aligned} x &= \sin\theta\cos\phi \\ y &= \sin\theta\sin\phi \\ z &= \cos\theta. \end{aligned} \tag{3.50}$$

In the convention we work with, on the sky map θ goes from 0 on the north pole to π on the south pole, and ϕ goes from 0 on the center and increases towards the left, becoming 2π after one rotation.
The spherical harmonics are related to the cartesian coordinates through

$$\begin{aligned} Y_{10} &= \frac{1}{2}\sqrt{\frac{3}{\pi}} z \\ Y_{11} &= -\frac{1}{2}\sqrt{\frac{3}{2\pi}}(x+iy), \end{aligned} \tag{3.51}$$

when we assume a unit sphere. Our modulation field is decomposed in the usual way,

$$f(\theta,\phi) = \sum_{m=-1}^{1} h_{1m} Y_{1m}(\theta,\phi), \tag{3.52}$$

and writing out the sum, applying 3.51, we obtain

$$f(x,y,z) = -\sqrt{\frac{3}{2\pi}}\mathrm{Re}[\hat{h}_{11}]x + \sqrt{\frac{3}{2\pi}}\mathrm{Im}[\hat{h}_{11}]y + \frac{1}{2}\sqrt{\frac{3}{\pi}}\mathrm{Re}[\hat{h}_{10}]z. \tag{3.53}$$

From this we see that we have

$$\begin{aligned} x &= -\sqrt{\frac{3}{2\pi}}\mathrm{Re}[\hat{h}_{11}] \\ y &= \sqrt{\frac{3}{2\pi}}\mathrm{Im}[\hat{h}_{11}] \\ z &= \frac{1}{2}\sqrt{\frac{3}{\pi}}\mathrm{Re}[\hat{h}_{10}]. \end{aligned} \tag{3.54}$$

However, since we don't necessarily get these normalized after applying the estimator, they will need to be normalized (since the direction vector $\hat{\mathbf{p}}$ is a unit vector) through

$$\hat{x} = \frac{x_{\mathrm{old}}}{\sqrt{x_{\mathrm{old}}^2 + y_{\mathrm{old}}^2 + z_{\mathrm{old}}^2}} \tag{3.55}$$

and so on. The angles can thus be readily calculated using equation 3.50. The amplitude $\hat{A}$ of the dipole is found by taking the norm of the un-normalized vector,

$$\hat{A} = \sqrt{x_{\mathrm{old}}^2 + y_{\mathrm{old}}^2 + z_{\mathrm{old}}^2}. \tag{3.56}$$

3.2 Optimal maximum-likelihood (Eriksen, Hoftuft)

The exact methods of finding the parameters which maximizes the likelihood is based on Metropolis-Hastings Monte Carlo methods, which explores the whole parameter space and eventually settles around the true values. They involve explicitly calculating the full likelihood in pixel space, which entails the factorization of the full covariance matrix for each Monte Carlo cycle.

With the parameters denoted by $\hat{\lambda}$, the Bayesian analysis framework is to find the posterior probability distribution of the parameters $P(\hat{\lambda}|\mathbf{d})$, where $\mathbf{d}$ is the data, through Bayes' theorem

$$P(\hat{\lambda}|\mathbf{d}) = \frac{P(\mathbf{d}|\hat{\lambda})P(\hat{\lambda})}{P(\mathbf{d})}, \tag{3.57}$$

where $P(\hat{\lambda})$ is the prior distribution of the parameters (often chosen to be uniform, when assuming no prior knowledge of the parameters) and $P(\mathbf{d}|\hat{\lambda})$ is the likelihood

$\mathcal{L}(\hat{\lambda})$. The probability $P(\mathbf{d})$ is a normalization factor, which can be ignored when the focus is to simply maximize the likelihood. The posterior distribution is then mapped out by maximizing $\mathcal{L}$ using a Markov Chain Monte Carlo method. This is done using the Metropolis algorithm scheme: The next step in the random walk of parameter exploration is always accepted if the new, proposal, likelihood has a larger value than the previous one, and if it is smaller there is a certain probability that it may yet be accepted. After a burn-in period the parameters can be estimated from the Markov Chain, with the best-fit value being the mean and with variance corresponding to the confidence of the estimation.

These methods have traditionally been strongly limited by computing power. Hoftuft et al. (2009) [12] stated a time usage of ~ 50000 CPU hours at a resolution of $N_{\mathrm{side}} = 32$, and the $\mathcal{O}(N_{\mathrm{side}}^6)$ scaling thus makes it virtually impossible to go much higher with today's resources (even with a cluster of 10,000 CPU's available you would need two weeks to complete the analysis for $N_{\mathrm{side}} = 64$).

Chapter 4

Implementation

In the previous chapter we outlined the mathematical formulas for estimating the dipole modulation. Here we will present some of the more computational aspects of the estimation procedure. Even though the QML approach is a lightweight method, there are still concerns to be addressed, mainly with regards to how we solve the inversion of the covariance matrix. This will be accomplished using the conjugate gradients algorithm. We will also investigate the computational expense of our program.

4.1 The conjugate gradients algorithm

In the estimator we need the inverse-variance filtered signal, but in order to calculate it we would need the inverse of the full covariance. This would be nearly impossible to compute, as the matrix to be inverted would be huge. Instead, we use the *conjugate gradients* algorithm, for which you can find a great introduction in Shewchuk (1994) [27], on which this section is based. It is an algorithm for solving a linear matrix-vector equation $Ax = b$ iteratively, and we will briefly outline the concept below. What we wish to solve and obtain is the inverse-variance weighted signal in equation 3.46. It is given by

$$\bar{\Theta} = (C^{\text{full}})^{-1}\Theta, \tag{4.1}$$

where C^{full} is the full covariance of the observed data. We can write this as

$$\begin{aligned} & C^{\text{full}}\bar{\Theta} = \Theta \\ \Rightarrow & [C + N]\bar{\Theta} = \Theta \\ \Rightarrow & [N^{-1}C + \mathrm{I}]\bar{\Theta} = N^{-1}\Theta \\ \Rightarrow & [C^{-1} + N^{-1}]C\bar{\Theta} = N^{-1}\Theta, \end{aligned}$$

ignoring for the moment beams and mappings between harmonic and pixel space. C is here the isotropic covariance, diagonal and equal to C_l in harmonic space, and N is the noise (including both instrumental noise, a sky-cut and removal of the monopole/dipole

modes[1]), diagonal in pixel space. The CG method requires that the matrix A on the left-hand side of $Ax = b$ is symmetric and positive-definite. That's not necessarily the case the way the equation is written here, with $\bar{\Theta}$ as our x, so we need to modify it. Only letting the expression in brackets be our A would solve the problem, but that expression might not be numerically well-conditioned for the search. We modify the equation through

$$\begin{aligned}&\Rightarrow[C^{-1/2} + N^{-1}C^{1/2}]C^{1/2}\bar{\Theta} = N^{-1}\Theta\\&\Rightarrow[\mathrm{I} + C^{1/2}N^{-1}C^{1/2}]C^{1/2}\bar{\Theta} = C^{1/2}N^{-1}\Theta.\end{aligned}$$

The A in our equation is here the expression inside brackets, and our $x = C^{1/2}\bar{\Theta}$. The beams and conversion between harmonic and pixel space is incorporated as

$$[\mathrm{I} + C^{1/2}b_l Y^\dagger N^{-1} Y b_l C^{1/2}]C^{1/2}\bar{\Theta} = C^{1/2}b_l Y^\dagger N^{-1}\Theta, \tag{4.2}$$

where b_l is the beam function in harmonic space and Y denotes mapping from harmonic to pixel space, with $Y^\dagger$ being the opposite operation. These operations are done using the HEALPix functions `alm2map` and `map2alm`, respectively. We thus solve for $\bar{\Theta}$ in harmonic space, which is what we need in the estimator as well. Of note is that the equation is solved for *real-valued spherical harmonics*[2], before we convert $\bar{\Theta}$ to the complex domain for inclusion in the estimator. The vectors are stored so that each l takes up $2l + 1$ components, with the first of those components corresponding to harmonic indices $l, m{=}{-}l$, the next $l, m{=}{-}l{+}1$, and so on.

The basic concept of the algorithm is to minimize the quadratic form

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} + c, \tag{4.3}$$

whose derivative is

$$f'(\mathbf{x}) = \frac{1}{2}A^T\mathbf{x} + \frac{1}{2}A\mathbf{x} - \mathbf{b}. \tag{4.4}$$

For symmetric A, $f'(\mathbf{x}) = 0$ yields the solution $A\mathbf{x} = \mathbf{b}$. In order for the solution to be a minimum of $f(\mathbf{x})$, and not just a critical point, the matrix A needs to be positive-definite. The CG algorithm is an improvement over the Method of Steepest Descent, which seeks to find the minimum by always going in the opposite direction of the gradient vector of f (which points toward where the function increases most rapidly). The step length α is chosen so that we reach the point which minimizes f along the line of the gradient. Depending on where we start and how the function looks, the number of steps needed to converge could vary enormously. The Conjugate Gradients method improves upon this by ensuring, through various techniques (mainly by Gram-Schmidt conjugation and concepts such as A-orthogonality), that each new search direction is orthogonal to all the previous ones, so that we never go in the same direction twice. This means that if the dimension of the problem at hand is n, the

[1] See section 6.1

[2] See the appendix, section A.3.1

search will be complete after at most n iterations (in practice there are roundoff errors which complicates matters). However we do need a criterion for when to stop the search and be satisfied, since the n in our case is $(l_{\max}+1)^2$, which can be more than 1 million and therefore the search would take a vast amount of time. How quickly the search converges depends in large part on the *condition number* κ of the matrix A, which is the ratio of its largest to smallest eigenvalue. The closer this number is to 1, the better (in the 2-dimensional case a small κ makes the quadratic form appear like a spherical pit with a well-defined minimum point, a large κ could instead make the function resemble a long trough or valley). This is also why *preconditioning* the system is so important when the dimensions are large: The technique is to multiply the equation by another matrix M^{-1} on each side, so that we get

$$M^{-1}Ax = M^{-1}b. \tag{4.5}$$

The preconditioner matrix M should resemble A, so that the product $M^{-1}A$ becomes close to the identity matrix (which has a condition number of precisely 1). A perfect preconditioner would equal A, but then again, this wouldn't help us at all since our problem to solve would remain the same and we would need to invert A, which is what we seek to avoid. So M should be chosen so that it is easily invertible but still resemble A, by and large.

Note that the dimensionality of the problem, $n = (l_{\max}+1)^2$, makes it impossible to store the full matrices A and M explicitly. However, the act of multiplying with these matrices are preserved, by storing the components making up the matrix or exploiting its sparsity.

4.1.1 The algorithm

The following is the (non-preconditioned) algorithm for solving $Ax = b$ iteratively using the CG approach, as outlined in Shewchuk:

Initial conditions:

$$\begin{aligned} r_0 &= b - Ax_0 \\ d_0 &= r_0 \end{aligned}$$

Iterative scheme:

$$\begin{aligned} \alpha_i &= \frac{r_i^T r_i}{d_i^T A d_i} \\ x_{i+1} &= x_i + \alpha_i d_i \\ r_{i+1} &= r_i - \alpha_i A d_i \\ \beta_{i+1} &= \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} \\ d_{i+1} &= r_{i+1} + \beta_{i+1} d_i. \end{aligned} \tag{4.6}$$

The solution vector is x_{i+1}, and the iterations are run until the norm of the residual vector r_{i+1} (which tells how far we are from the true solution) is smaller than some

convergence criterion,

$$\sqrt{\frac{r_{i+1}^T r_{i+1}}{r_0^T r_0}} \leq \epsilon, \tag{4.7}$$

where ϵ is chosen as a small number (in this work we set $\epsilon = 1.0 \cdot 10^{-6}$). The vector d_i gives the direction for the next step, and β_i is a constant which makes sure it is orthogonal to the previous directions. The preconditioned scheme is

$$\begin{aligned}
&\text{Initial conditions:} \\
&r_0 = b - Ax_0 \\
&d_0 = M^{-1} r_0 \\
&\text{Iterative scheme:} \\
&\alpha_i = \frac{r_i^T M^{-1} r_i}{d_i^T A d_i} \\
&x_{i+1} = x_i + \alpha_i d_i \\
&r_{i+1} = r_i - \alpha_i A d_i \\
&\beta_{i+1} = \frac{r_{i+1}^T M^{-1} r_{i+1}}{r_i^T M^{-1} r_i} \\
&d_{i+1} = M^{-1} r_{i+1} + \beta_{i+1} d_i.
\end{aligned} \tag{4.8}$$

4.1.2 Preconditioner

We will utilize the *block preconditioner* of Eriksen et al. (2004) [28] in this work. It is not the fastest method out there, there is for instance the multigrid preconditioner of Smith, Zahn and Dore (2007) [29]. The block preconditioner is an improvement over the simple *diagonal* one. There, we approximate $M \sim A$ as merely the diagonal of A. This works quite well when A has very low values on the off-diagonal, but in other cases we need improvements. The block preconditioner does this by approximating M as a dense block equal to A, M_{block}, for a certain number of l's, and as a diagonal, M_{diag}, for the rest. If the block is too large, the time needed to compute, invert and apply it would outweigh the gain in number of iterations needed to reach our goal, so it is important to find a good balance here. In addition a large block demands a lot of computer memory. Our preconditioner matrix M would schematically look like

$$M = \begin{pmatrix} [M_{\text{block}}] & 0 \\ 0 & M_{\text{diag}} \end{pmatrix}. \tag{4.9}$$

The size of l_{block} is chosen between 50 and 70 in this work, with the time usage shown in the coming section. The inversion of M is done by separately inverting the block and the diagonal, and then putting them together again. The program uses either Cholesky or LU factorization in the inversion of the block, and the size is not large enough so that it makes much of a difference which one is used. The expression for the preconditioner

l_{mod}	Time (s)
40	0.001
64	0.004
100	0.015
128	0.027
256	0.18
512	1.3
1024	10.1

Table 4.1: The CPU time for the estimator on different l_{mod}.

in harmonic space is

$$M_{lm,l'm'} = \mathrm{I}_{lm,l'm'} + \sqrt{C_l}(N^{-1}_{lm,l'm'})^R\sqrt{C_l}b_l^2, \tag{4.10}$$

where the expression for $N^{-1}_{lm,l'm'}$ is shown below in equation 4.11 and the ^{R} denotes conversion to *real spherical harmonics*. The $\mathrm{I}_{lm,l'm'} = \delta_{ll'}\delta_{mm'}$ is the identity matrix. Note that this preconditioner does resemble the A in our equation, as is our wish.

4.2 Computational expense

4.2.1 The estimator

The computational time of the estimator in equation 3.46 varies mainly with the size of l_{mod} used. I have listed the CPU times[3] in table 4.1, for estimation of a single $\tilde{h}_{1m}$ (the Fisher matrix of equation 3.49 was also calculated at the same time, to negligible additional cost). The way the program is written the estimator should scale roughly as $\mathcal{O}(l^3_{\mathrm{mod}})$, and this is reflected in the recorded times. The most important factor in the calculation when it comes to computational expense is the calculation of the second Wigner 3-j symbol. Regardless, the time usage is really rather insubstantial, with only 10 seconds on our highest l_{mod}. Compared to the time it takes to run the entire program, this is really a mere drop in the sea.

4.2.2 The CG search

The time usage of the CG search varies strongly with the input data and with the size of the block preconditioner. The construction of the dense $N^{-1}_{lm,l'm'}$ part of the block scales roughly as $\mathcal{O}(l^4_{\mathrm{block}})$, so a small increase in l_{block} can cause a huge boost in computational time. The size of the block also has the same scaling, so it eats up memory rather quickly. The expression for the noise covariance in harmonic space is

[3]The program was run on an Intel® Core™ 2 Quad Q9400 2.66 GHz processor, clocked at 2 GHz and utilizing only 1 core. The language was Fortran 95, compiled using the Intel® Fortran Compiler.

l_{block}	Time (s)
10	0.05
20	0.7
30	3.6
40	12
50	30
60	65
70	127
80	225

Table 4.2: The CPU times for the creation of the dense harmonic N^{-1} block on different sizes l_{block}.

given by, as shown in Hivon et al. (2002) [30] and Eriksen et al. (2004) [28],

$$N^{-1}_{lm,l'm'} \approx \frac{N_{\text{pix}}}{4\pi}(-1)^m \sum_{l''m''} \eta_{l''m''} \sqrt{\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi}} \times \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix}, \tag{4.11}$$

where the sum over l'' goes up to $l_{\max}$ and $\eta_{l''m''}$ are the "a_{lm}'s" of the inverse noise model map. The CPU times for the calculations of this entire block up to different sizes l_{block} are shown in table 4.2, where the $\mathcal{O}(l^4)$ scaling is apparent. In table 4.3 is shown the time and number of iterations it takes to complete the CG search for different l_{block}. Note that these figures are for a lower convergence criterion than what is used when obtaining the analysis data, so the differences shown are smaller than what they would be for a more strict criterion (for comparison, running the search with a convergence criterion of $\epsilon = 10^{-6}$ in the case of $l_{\text{block}} = 80$ takes 202 iterations and 31min 20s to complete). A plot of the convergence as a function of CG step is shown in figure 4.1. The CPU cost to invert the block part of the preconditioner is shown in table 4.4, for both LU and Cholesky decomposition (where LAPACK has been used to perform the decompositions). Both methods scale as $\mathcal{O}(l^6_{\text{block}})$, with Cholesky being about a factor 2 faster than LU. Both these properties are reflected in the times shown.

Calculating the diagonal part, $N^{-1}_{lm,lm}$, of the harmonic-space inverse noise covariance for all l up to $l_{\max}$ also takes some time, though it only scales as roughly $\mathcal{O}(l^3_{\max})$ and therefore is not very important in the bigger picture. These times are shown in table 4.5.

The most important functions when it comes to the time usage of a single CG step are those for conversion between pixel space and harmonic space; `map2alm` and `alm2map` in HEALPix. With each step there are two of these, occurring when multiplying with N^{-1}, and they take roughly 8.5s combined for $N_{\text{side}} = 512$, $l_{\max} = 1024$. Both functions are found to scale roughly as $l^3_{\max}$.

l_{block}	Time (min)	Iterations
No block	16min 17s	105
50	8min 6s	51
70	7min 36s	46
80	6min 31s	41

Table 4.3: Number of iterations and CPU time to perform the full CG search on different l_{block}. The data used is the foreground-reduced 7-year W band data (See section 6.2) from WMAP on $N_{\text{side}} = 512$ masked with the KQ85 mask, and the convergence criterion is $\epsilon = 1.0 \cdot 10^{-4}$.

l_{block}	Time LU	Time Cholesky
No block	0s	0s
50	0min 20s	0min 10s
70	2min 36s	1min 20s
80	6min 0s	2min 53s

Table 4.4: CPU time spent to invert the block part of the preconditioner using either LU or Cholesky decomposition.

Note also that when we are running the program on an ensemble of simulations, it is enough to compute the preconditioner only *once*, since it will be equal for all simulations. In practice it is precomputed beforehand.

l_{max}	Time (s)
128	0.1
256	0.7
512	4.8
1024	35.7

Table 4.5: The CPU times for the construction of the diagonal harmonic $N^{-1}_{lm,lm}$ on different sizes l_{max}.

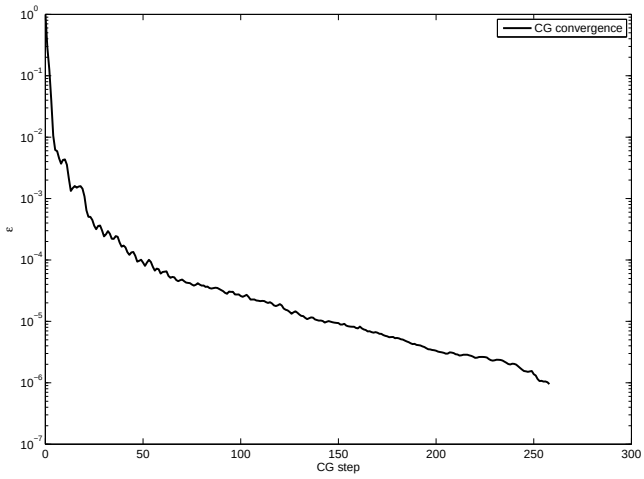


Figure 4.1: Plot of the CG convergence $\sqrt{\frac{r_i^T r_i}{r_0^T r_0}}$ as a function of step i, for the 7-year foreground-reduced V-band on $N_{\mathrm{side}} = 512$ with the KQ85 mask, using an l_{block} of 70.

Chapter 5

Analysis of simulations

In order to trust our code and algorithms it is pivotal that we test them on realistic simulated maps, and see if the analysis turns out as expected. In addition, simulations will need to be performed in order to calculate the mean estimator noise and to construct the Fisher normalization matrix, to be used with further analysis.

5.1 Creating simulated maps

Recall that the CMB temperature field map is decomposed into spherical harmonics through (equation 2.1)

$$\Delta T(\theta,\phi) = \Theta(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}Y_{lm}(\theta,\phi).$$

In an isotropic model, we assume our a_{lm}'s to be drawn from a Gaussian distribution with variance C_l. Since the modulated map is constructed from an inherent isotropic signal the first part of the simulation procedure is the same. Our simulated CMB map will be created from a_{lm}'s calculated using a Gaussian random number generator, and we use the WMAP best-fit for the standard ΛCDM model as our C_l's. We only need to calculate a_{lm}'s for positive m, since $a_{l-m} = (-1)^m a^*_{lm}$. We draw two real numbers x and y and construct the complex a_{lm}:

$$\begin{aligned}
&\text{if } m > 0 \text{ then}\\
&\quad x = \text{gauss}(0,1), \quad y = \text{gauss}(0,1)\\
&\quad a_{lm} = \sqrt{\frac{C_l}{2}}(x+iy)\\
&\text{else if } m = 0 \text{ then}\\
&\quad x = \text{gauss}(0,1)\\
&\quad a_{lm} = \sqrt{C_l}x\\
&\text{end if,}
\end{aligned} \tag{5.1}$$

where gauss$(0,1)$ describes a random normal Gaussian number with mean 0 and variance 1. We used the HEALPix function `rand_gauss` as the random number generator. Having these a_{lm}, we can construct the map with HEALPix using `alm2map`. Our range for l is $l_{\max} = 2 \cdot N_{\text{side}}$ throughout the thesis, N_{side} being the resolution parameter for the map. A dipole modulation is created simply by choosing an amplitude A and direction $\hat{\mathbf{p}}$, and then modifying the isotropic map Θ^{iso} in pixel space through

$$\Theta(\hat{\mathbf{n}}) = (1 + A\hat{\mathbf{p}} \cdot \hat{\mathbf{n}})\Theta^{\text{iso}}(\hat{\mathbf{n}}). \tag{5.2}$$

Next we need to include beam smoothing and pixel window, which is done by moving back into harmonic space (using the HEALPix function `map2alm`) and multiplying,

$$a_{lm} = a_{lm} \cdot b_l p_l, \tag{5.3}$$

with b_l being the instrumental beam (see equation 2.6), and the pixel window p_l is provided by the function `pixel_window` for different resolutions N_{side}. Finally, we need to add random Gaussian pixel noise of a certain strength,

$$\Theta(\hat{\mathbf{n}}) = \Theta(\hat{\mathbf{n}}) + \sigma_n(\hat{\mathbf{n}}) \cdot \text{gauss}(0,1), \tag{5.4}$$

where σ_n is the noise rms (100 μK or less in our usage), usually homogeneous for simulations unless we use the exact noise model from the data maps. Example simulations (without noise) were shown in figure 2.2 on page 21. All our simulations use an instrumental beam of 0.35 degrees, equivalent to the WMAP V-band, unless noted otherwise.

5.2 Verification of unbiasedness

We construct a number of simulations and create histograms of the distribution in analyzed values of the dipole parameters A, θ and ϕ. Ideally the difference between estimated and true values, $\hat{A} - A_{\text{true}}$, should have mean 0 and be normally distributed about the mean with standard deviation resembling the uncertainty of the estimator and the data. We will also look at how the estimator behaves with the number of modes we use in the estimation. Unless noted otherwise, we use $l_{\text{mod}} = l_{\max} = 2 \cdot N_{\text{side}}$.

In the first two sections, simulating full-sky maps, we have used the analytical Fisher matrix of equation 3.49 and not subtracted the estimator noise $\langle \tilde{h}_{lm} \rangle$. The analytical expression is found to match well those obtained from simulations, and the estimator noise should be negligible for purely isotropic full-sky maps.

We then construct simulations with a sky-cut present, where the Fisher matrix and the estimator noise are created from the $\tilde{h}_{lm}$ obtained from the estimator.

Parameter	Estimated value	Relative error	Standard deviation
$\hat{A}$	0.100005	$5.0 \cdot 10^{-5}$	$3.25 \cdot 10^{-3}$
$\hat{\theta}$	1.90148	$7.8 \cdot 10^{-4}$	$3.19 \cdot 10^{-2}$
$\hat{\phi}$	4.05126	$3.1 \cdot 10^{-4}$	$3.33 \cdot 10^{-2}$

Table 5.1: The mean estimated values of A, θ and ϕ for 1000 simulations and $N_{\text{side}} = 512$. The l_{mod} in the estimator is $2 \cdot N_{\text{side}} = 1024$. Input is $(A, \theta, \phi) = (0.1, 1.9, 4.05)$.

5.2.1 Simulations with a dipole modulation

Here we have constructed simulations with a dipole of amplitude 0.1, and direction[1] $(\theta, \phi) = (1.90, 4.05) \simeq (109°, 232°)$ (which is to the south-east on the projected sky-maps), having $N_{\text{side}} = 512$ and an instrumental beam of 0.35 degrees. Table 5.1 shows the mean estimated values for 1000 simulations using the isotropic Fisher normalization. The noise model used is the average noise rms of the W band[2], equal to 67.2 μK per pixel. We see that the input values are very well estimated indeed. The standard deviation was calculated from $\text{Var}(x) = \frac{1}{n}\sum_i (x_i - \mu_x)^2$. Calculating the 68% confidence intervals directly from the simulations by taking the 16% and 84% highest values estimated the results are

$$\begin{aligned} A &= (0.09686, 0.10316) \\ \theta &= (1.86853, 1.93384) \\ \phi &= (4.01990, 4.08151). \end{aligned} \tag{5.5}$$

Histograms from the simulations are shown in figure 5.1. These simulations indicate that the estimator is unbiased as it should be. In figure 5.5 histograms for 10000 similar simulations on $N_{\text{side}} = 32$ are shown, and again we see that the estimator seems to be unbiased. The 68% confidence intervals match fairly well the standard deviation, indicating that the estimator yields a Gaussian distribution of estimates.

5.2.2 Simulations without a modulation

Without a modulation present, the estimator still yields a non-zero value for the amplitude A. This is to be expected, since the amplitude by design is chosen to be non-negative; we are in effect looking at the absolute value. The distribution of the estimated amplitudes in this case resembles more the Rayleigh distribution[3]. We ran 1000 simulations with $N_{\text{side}} = 512$ and again using the average noise rms of the W band, and employing the analytical Fisher normalization. The mean estimated value of A was $\hat{A} = 5.0322 \cdot 10^{-3}$, and the 68% confidence interval (calculated as for the dipole

[1] The directions are given in the HEALPix convention, where θ is the co-latitude (0° at the north pole of the map, 180° at the south), and ϕ the longitude (being 0° at the center and moving leftwards on the map, reaching the center again at 360° after one rotation).

[2] See section 6.2.1

[3] The Rayleigh distribution goes as the normal distribution multiplied by x, $f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/\sigma^2}$, for non-negative x. It thus rises rapidly from 0 (depending on σ), and exhibits a drawn-out tail.

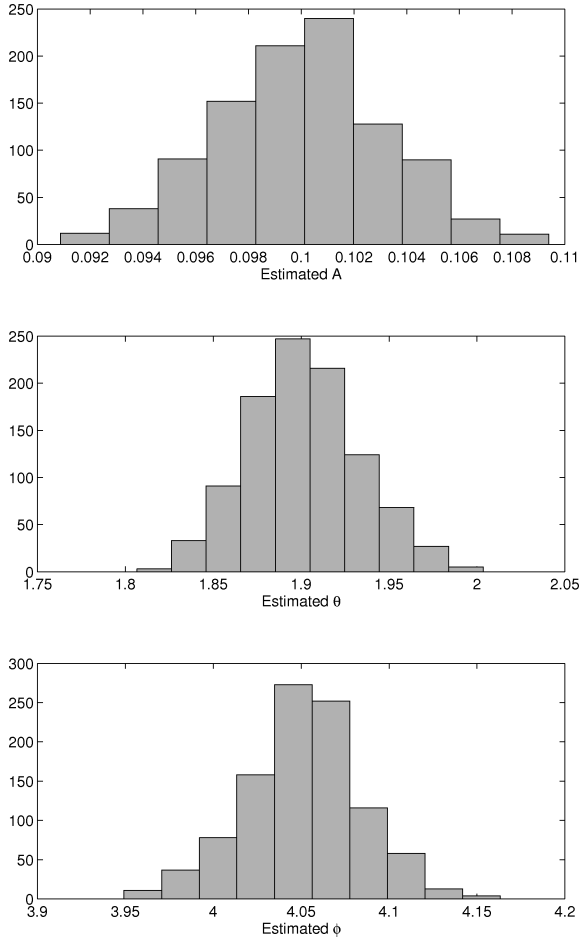


Figure 5.1: Histograms of 1000 simulations on $N_{\text{side}} = 512$ with input modulation $(A, \theta, \phi) = (0.1, 1.9, 4.05)$.

l_{mod}	$\langle\hat{A}\rangle$	68%	95%
10	0.186	0.22	0.34
40	0.0484	0.056	0.085
64	0.0304	0.036	0.053
100	0.0179	0.023	0.035
128	0.0153	0.018	0.026
256	$7.58 \cdot 10^{-3}$	$8.8 \cdot 10^{-3}$	0.014
512	$5.29 \cdot 10^{-3}$	$6.2 \cdot 10^{-3}$	$9.3 \cdot 10^{-3}$
1024	$5.03 \cdot 10^{-3}$	$6.0 \cdot 10^{-3}$	$9.0 \cdot 10^{-3}$

Table 5.2: Expectation value of $\hat{A}$ for isotropic simulations on different l_{mod}, using average W-band noise rms. Also shown are the 68% and 95% highest $\hat{A}$ measured.

simulations above)

$$A = (2.766 \cdot 10^{-3}, 7.210 \cdot 10^{-3}). \tag{5.6}$$

The standard deviation was $2.181 \cdot 10^{-3}$. A histogram of the estimated $\hat{A}$ is shown in figure 5.2, and we can see, as expected, that there is somewhat of a bias in the reported amplitude for no modulation. The estimated direction, however, is more or less random in the case of no modulation. To show this I have repeated the estimation on a much smaller N_{side} of 32 and done 10000 simulations with pixel noise a hundredth of the full case (to get similar signal-to-noise, since the harmonic-space noise model is $N_l = \sigma_0^2 \cdot 4\pi/N_{\mathrm{pix}}$, where σ_0 is the pixel noise rms. This gives a scaling by N_{pix}.). A map of the distribution of estimated dipole directions is shown in figure 5.3, and it pretty much resembles random noise. Compare this to the case when we *do* have an input modulation, shown in figure 5.4. Note that the variance of the estimates on this resolution is a lot higher than in the $N_{\mathrm{side}} = 512$ case (as evidenced from the histograms of figure 5.5 vs figure 5.1), so a similar distribution for that size would be even more focused around the true values than here. What seems to be clear is that if we estimate an amplitude substantially larger than 0.01 on an $N_{\mathrm{side}} = 512$ map using all available l in the estimation, this should indicate evidence of a true dipole modulation being present, and we should also have some trust in the estimated direction. For lower l_{mod}, the uncertainty is higher, so a higher value of $\hat{A}$ is expected for isotropic signals. Running 10000 simulations on the $N_{\mathrm{side}} = 32$ map with no modulation, we get a mean $\hat{A} = 3.04 \cdot 10^{-2}$, which is quite a bit higher than in the $l_{\mathrm{mod}} = 1024$ case (using an $N_{\mathrm{side}} = 32$ instead of an $N_{\mathrm{side}} = 512$ map gives similar results for the same l_{mod} as long as the noise is similar, evidenced by comparing with 500 simulations of the full case, see figure 5.6). Table 5.2 shows the expectation value for $\hat{A}$ for isotropic simulations on different l_{mod}.

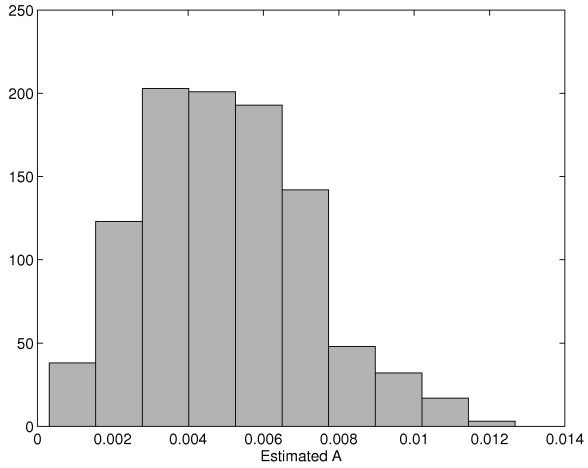


Figure 5.2: Histogram of 1000 simulations on $N_{\text{side}} = 512$ with input $A = 0$.

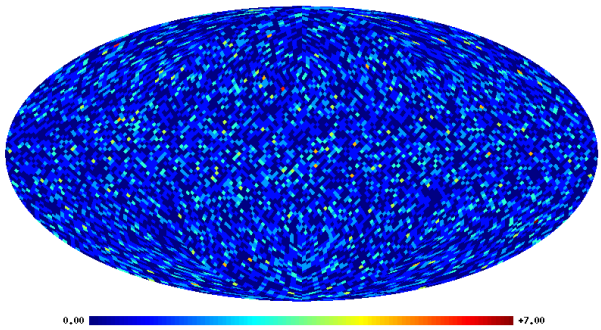


Figure 5.3: Map of estimated dipole directions for $N_{\text{side}} = 32$ and input $A = 0$.

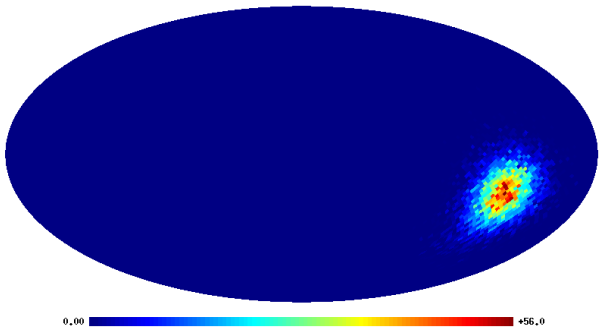


Figure 5.4: Map of estimated dipole directions for $N_{\text{side}} = 32$ and input $(A, \theta, \phi) = (0.1, 1.9, 4.05)$.

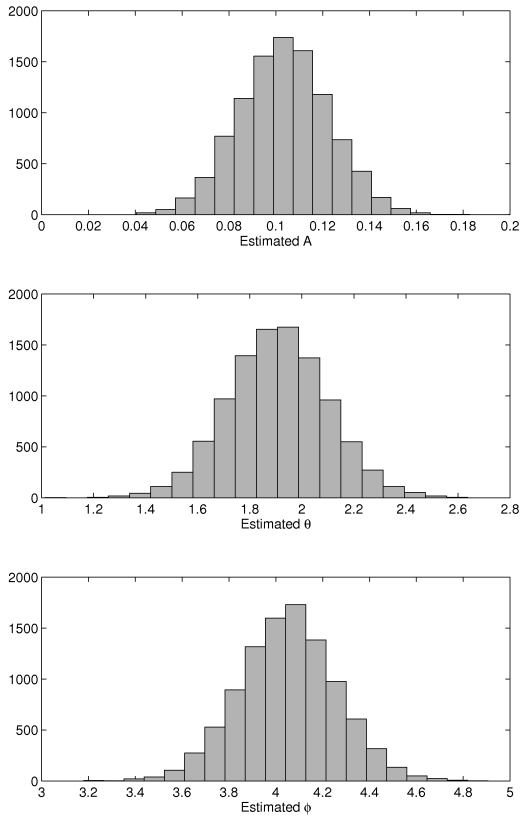


Figure 5.5: Histograms of 10000 simulations on $N_{\text{side}} = 32$ with $(A, \theta, \phi) = (0.1, 1.9, 4.05)$.

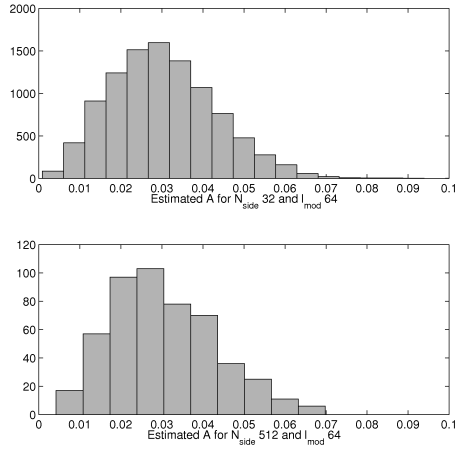


Figure 5.6: Histogram of 10000 simulations on $N_{\rm side} = 32$ (upper) and 500 simulations on $N_{\rm side} = 512$ (lower) with input $A = 0$.

5.2.3 Simulations with a sky-cut

Here we perform simulations using the same sky-cut as we will employ for analyzing the real data, the KQ85 mask (figure 2.4). Earlier we only looked at full-sky simulations, but since the data will need to be analyzed with a mask we should take this into consideration when comparing with simulations. These simulations, for no dipole modulation, will in turn be used to construct the Fisher matrix and estimator noise to use in the analysis. For 1280 isotropic simulations using complete 7-year W-band rms noise and a beam of 0.22 degrees, the mean estimated values of $\hat{A}$ as well as the 68% confidence intervals (calculated from the 16% and 84% highest estimated values) are shown in table 5.3. Similar results were obtained on the V- and Q-bands. Histograms of the W-band simulations for a variety of $l_{\rm mod}$ are shown in figure 5.7.

For simulations with a modulation and a sky-cut results are shown in table 5.4 with histograms in figure 5.8. Here, we have constructed 1000 simulations using V-band noise and beam, and employing an input modulation of 0.05 amplitude. The estimator is largely unbiased also with the sky-cut present, though the estimates are shifted to slightly larger values for higher l. These higher l are also where the noise becomes dominant over the signal. The significances of the estimates compared with the pure isotropic case increase with l however, also when looking at the low end of the confidence interval. Also, as expected, the confidence increases with l.

Similarly to the simulations with no sky-cut, I have created maps of the estimated dipole directions for modulated and unmodulated signals. For clarity the directions have been projected onto a low-resolution $N_{\rm side} = 32$ map (since the sample size of 1000 simulations is not very large compared to the number of pixels at full resolution). These maps are shown in figure 5.9.

l_{mod}	$\langle \hat{A} \rangle$	68% confidence
10	0.197	(0.115, 0.277)
25	0.0858	(0.0488, 0.1205)
40	0.0547	(0.0310, 0.0784)
64	0.0343	(0.0197, 0.0481)
100	0.0216	(0.0124, 0.0308)
200	0.0112	(0.0067, 0.0161)
500	0.0054	(0.0031, 0.0076)
1000	0.0046	(0.0026, 0.0067)

Table 5.3: Expectation value of $\hat{A}$ for isotropic simulations with the KQ85 sky-cut on different l_{mod}, using W-band noise rms. Also shown is the 68% confidence intervals.

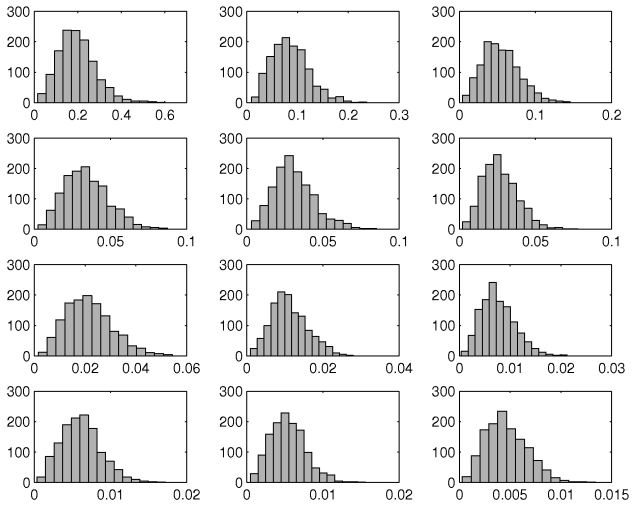


Figure 5.7: Histograms of 1280 isotropic simulations equivalent to the 7-year W-band. The l_{mod} are, from left to right starting at the top: 10, 25, 40, 64, 70, 80, 100, 200, 300, 400, 500 and 1000.

$l_{\rm mod}$	$\langle\hat{A}\rangle$	68% confidence	$\langle\hat{\theta}\rangle$	Confidence	$\langle\hat{\phi}\rangle$	Confidence
10	0.210	(0.122, 0.294)	1.69	(0.88, 2.45)	3.46	(1.62, 5.23)
25	0.100	(0.0562, 0.1432)	1.74	(1.16, 2.31)	3.77	(2.74, 4.89)
40	0.0778	(0.0470, 0.1075)	1.82	(1.34, 2.28)	3.98	(3.35, 4.69)
64	0.0656	(0.0437, 0.0878)	1.91	(1.60, 2.22)	4.08	(3.66, 4.51)
100	0.0566	(0.0424, 0.0704)	1.91	(1.71, 2.11)	4.07	(3.78, 4.36)
200	0.0495	(0.0427, 0.0564)	1.89	(1.78, 2.01)	4.10	(3.95, 4.25)
500	0.0587	(0.0546, 0.0631)	1.83	(1.77, 1.89)	4.11	(4.07, 4.15)
1000	0.0557	(0.0515, 0.0597)	1.88	(1.81, 1.93)	3.77	(3.74, 3.79)

Table 5.4: Expectation value of $\hat{A}$, $\hat{\theta}$ and $\hat{\phi}$ for modulated simulations with the KQ85 sky-cut on different $l_{\rm mod}$, using V-band noise rms. Also shown is the 68% confidence intervals. The input modulation was $A = 0.05$, $\theta = 1.9$, $\phi = 4.05$.

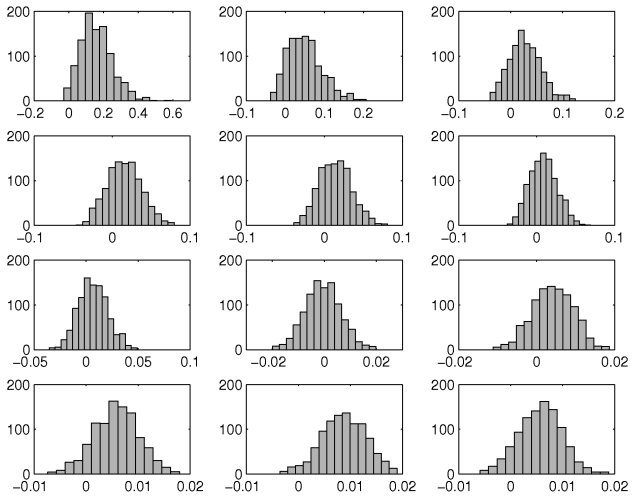
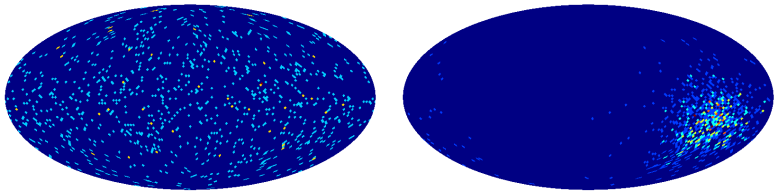
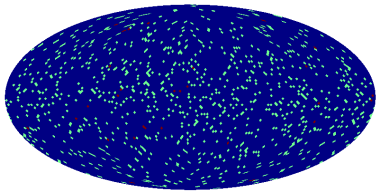


Figure 5.8: Histograms of 1000 modulated simulations with 7-year V-band noise. The $l_{\rm mod}$ are, from left to right starting at the top: 10, 25, 40, 64, 70, 80, 100, 200, 300, 400, 500 and 1000. The input modulation is $A_{\rm in} = 0.05$, and the histograms are of $\hat{A} - A_{\rm in}$.

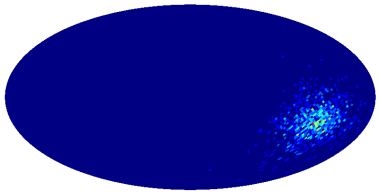


(a) $l_{\mathrm{mod}} = 64$, no modulation.

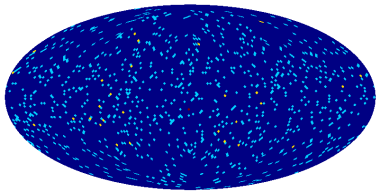
(b) $l_{\mathrm{mod}} = 64$, $A = 0.05$.



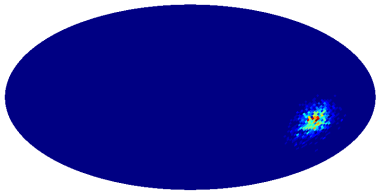
(c) $l_{\mathrm{mod}} = 100$, no modulation.



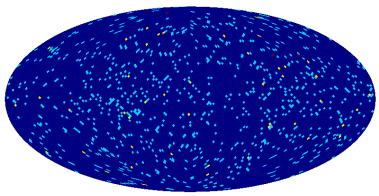
(d) $l_{\mathrm{mod}} = 100$, $A = 0.05$.



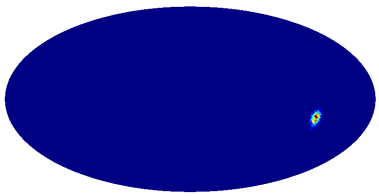
(e) $l_{\mathrm{mod}} = 200$, no modulation.



(f) $l_{\mathrm{mod}} = 200$, $A = 0.05$.



(g) $l_{\mathrm{mod}} = 500$, no modulation.



(h) $l_{\mathrm{mod}} = 500$, $A = 0.05$.

Figure 5.9: Maps of the estimated dipole directions for simulations using V-band noise, at different l_{mod}.

5.3 Estimating confidence and statistical significance

For isotropic simulations, we will look at the 68% and 95% levels of estimated values of A, to later compare with what we find in the real data. The significances of possible detections of dipolar assymetries in the data will be calculated in the pure frequentist manner of calculating the probability of a simulation to yield the same amplitude, simply by adding up the number of simulations giving higher or equal amplitude and dividing by the total number of simulations. This method requires a large sample of simulations in order to obtain reliable results, and for the most part we have constructed 1000 simulations corresponding to each dataset.

5.4 Determining the full Fisher matrix

We determine the full Fisher matrix for use with analyzing data by calculating it from isotropic simulations with a sky-cut and noise similar to the data and with the appropriate l_{mod} in the estimator. The Fisher matrix is given as the covariance matrix of the variables. The covariance of two complex random variables x and y is given by

$$\begin{aligned} \mathrm{Cov}(x,y) &= E\left[(x-\mu_x)(y-\mu_y)^*\right] \\ &= E\left[xy^*\right] - \mu_x\mu_y^*, \end{aligned} \tag{5.7}$$

where $\mu_x = E[x]$. We estimate the covariances through

$$\mathrm{Cov}(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)^*, \tag{5.8}$$

where the mean has been estimated by $\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i$. Our x's and y's are the harmonic modulation parameters $\tilde{h}_{1m}$, ordered in vector form as

$$\tilde{\mathbf{h}} = \begin{bmatrix} \tilde{h}_{1-1} \\ \tilde{h}_{10} \\ \tilde{h}_{11} \end{bmatrix}.$$

Our full Fisher matrix then becomes

$$\mathcal{F} = \begin{pmatrix} \mathrm{Cov}(\tilde{h}_{1-1},\tilde{h}_{1-1}) & \mathrm{Cov}(\tilde{h}_{1-1},\tilde{h}_{10}) & \mathrm{Cov}(\tilde{h}_{1-1},\tilde{h}_{11}) \\ \mathrm{Cov}(\tilde{h}_{10},\tilde{h}_{1-1}) & \mathrm{Cov}(\tilde{h}_{10},\tilde{h}_{10}) & \mathrm{Cov}(\tilde{h}_{10},\tilde{h}_{11}) \\ \mathrm{Cov}(\tilde{h}_{11},\tilde{h}_{1-1}) & \mathrm{Cov}(\tilde{h}_{11},\tilde{h}_{10}) & \mathrm{Cov}(\tilde{h}_{11},\tilde{h}_{11}) \end{pmatrix}.$$

We invert this matrix of complex numbers using the LAPACK utilities `ZGETRF` and `ZGETRI`.

Chapter 6

Analysis of WMAP7 data

The WMAP team has argued, using their version of the quadratic estimator outlined in this work, that there is no significant evidence of a dipolar asymmetry in the CMB data. In this chapter we will do our own analyses and see if the WMAP team is correct, or if the claims from using other methods are still significant.

6.1 Template cleaning

In chapter 4 we described the conjugate gradients method for finding the inverse-variance weighted signal. The noise model for our CG search includes, in addition to the uncorrelated pixel noise, a template cleaning scheme. This serves to remove the monopole and dipole modes from the map, which are contaminated by the kinematic movement across the sky. The noise data model can then be written as

$$n_{\text{tot}} = n + T^T a, \tag{6.1}$$

where n is the pixel noise, T is a matrix of template maps (monopole and dipole) of size $4 \times N_{\text{pix}}$, with a being a coefficient vector denoting the strength of the templates. Taking the covariance of this we get

$$N_{\text{tot}} = \left\langle n_{\text{tot}} n_{\text{tot}}^T \right\rangle = \left\langle (n + T^T a)(n + T^T a)^T \right\rangle = \left\langle nn^T \right\rangle + T^T \left\langle aa^T \right\rangle T. \tag{6.2}$$

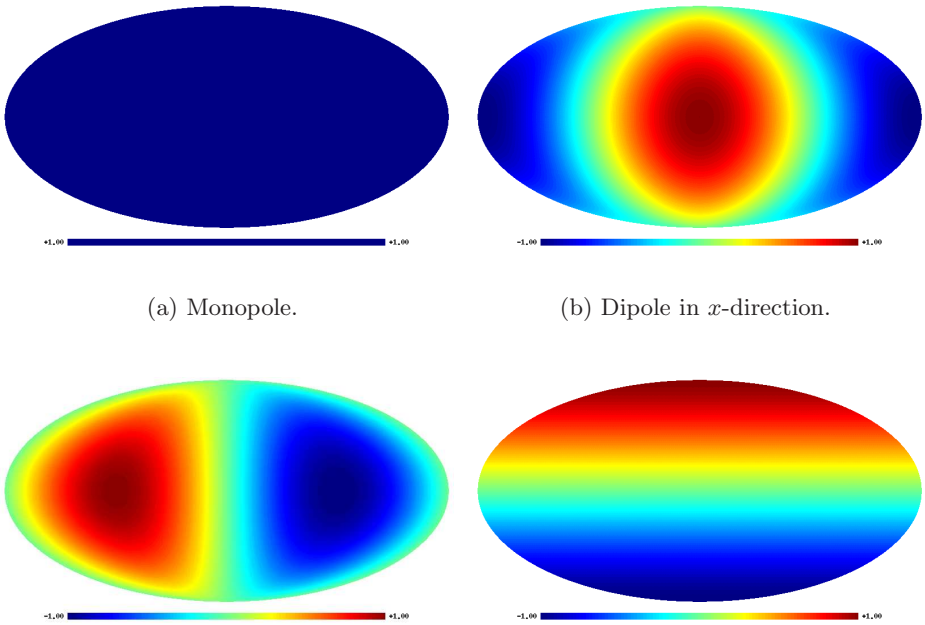
Here $\left\langle nn^T \right\rangle = N$ is the ordinary covariance of the pixel noise. The covariance of the template coefficients, $\left\langle aa^T \right\rangle = C_t$, is taken do be $\sim \infty$. Thus we have

$$N_{\text{tot}} = N + T^T C_t T, \tag{6.3}$$

with the inverse given by

$$N_{\text{tot}}^{-1} = \left(N + T^T C_t T\right)^{-1} = N^{-1} - N^{-1} T^T \left(C_t^{-1} + T N^{-1} T^T\right)^{-1} T N^{-1} \tag{6.4}$$

$$\simeq N^{-1} - N^{-1} T^T \left(T N^{-1} T^T\right)^{-1} T N^{-1}, \tag{6.5}$$



(a) Monopole.

(b) Dipole in x-direction.

(c) Dipole in y-direction.

(d) Dipole in z-direction.

Figure 6.1: Template maps.

where the decomposition of the inverse in the second equality has been done using the *Sherman-Morrison-Woodbury formula* [31]. The four templates are shown in figure 6.1. By adding a quite large monopole and dipoles to isotropic simulations, we find that this scheme works as it should. Note that this only removes the kinematic dipole; the (possibly primordial) dipole modulation we are interested in analyzing should extend to all scales. The ordinary pixel noise covariance N also incorporates the mask, setting N^{-1} to zero for masked pixels.

6.2 Analyzing the full data set

The QML estimator allows us to analyze the full data set, without having to downgrade and smooth the data. Since the original data is strongly contaminated by foregrounds, we mainly use the foreground-reduced data sets provided by the WMAP team on LAMBDA[1]. We will use the KQ85 mask and 7-year data unless otherwise noted. To more directly compare with the results obtained by Hanson & Lewis (2009) and Hoftuft et al. (2009), the 5-year data will also be analyzed.

[1]LAMBDA is NASA's web page for WMAP and other CMB projects, `http://lambda.gsfc.nasa.gov/`.

Frequency	Q-band	V-band	W-band
σ_o	$2197\,\mu$K	$3137\,\mu$K	$6549\,\mu$K

Table 6.1: The values of σ_0 for each frequency band.

6.2.1 Creating the noise rms maps

When analyzing the full data set, we need to use the real noise data. The WMAP sky maps are released every two years, and so far the 3-year, 5-year and 7-year data are available. There is a certain inherent electronic noise to each pixel, but with each release the impact of this noise is lowered, since there are more observations of each point on the sky. The relationship between pixel noise and number of observations is

$$\sigma = \frac{\sigma_0}{\sqrt{N_{\rm obs}}}, \tag{6.6}$$

where σ is the final noise rms, σ_0 the inherent noise rms of each pixel, and $N_{\rm obs}$ the number of observations. The values of σ_0 are provided on LAMBDA, and are shown in table 6.1. The value of $N_{\rm obs}$ is given in the sky maps, and calculated 7-year noise rms maps for the three frequencies Q (41 GHz), V (61 GHz) and W (94 GHz) are shown in figure 6.2. The complicated structure is due to how the WMAP satellite scans the sky, so some pixels get more observations than others. For instance, the ecliptic poles get the highest number of observations as they are scanned with each hourly spacecraft precession, and this is reflected in the blue regions on the noise maps. For details on the scanning strategy see [23].

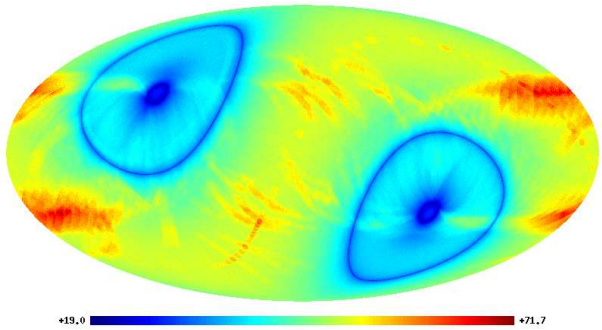
6.2.2 Analysis results

The full data is analyzed by creating the Fisher matrix from an ensemble of isotropic Monte Carlo simulations, where the simulations employ the same mask and noise model as the data in question. We have performed 1000 simulations unless noted otherwise, and the CG search for both the data and the simulations has been done to a precision of $\epsilon = 1.0 \cdot 10^{-6}$. We first compare our results at low $l_{\rm mod}$ with those obtained by Hoftuft et al. [12], and then extend our analysis to include all possible l, to compare with Hanson & Lewis [18] and Bennett et al. [17].

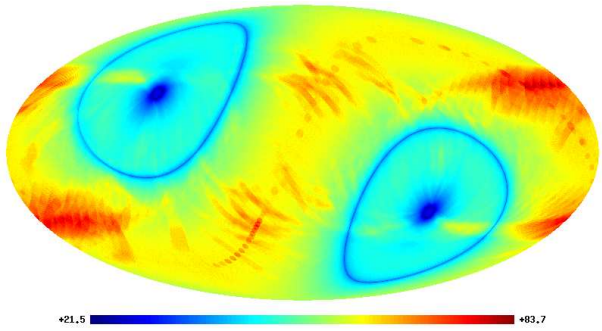
Results from a low-l analysis

The reported results from Hoftuft et al. are shown in table 6.2, for the data sets directly comparable to those in our analysis. It should be kept in mind that their analysis is for the 5-year data, however.

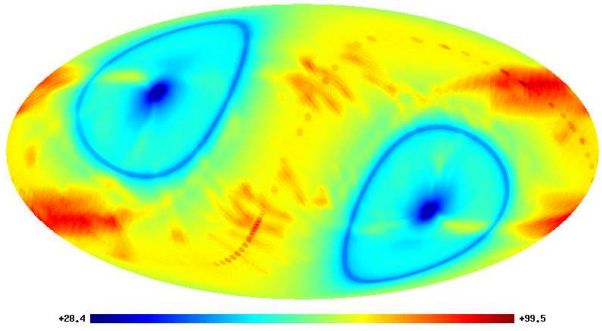
Running our estimator on the W-band up to an $l_{\rm mod} = 64$ we get an amplitude $\hat{A} = 0.0696$ and direction $(\hat{\theta}, \hat{\phi}) = (111^\circ, 223^\circ)$, which is very consistent with their



(a) Q-band.



(b) V-band.



(c) W-band.

Figure 6.2: Noise rms maps for 7-year WMAP data, in units of μK.

Data	l_{mod}	$\hat{A}$	$(\hat{\theta}, \hat{\phi})$
V-band	40	0.119 ± 0.034	$(112^\circ, 224^\circ) \pm 24^\circ$
V-band	64	0.080 ± 0.021	$(112^\circ, 232^\circ) \pm 23^\circ$
V-band	80	0.070 ± 0.019	$(107^\circ, 235^\circ) \pm 22^\circ$
W-band	64	0.074 ± 0.021	$(112^\circ, 232^\circ) \pm 24^\circ$

Table 6.2: The reported results from Hoftuft et al. (2009) [12], using maps downgraded to $N_{\mathrm{side}} = 32$. The downgraded mask was created as outlined in section 2.2.4, and is based on the KQ85 mask. The confidence regions are 68%.

results. On the V-band we get, where the subscript denotes l_{mod},

$$\hat{A}_{40} = 0.0918, \quad (\hat{\theta}, \hat{\phi})_{40} = (104^\circ, 209^\circ)$$
$$\hat{A}_{64} = 0.0705, \quad (\hat{\theta}, \hat{\phi})_{64} = (113^\circ, 222^\circ)$$
$$\hat{A}_{80} = 0.0423, \quad (\hat{\theta}, \hat{\phi})_{80} = (102^\circ, 228^\circ).$$

These results are for the most part consistent with their given confidence regions, except for the amplitude value at $l_{\mathrm{mod}} = 80$, which is off by 0.009, or roughly 18%. They also performed analyses using an extended mask[2] based on KQ85, and comparing with those we again get consistent results across all bands. For the Q-band on $l_{\mathrm{mod}} = 64$ they reported an amplitude $\hat{A} = 0.088 \pm 0.022$ with direction $(\hat{\theta}, \hat{\phi}) = (111^\circ, 245^\circ) \pm 23^\circ$, while we get

$$\hat{A} = 0.075, \quad (\hat{\theta}, \hat{\phi}) = (113^\circ, 224^\circ).$$

Our results for the amplitudes are well outside the 68% confidence regions calculated from our isotropic simulations. For $l_{\mathrm{mod}} = 64$ they are also outside the 95% region for all bands (the upper limit is 0.0656, 0.0650 and 0.0656 for the V-, W- and Q-bands, respectively).

For a more direct comparison we have also analyzed the 5-year data on the V-band, using the 5-year KQ85 mask. The results are

$$\hat{A}_{40}^{5\text{ year}} = 0.1090, \quad (\hat{\theta}, \hat{\phi})_{40}^{5\text{ year}} = (100^\circ, 214^\circ)$$
$$\hat{A}_{64}^{5\text{ year}} = 0.0711, \quad (\hat{\theta}, \hat{\phi})_{64}^{5\text{ year}} = (109^\circ, 219^\circ)$$
$$\hat{A}_{80}^{5\text{ year}} = 0.0454, \quad (\hat{\theta}, \hat{\phi})_{80}^{5\text{ year}} = (98^\circ, 227^\circ),$$

which is also consistent with their analyses, again apart from $l_{\mathrm{mod}} = 80$. Also here the reported amplitudes are well outside the 68% confidence levels, and for the two lowest l_{mod} they are also outside the 95% level.

There is an indication, however, that the amplitude tends to decay with l_{mod}, which is what will be explored further in the full analysis below. This behaviour was only

[2]Their extended mask was constructed by smoothing with the same beam as the data, and setting all resulting pixels having a value less than 0.99 to 0, and the rest to 1. Compared to the other approach this resulted in maps that masked out an additional 10% of the sky.

hinted at by Hoftuft et al., while both Hansen et al. and Hanson & Lewis clearly showed signs of such a trend being present.

The fact that our results deviate from those of Hoftuft et al. when the l_{mod} is as high as 80, could possibly point to that the approximation of one iteration in the quadratic estimator is not valid for higher l_{mod}. The likelihood gets more focused with including more data, so the iterative method might perform worse on the resulting, steeper function.

Results from a full analysis

Here we perform the full analysis for several different l_{mod}, extending all the way up to the maximum l available. We add that for the low-l analysis above our values also seemed to be consistent with those shown in the graph of FIG. 3 in Hanson & Lewis. Of special note in that graph is the big drop in amplitude from $l_{\mathrm{mod}} = 64$ to $l_{\mathrm{mod}} = 100$. For the 7-year W-band the values we obtain are

$$\begin{aligned}\hat{A}_{100} &= 0.0262\\ \hat{\theta}_{100} &= 1.72 = 99^\circ\\ \hat{\phi}_{100} &= 4.38 = 251^\circ,\end{aligned}$$

where the amplitude seems to be in close agreement with the results of H&L. With the Fisher matrix given from 1280 simulations with the KQ85 mask and W-band rms noise, the results for a number of l_{mod} are given in table 6.3. These results seem to confirm the indication that the reported amplitude drops considerably the more l's we include in the analysis, though we note that the direction seems to be relatively stable across all 13 different values of l_{mod} in this analysis (in fact, only 0.16% of the simulations show an equally compact clustering[3] of directions with different l_{mod}. For the Q- and V-bands the result is similar; 0.2% and 0.4% respectively.). The estimated amplitude together with that expected from simulations is shown in figure 6.3, along with the corresponding 68% confidence intervals of the simulations. Maps of the dipole directions are shown in figure 6.4. The amplitude "hovers" around the edge of the confidence interval, except in a few places where it is notably higher. Based on the simulations we have also calculated the probability for a simulation to have higher reported dipole modulation amplitude than the data (simply by adding up how many give a higher amplitude and dividing by the total number of simulations), and the result is shown as the red curve in figure 6.5(b). This matches somewhat the trend shown in Bennett et al. and the significance plots in H& L. We do however get some weird results for higher l_{mod},

[3]How clustered the different estimated directions are have been found by calculating the highest angular separation on the spherical surface between two directions. The maximum separation is 180°, so this simple method does not serve to finding out if the directions extend over the entire sphere. It also does not distinguish between cases where only a few outliers make the separation much larger than for most of the estimated directions, or cases where the directions are completely randomly spread. However, remedying this by e.g. throwing away the 20% highest separations in each case does not give particularly different results.

namely the steep increase in amplitude / decrease in probability towards the end. This could possibly be related to there being some residual signal power above our chosen $l_{\max} = 2 \cdot N_{\text{side}} = 1024$ for this band, which could be tested by repeating the analysis for e.g. $l_{\max} = 1400$. Regardless, the probability is highly dependent on the choice of l_{mod} in the estimator. Interpreting such a probability as statistical significance, this seems to support Bennett et al.'s notion that the *a posteriori* choice of l_{mod} could inflate the statistical significance of the asymmetry.

When analyzing the V-band data this increase towards the end is not seen, and our results seem to be in good agreement with H&L throughout. Plots of the estimated amplitude as function of l_{mod} and the probability of a simulation to report higher amplitude for all three bands are shown in figure 6.5, and confidence regions are included in figure 6.6. The lowest probabilities we see are for $l_{\text{mod}} = 40$ and $l_{\text{mod}} = 64$, though the grid in l_{mod} is quite coarse. Bennett et al. did a similar calculation for a finer grid and found the probability to be lowest at $l_{\text{mod}} = 67$. The lowest probabilities we find are 0.007 (Q-band), 0.01 (V-band) and 0.014 (W-band) using our coarse grid.

Repeating the analysis for a fine grid in the V-band, to compare with Bennett et al., we find like them the probability to be lowest at $l_{\text{mod}} = 67$, with only 1% of the simulations showing as high amplitude. A plot of this is shown in figure 6.7, which matches very well their result up to at least $l_{\text{mod}} = 200$ (the differences at higher multipoles could be due to the fact that their plot is for the coadded V+W band while we have analyzed the V-band only). The other low spikes occur at $l_{\text{mod}} = 15$ and $l_{\text{mod}} = 37$, with probabilities of 0.013 and 0.016, respectively. Comparing the estimated directions on this fine grid, only 0.2% of our 1000 simulations show a comparably clustered set of estimated directions at different l_{mod}. The data shows as little as 75° separation, and throwing away the 20% largest separations for the data and each simulation, to minimize the effect of outliers, results in the probability only staying at 0.2%. For the Q-band the behaviours are similar, however, *none* of the 1000 simulations have as clustered estimates of the direction as the data. This increases to 0.1% when discarding the 20% largest separations. For the W-band as well, none of the simulations show as clustered directions as the data.

The same trends are seen when applying the KQ75 mask, where we have restricted ourselves to the V-band. The lowest probabilities are at $l = 40$ and $l = 64$ (for the coarse grid of 13 different l_{mod}) and there is a decreasing amplitude. Also here the reported direction stays more or less constant, with only 0.1% of the simulations showing a similarly dense clustering on the coarse grid and 0.5% on the fine grid. The direction is consistent with the other results, with $(\hat{\theta}, \hat{\phi})$ being $(109°, 203°)$ at $l_{\text{mod}} = 64$, and $(109°, 227°)$ at $l_{\text{mod}} = 100$. The maximum separation between two directions is 79°. Plots of the amplitude and significance compared with the KQ85 case are shown in figure 6.8.

The 5-year data, where we also have restricted ourselves to the V-band, show the same trends with regards to the clustering of directions, with only 0.1% of the simulations being as clustered for the fine grid. The estimated directions are also similar to the other cases. The p-values for the amplitudes are generally somewhat lower, but follow the same pattern, specifically they are 0.4% at $l_{\text{mod}} = 37$ and 0.1% at $l_{\text{mod}} = 68$.

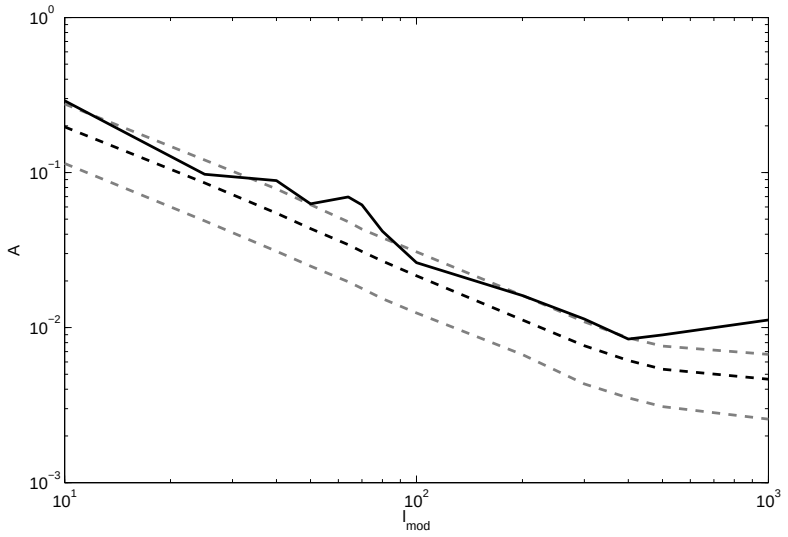


Figure 6.3: Amplitude vs l_{mod} for 7-year W-band data (solid black) compared to the expectation from simulations (dashed black) and their 68% confidence intervals (dashed grey) on $N_{\mathrm{side}} = 512$.

In general the differences are much smaller between the year-sets than for these two spikes.

These results suggest that something is indeed going on with regards to an asymmetry in a certain direction on the sky, but that it should be described as something else than the simple dipole modulation extending to all scales, as that model does not yield particularly strong significances on the amplitude when l_{mod} is increased above ~ 80. It can be noted that the amplitude is consistently estimated as higher than the expectation value, however (apart from a few drops below in some of the data sets).

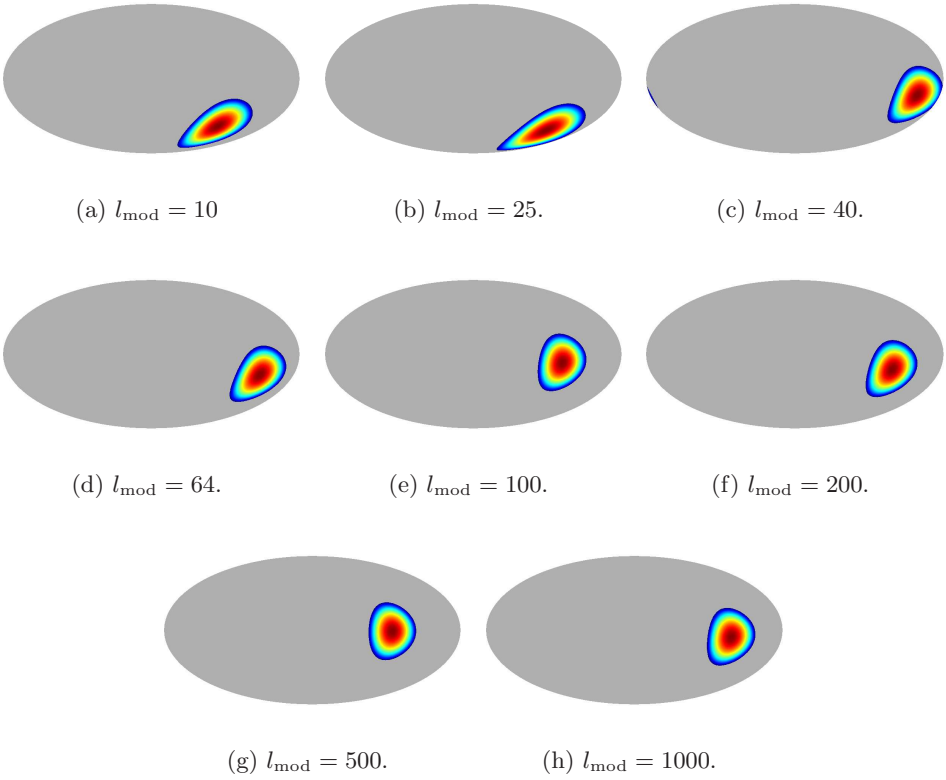
Results for raw maps

To investigate if the foreground-subtraction has an effect on the observed asymmetry we should also analyse the raw data maps. This has been done in detail for the V-band, and the data again give similar results (somewhat higher amplitude), and consistent directions across all l_{mod}, with none of the simulations being as clustered for the coarse grid, and 0.1% for the fine grid. The significances are consistently higher than for the foreground-reduced case, confirming the results of H&L, and plots of both the amplitude and the significance are shown in figure 6.9. Although the results are shown explicitly only for the V-band the trends are the same also for the other bands.

Using the KQ75 mask the behaviour is similar, but we see from the plot that the differences in significance between the foreground-reduced and the raw maps are smaller. The difference in amplitude is also smaller. For this mask 0.3% of the simulations have

l_{mod}	$\hat{A}$	$\hat{\theta}$	$\hat{\phi}$
10	0.290	139°	256°
25	0.0974	143°	237°
40	0.0888	107°	208°
50	0.0628	96°	219°
64	0.0696	111°	223°
100	0.0262	99°	251°
200	0.0161	105°	240°
500	0.0090	91°	263°
1000	0.0112	97°	242°

Table 6.3: Estimated dipole values for W-band 7yr data on different l_{mod}.

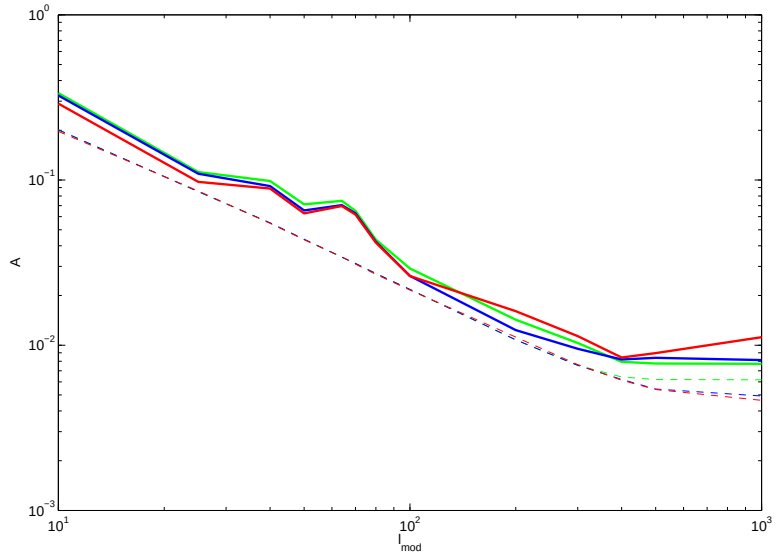


(a) $l_{\mathrm{mod}} = 10$ (b) $l_{\mathrm{mod}} = 25$. (c) $l_{\mathrm{mod}} = 40$.

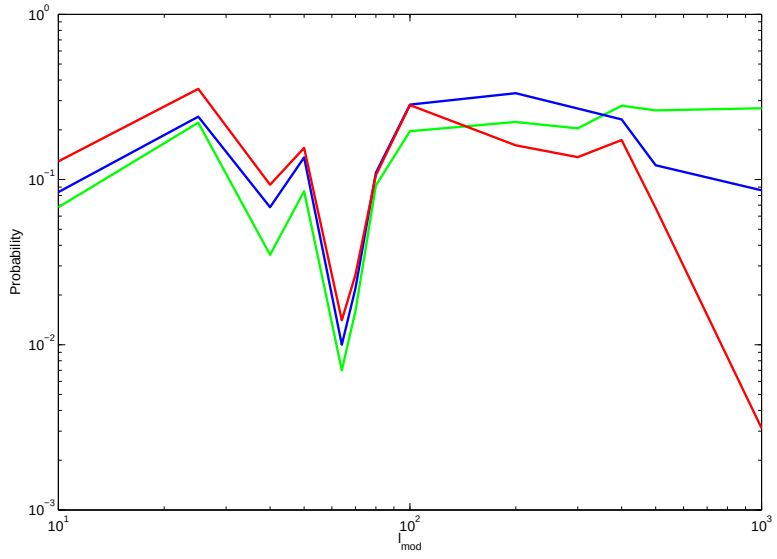
(d) $l_{\mathrm{mod}} = 64$. (e) $l_{\mathrm{mod}} = 100$. (f) $l_{\mathrm{mod}} = 200$.

(g) $l_{\mathrm{mod}} = 500$. (h) $l_{\mathrm{mod}} = 1000$.

Figure 6.4: Maps of the dipole directions on 7-yr W-band data estimated for different l_{mod}. For convenience the maps are greyed out where $\hat{\mathbf{p}} \cdot \hat{\mathbf{n}} > 0.25$.

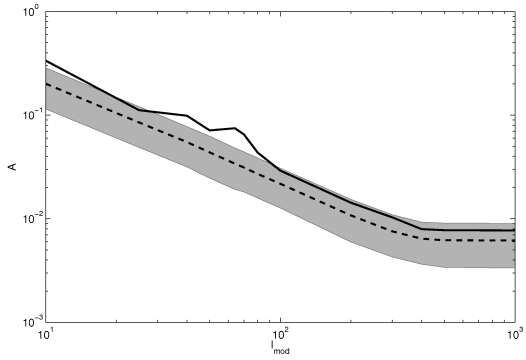


(a) Estimated amplitude as function of l_{mod}. The dashed lines give the expected value for isotropic simulations.

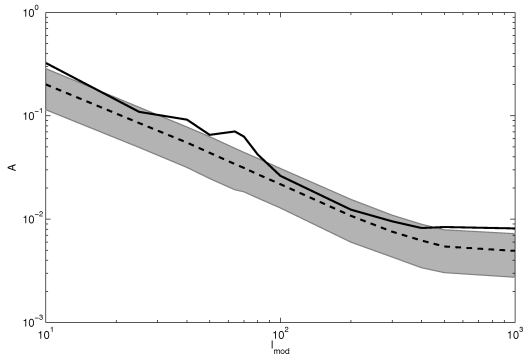


(b) Probability of a simulation reporting higher amplitude than the data, as a function of l_{mod}.

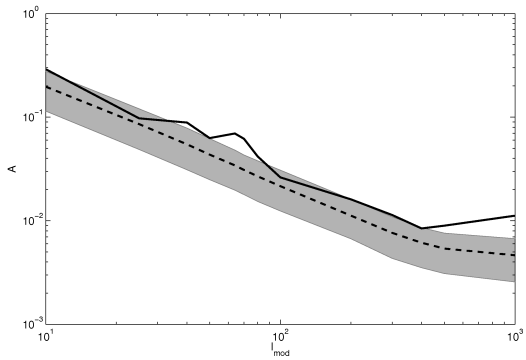
Figure 6.5: Comparisons of the estimated amplitude for simulations vs 7-year data. In both figures the Q-band is green, V-band blue and W-band red.



(a) 7-year Q-band.



(b) 7-year V-band.



(c) 7-year W-band.

Figure 6.6: The estimated amplitude (black) versus the expectation from simulations (dashed black) and the respective 68% confidence regions (grey-shaded).

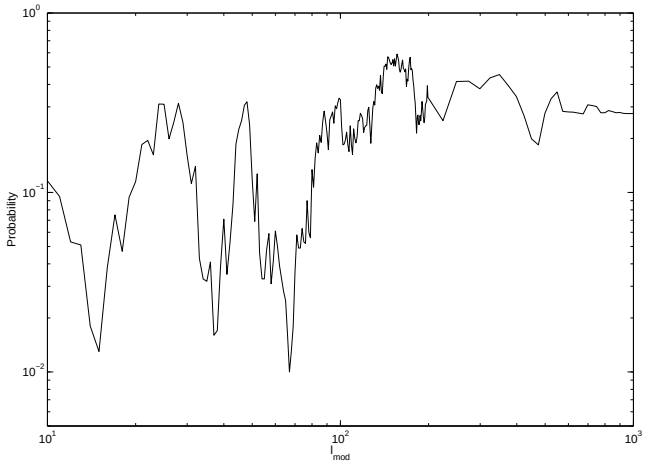
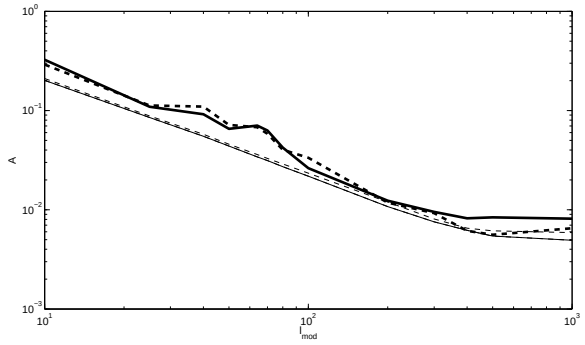


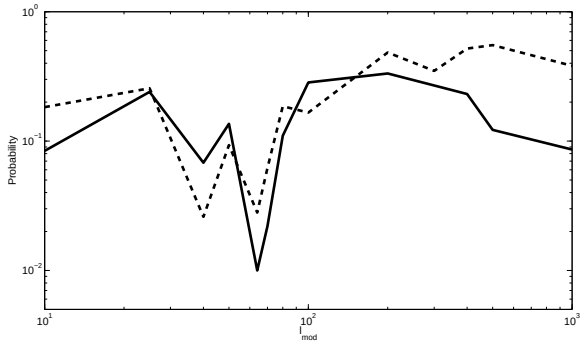
Figure 6.7: Probability of a simulation to yield as high estimates of A as the 7-year V-band data, for different l_{mod} using a fine grid.

	V-band			W-band			Q-band		
l_{mod}	$\hat{A}$	$\hat{\theta}$	$\hat{\phi}$	$\hat{A}$	$\hat{\theta}$	$\hat{\phi}$	$\hat{A}$	$\hat{\theta}$	$\hat{\phi}$
40	0.0918	104°	209°	0.0888	107°	208°	0.0987	105°	209°
64	0.0705	113°	222°	0.0696	111°	223°	0.0748	113°	224°
100	0.0263	111°	243°	0.0262	99°	251°	0.0292	106°	238°
200	0.0123	113°	237°	0.0161	105°	240°	0.0143	105°	224°
500	0.0084	94°	234°	0.0090	91°	263°	0.0078	114°	226°
1000	0.0081	91°	204°	0.0112	97°	242°	0.0077	114°	227°

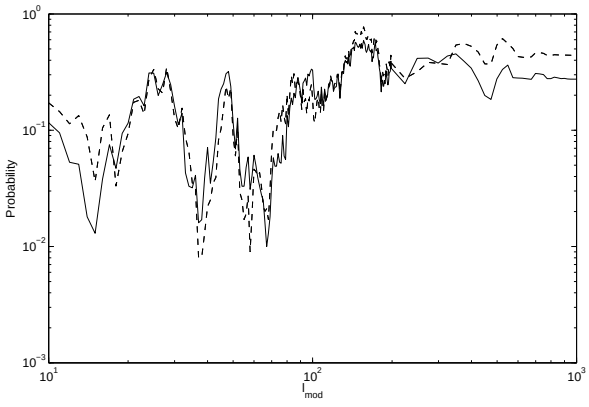
Table 6.4: Estimated dipole values and directions for various 7-year data using the KQ85 mask, on different l_{mod}.



(a) Plot of amplitude.

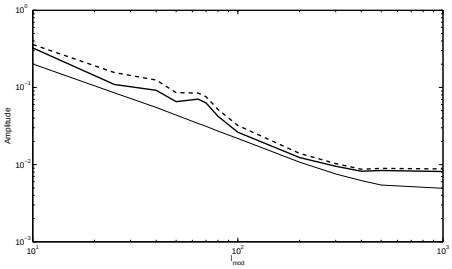


(b) Plot of significance.

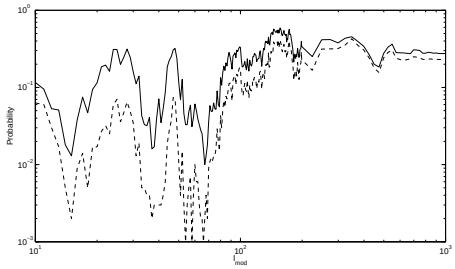


(c) Plot of significance (fine grid, different simulations).

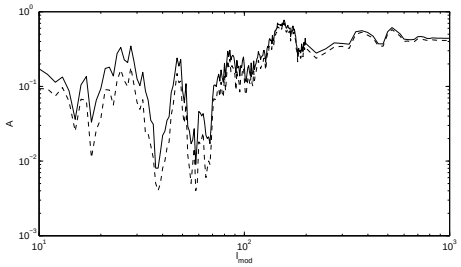
Figure 6.8: Comparison of results on the 7-year V-band using KQ85 (solid) or KQ75 (dashed) masking.



(a) Plot of amplitude (KQ85).



(b) Plot of significance (KQ85).



(c) Plot of significance (KQ75).

Figure 6.9: Comparison of results on the 7-year V-band for foreground-reduced (solid) or raw (dashed) maps. The thin line in (a) is the expected value.

an equally compact set of estimated directions. This analysis has only been performed on the V-band.

As a way to summarize, figure 6.10 shows all the different estimated directions across all l_{mod} for the different bands, projected onto an $N_{\mathrm{side}} = 16$ map, and table 6.5 gives an overview of the clustering compared to simulations.

6.2.3 Using more than one iteration in the estimator

The assumption that we should get good results using only one iteration ought to be tested. We won't be doing that in this thesis, but leave it as a problem for future work, and we will present the basic concept here. The iterative estimator for the dipole modulation is

$$\hat{\mathbf{h}}_{n+1} = \hat{\mathbf{h}}_n + \mathcal{F}_n^{-1}\big[\tilde{\mathbf{h}}_n - \langle\tilde{\mathbf{h}}_n\rangle\big], \tag{6.7}$$

where we have ended it after one iteration assuming no initial modulation $\hat{\mathbf{h}}_0$.

In equation 3.20 we said that the inverse-variance weighted signal was

$$\bar{\Theta} = C^{-1}|_{\boldsymbol{\lambda}=0}\Theta,$$

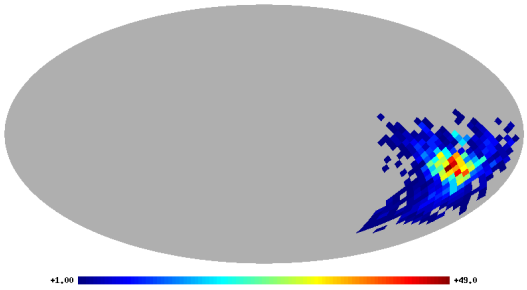


Figure 6.10: All the different estimated directions across 222 values of l_{mod}. The data included are the 7-year V-, Q- and W-bands, both foreground-reduced and raw, employing the KQ85 mask, as well as the raw and reduced 7-year V-band using the KQ75 mask, and the 5-year reduced V-band using the KQ85yr5 mask.

Frequency	V-band				W-band		Q-band	
Data set	85_{red}	85_{raw}	75_{red}	$85^{\mathrm{yr5}}_{\mathrm{red}}$	85_{red}	85_{raw}	85_{red}	85_{raw}
$\vartheta^{\mathrm{data}}_{\mathrm{max}}$	75°	72°	79°	80°	75°	73°	71°	51°
$P(\vartheta^{\mathrm{sim}}_{\mathrm{max}} \leq \vartheta^{\mathrm{data}}_{\mathrm{max}})$	0.2%	0.1%	0.5%	0.1%	0	0	0	0

Table 6.5: Clustering of estimated directions across 222 different values of l_{mod}, where ϑ is the highest angular separation between two directions. The number details the mask used, and the subscript if the data is raw or foreground-reduced. Unless specified the data is 7-year. A probability of 0 means it could not be inferred since no simulations were as clustered as the data.

assuming no modulation. This is correct for the first iteration. For the second iteration, we need to assume the modulation we got from the first iteration, thereby repeating the CG search with the intrinsic covariance not being quite equal to the isotropic C_l. The new Fisher matrix and mean must then be computed from an ensemble of simulations using the results from their second iterations. Thus, the hard part is computing the inverse-variance weighted signal from a non-isotropic intrinsic signal covariance. We have already computed an expression for this new power spectrum, to first order in the modulation, in chapter 3 (equation 3.35), if we set $l = 1$. Computing the Gaunt integral we get

$$\begin{aligned} C_{l'm',l''m''} =& \delta_{l'l''}\delta_{m'm''}C_{l'} + (-1)^{m'}\left[C_{l'} + C_{l''}\right]\sqrt{\frac{(2\cdot 1+1)(2l'+1)(2l''+1)}{4\pi}} \\ &\times \begin{pmatrix} 1 & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \sum_{m=-1}^{1} f_{1m} \begin{pmatrix} 1 & l' & l'' \\ -m & m' & m'' \end{pmatrix}. \end{aligned} \tag{6.8}$$

For the f_{1m} we then insert our first estimate of the dipole modulation coefficients. On the diagonal this new covariance turns out to be the isotropic C_l, since the 3-j symbol $\left(\begin{smallmatrix} 1 & l' & l'' \\ 0 & 0 & 0 \end{smallmatrix}\right)$ ends up being 0 when $l' = l''$, by virtue of A.8. For $l' \neq l''$ the first term vanishes and we are left with

$$\begin{aligned} C_{l'm',l''m''} =& (-1)^{m'}\left[C_{l'} + C_{l''}\right]\sqrt{\frac{(2\cdot 1+1)(2l'+1)(2l''+1)}{4\pi}} \begin{pmatrix} 1 & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \\ &\times \sum_{m=-1}^{1} f_{1m} \begin{pmatrix} 1 & l' & l'' \\ -m & m' & m'' \end{pmatrix}. \end{aligned} \tag{6.9}$$

From the last selection rule in A.4 we have that

$$|l' - l''| \leq 1 \leq |l' + l''|,$$

which means that the only non-vanishing off-diagonal elements are those where l' and l'' differ by just 1. So the matrix will be sparse, but still pretty huge and not quite easy to work with.

Chapter 7

Conclusions / Final words

In this thesis we have explored the possibility of there being a hemispherical power asymmetry in the WMAP data, using a Quadratic Maximum-Likelihood estimator. The power asymmetry has been approximated as a simple dipolar modulation of an intrinsic statistically isotropic signal, a model that has been recently studied in the literature. The QML estimator was developed for this model in Hanson & Lewis (2009), and in chapter 3 we performed our own derivation. We have reimplemented this estimator for our own analyses, with the results from analyzing the 7-year WMAP data shown in chapter 6.

7.1 Is there evidence of a dipolar asymmetry in the WMAP data?

The WMAP team, in Bennett et al. (2011), argue that there is no significant evidence for an anomalous dipolar power asymmetry in the WMAP data, and that the previously reported claims have un-naturally high significance values due to an *a posteriori* bias in the choice of l_{mod}. Based on our results using the QML estimator outlined in this work the reported amplitude of the dipole modulation does not yield very high significances across a large range of l_{mod}, however, we note that the reported direction stays remarkably constant. If there was no modulation, we would expect the reported directions to be more scattered, and indeed, that is what we see when comparing with simulations. The estimated direction is consistent with those of previous works using other methods, and it is consistent with different data (our own analyses show this to be the case for the different bands of WMAP, different sky-cuts and raw/reduced maps, and previous studies have given similar directions for different year-sets, and also partly the COBE data). The reported amplitude is diminishing with the choice of l_{mod} in the estimators however (though it lies above the expected value at nearly all multipoles), and the significance is found to be highly unstable, so the simple hemispherical dipolar asymmetry model is seemingly not good enough for describing the phenomenon. The amplitudes we get are close to the results of Hanson & Lewis. Hansen et al. (2009), who analyzed differences between local power spectra, also noted the diminishing effect, but

again, the reported direction across various l_{mod} stayed more or less constant.

Based on our simulations we find the probability of an isotropic map to report as clustered directions of dipolar asymmetry to lie between 0.1% and 1.0%. Indeed, using a fine grid of 222 different values for l_{mod} the statistical significance can in some cases not be inferred since none of our 1000 simulations show an equally compact clustering.

For the l_{mod} reporting the highest significance of a dipole modulation the probability of isotropic maps to report as high an amplitude is close to 1%, while for higher l_{mod} this number gets closer to 30%. We find that the effect persists, and that both the amplitude and the direction do not vary too much, with regards to either frequency band, sky-cut or raw/reduced maps.

The WMAP team's claim of there being no significant evidence for an anomalous dipolar power anisotropy in the WMAP data seems to be premature. More work still needs to be done. If there is an anomaly, it still might not be of cosmological origin, though systematics are an unlikely explanation since the effect was also observed on the COBE data, and the methods for foreground-subtraction have changed with different year-releases. According to Gold et al. (2011), they find no evidence for foreground contamination outside the current KQ85yr7 analysis mask.

Bennett et al. make the argument that searching specifically for a dipolar or hemispherical power asymmetry is in itself an *a posteriori* choice, and that oddities are bound to occur in any data set. Even though none of our simulated maps may exhibit *this particular* oddity to equal effect, many of them would have *other* highly unlikely oddities. We would however be ill-advised not to keep investigating such possible anomalies; if new physical models can be found that give better fits to the entire data set science will have progressed further.

7.2 Future projects

There still seems to be an effect going on, but the simple model of a scale-invariant dipole modulation should be replaced with something else to make a better fit of the data. Testing models where the modulation amplitude decreases with l would be the next step forward. When it comes to the QML estimator the assumption of one iteration being a sufficiently good approximation is still not tested, and this should be done as well. Comparisons with the results obtained from other methods suggest the approximation is valid up to about $l_{\mathrm{mod}} = 80$, for higher l_{mod} we need new results from exact analyses to compare with. The MCMC analysis methods could be repeated for an l_{mod} of 50 to confirm the pattern of the significance values.

The WMAP experiment is now over, with the final nine-year data sets to be released later this year or early 2012. This will gives us an additional two years of data to lessen the impact of instrumental noise and obtain better foreground models. It is the first publically released data of Planck, however, that is most interesting in the coming years. If the hemispherical power asymmetry is observed to similar confidence also in those data sets, the explanation would almost certainly be physical. The important questions then would be if the anomaly is relevant and if we can find better fits to the

data than the standard ΛCDM-model.

Comparisons with an optimal maximum-likelihood estimator

The traditional methods of performing the full ML-analysis have had poor computational scaling, going as $\mathcal{O}(N_{\text{pix}}^3)$. To make a complete comparison with the quadratic estimator improved methods must be derived. Dag Sverre Seljebotn, who published his Master's Thesis [32] at the University of Oslo last year, developed a Gibbs sampling algorithm for testing the hemispherical asymmetry. This new method scales as $\mathcal{O}(N_{\text{pix}}^{3/2})$, and should be well suited for performing analyses on full-resolution maps.

Further tests of the quadratic estimator

The QML estimator should be tested with more iterations to confirm the assumption of one iteration being sufficient. We saw that we got results which were consistent with the optimal maximum-likelihood approach of Hoftuft et al. (2009) for the lower values of l_{mod}, however, at $l_{\text{mod}} = 80$ we were a little outside the lower bound of their confidence region in amplitude. A possible reason for this could be that the approximation fails when the likelihood function gets more steep and focused, so that the estimated value falls somewhat short of the true maximum. We emphasize, though, that the deviation from their reported confidence region was only slight.

We should also increase the number of simulations we compare with, especially when it comes to comparing the clustering of directions, to more accurately assess the significance.

Possible alternative models for describing the asymmetry

One modification to our simple model of a scale-invariant dipole modulation would of course be extending it to include a scale-dependent amplitude $A(l)$. Without adopting a new model explicitly one way of testing this could be to compute the estimator in bands of certain regions of l, and observe how the reported amplitudes vary.

Another possibility would be to consider other modulation patterns than the simple dipole. The quadratic estimator derived in this text, as written in equation 3.46, should be applicable for any field f which modulates the isotropic signal through

$$\Theta(\hat{\mathbf{n}}) = [1 + f(\hat{\mathbf{n}})]\Theta_{\text{iso}}(\hat{\mathbf{n}}).$$

However, most other fields would not have such an elegant representation in harmonic space as the dipole, where only a few parameters are enough to describe the entire field. So the quadratic estimator should rather be re-derived using a different approach.

Physical models to be given particular attention

We gave a short overview of a few proposed physical models in chapter 1, and several other possibilities have also been found in the literature. Based on our results it seems

that models allowing a scale-dependent amplitude are those which should be most purposefully investigated, of those that concern a type of dipolar modulation pattern.

Improved methods for foreground-subtraction

We see the asymmetric effect when analyzing raw maps, and it is there stronger than in the foreground-reduced case. Applying the KQ85 mask and using the foreground-reduced maps the impact of foreground contaminations should be negligible, but this may not be the case. Perhaps there is some residual foreground contaminations left in the preferred direction. Using the KQ75 mask we see less of a difference between the foreground-reduced and the raw case. This could indicate that the KQ85 mask lets some foreground contamination through which helps to increase the asymmetry effect, and the KQ75 mask effectively blocks this contamination.

Appendix A

Resource

A.1 The HEALPix software

HEALPix[1] (Hierarchical, Equal Area, and iso-Latitude Pixelization) is a software designed for a convenient pixelization scheme on the sphere. The HEALPix format is that used by NASA for their CMB maps taken by the WMAP satellite, and the software has become widely used in cosmology for work regarding the CMB. As its name suggests, the format is based on a few basic principles:

Hierarchical The base-pixels are divided into more pixels at higher resolution in a hierarchical structure in the data base. This ensures faster performance.

Equal Area The pixelization is designed so that each pixel takes up the same area on the surface of the sphere, thereby ensuring the sampling to have minimal regional dependence.

iso-Latitude The pixel centers are positioned on a certain number of rings of constant latitude around the sphere. This is advantageous for computations involving spherical harmonics operations.

The map, or sphere, is divided into 12 base-pixels. These base-pixels can again be divided further into several other pixels, and this is represented by the parameter N_{side}. It details how many new pixel sides the side of the base-pixel is divided into. This makes the total number of pixels in the map equal to

$$N_{\text{pix}} = 12 \cdot N_{\text{side}}^2.$$

The largest maps we use have an N_{side} of 512, or 3 145 728 number of pixels (3 Megapixels). With such a large number of data points the ability to perform spherical harmonics transforms and the like in an efficient and fast manner is crucial. The upcoming Planck datasets will have ~ 50 Megapixels, so this importance can not be overstated. Examples of the pixelization is shown in figure A.1, where the grey areas represent two base-pixels and conveniently shows how they are divided into higher-resolution pixels.

We rely heavily on the use of HEALPix, not only because it is the format the WMAP sky-maps are released in, but also because the software contains lots of useful routines used in our program. Perhaps the most important ones are those for conversion between pixel space and harmonic space; `map2alm` and `alm2map`. Also frequently used are the functions for conversion from the RING pixelization scheme to the NEST scheme, and of course its map plotting routine. The raw FITS maps are then converted to images using `map2gif`, which is a separate program in the HEALPix software which can be run directly from the command line.

[1] See Górski et al., [33] and [34], for documentation. The website is `http://healpix.jpl.nasa.gov/`.

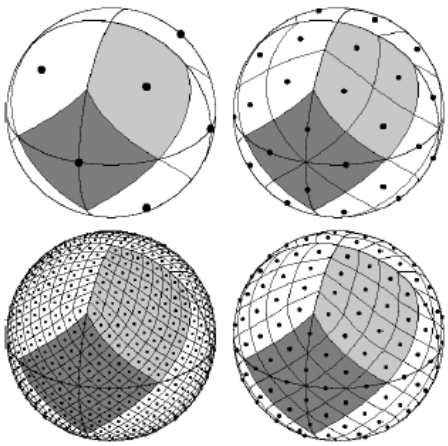


Figure A.1: Examples of the HEALPix pixelization, with (in clockwise order from the upper left) $N_{\mathrm{side}} = 1, 2, 4, 8$. The dots represent pixel centers. Image reprinted from the HEALPix Primer [33].

A.2 The 3-j symbols

The Wigner 3-j symbols come into play in this thesis as a cause of the integral of the products of three spherical harmonics functions, called the Gaunt integral. This integral is given as follows:

$$\begin{aligned}&\int Y_{lm}(\theta,\phi)Y_{l_1m_1}(\theta,\phi)Y_{l_2m_2}(\theta,\phi)\,\mathrm{d}\Omega\\&=\sqrt{\frac{(2l+1)(2l_1+1)(2l_2+1)}{4\pi}}\begin{pmatrix}l & l_1 & l_2\\0 & 0 & 0\end{pmatrix}\begin{pmatrix}l & l_1 & l_2\\m & m_1 & m_2\end{pmatrix}.\end{aligned}\tag{A.1}$$

The actual 3-j symbols are those given on the form

$$\begin{pmatrix}l & l_1 & l_2\\m & m_1 & m_2\end{pmatrix}.\tag{A.2}$$

The symbol arises in quantum mechanics and can be expressed as

$$\begin{pmatrix}l & l_1 & l_2\\m & m_1 & m_2\end{pmatrix}\equiv\frac{(-1)^{l-l_1-m_2}}{\sqrt{2l_2+1}}\langle lml_1m_1|l_2-m_2\rangle,\tag{A.3}$$

where the expression in brackets are the Clebsch-Gordan coefficients. The symbols can be calculated analytically but are more commonly found using recursive relations.

There are a number of selection rules which need to be satisfied in order for this symbol to not equal 0:

$$\begin{aligned}&m+m_1+m_2=0\\&l+l_1+l_2\text{ is an integer}\\&|m_i|\leq l_i\\&|l_1-l_2|\leq l\leq|l_1+l_2|.\end{aligned}\tag{A.4}$$

In addition there are also a number of symmetry relations, and the ones most relevant for this thesis are:

1. A 3-j symbol is invariant under an even permutation of its columns:

$$\begin{pmatrix}l & l_1 & l_2\\m & m_1 & m_2\end{pmatrix}=\begin{pmatrix}l_1 & l_2 & l\\m_1 & m_2 & m\end{pmatrix}=\begin{pmatrix}l_2 & l & l_1\\m_2 & m & m_1\end{pmatrix}.\tag{A.5}$$

2. An odd permutation of its columns gives a phase factor:

$$\begin{aligned}\begin{pmatrix}l & l_1 & l_2\\m & m_1 & m_2\end{pmatrix}&=(-1)^{l+l_1+l_2}\begin{pmatrix}l_1 & l & l_2\\m_1 & m & m_2\end{pmatrix}\\&=(-1)^{l+l_1+l_2}\begin{pmatrix}l & l_2 & l_1\\m & m_2 & m_1\end{pmatrix}.\end{aligned}\tag{A.6}$$

3. Changing the signs of the m's also give a phase factor:

$$\begin{pmatrix} l & l_1 & l_2 \\ m & m_1 & m_2 \end{pmatrix} = (-1)^{l+l_1+l_2} \begin{pmatrix} l & l_1 & l_2 \\ -m & -m_1 & -m_2 \end{pmatrix}. \tag{A.7}$$

The final symmetry relation given here also ensures that for a 3-j symbol with $m = m_1 = m_2 = 0$, the following selection rule holds:

$$\text{if } m = m_1 = m_2 = 0, \text{ then } l + l_1 + l_2 \text{ must be even.} \tag{A.8}$$

Also of note is that, in the Gaunt integral of equation A.1, if the spherical harmonics function Y_{lm} is complex conjugated, then that introduces a phase factor $(-1)^m$ and a change of the sign of the appropriate m in the second 3-j symbol.

Calculation of the 3-j symbols in the program are done using the `DRC3JJ` subroutine provided by SLATEC[2]. For further information on the symbols see, e.g., Varshalovich et al. (1988) [35].

[2] `http://www.netlib.org/slatec/`.

A.3 Spherical harmonics

The spherical harmonics are functions which come in as solution eigenfunctions to the *Laplace equation*, $\nabla^2\Psi = 0$, in spherical coordinates. See for instance [35] for an overview. The definition is

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_{lm}(\cos\theta)e^{im\phi}, \quad l \geq 0,\ |m| \leq l, \tag{A.9}$$

where the P_{lm} are the associated Legendre functions. For negative m we have the relation

$$Y_{l-m} = (-1)^m Y_{lm}^*.$$

The following are a number of properties relevant for this thesis:

The spherical harmonics are orthonormal,

$$\int_\Omega Y_{lm}Y_{l'm'}^*\,\mathrm{d}\Omega = \delta_{ll'}\delta_{mm'}. \tag{A.10}$$

"Any" field on the sphere can be expanded in terms of the spherical harmonics, just like "any" function in flat space can be expanded in terms of Fourier series:

$$f(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}Y_{lm}(\theta,\phi), \tag{A.11}$$

where the spherical harmonics coefficients are given by

$$a_{lm} = \int_\Omega f(\theta,\phi)Y_{lm}^*(\theta,\phi)\,\mathrm{d}\Omega. \tag{A.12}$$

For real fields $f(\theta,\phi)$, these coefficients satisfy the relation

$$a_{l-m} = (-1)^m a_{lm}^*. \tag{A.13}$$

There is a relationship between the spherical harmonics and the Legendre polynomials P_l, given by

$$P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{y}}) = \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y_{lm}(\hat{\mathbf{x}})Y_{lm}^*(\hat{\mathbf{y}}), \tag{A.14}$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit vectors.

A.3.1 Real-valued version

We also utilize the real-valued versions of the spherical harmonics in this work. For a complex harmonic function a_{lm}^C we have that

$$a_{lm}^R = \begin{cases} a_{lm}^C & \text{if } m = 0 \\ \sqrt{2}\cdot\mathrm{Re}(a_{lm}^C) & \text{if } m > 0 \\ \sqrt{2}\cdot\mathrm{Im}(a_{l-m}^C) & \text{if } m < 0. \end{cases} \tag{A.15}$$

These a_{lm}'s are stored in vectors of length $2l+1$ for each l, so that element 1 corresponds to $m = -l$, element 2 is $m = -l+1$, element 3 is $m = -l+2$, and so on. The mapping from complex to real space can be written as a transformation by a unitary matrix $\mathbf{U}$, which for the $l = 2$ case is given by

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} i & 0 & 0 & 0 & -i \\ 0 & -i & 0 & -i & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{A.16}$$

Transforming an entire matrix $\mathbf{A}$ of complex spherical harmonics to the real realm we compute

$$\mathbf{A}^R = \mathbf{U}\mathbf{A}\mathbf{U}^\dagger. \tag{A.17}$$

For any given l, the matrix $\mathbf{A}$ has dimension $(2l+1)\times(2l+1)$ with the elements ordered analogously to the scheme outlined earlier.

Bibliography

[1] S. Dodelson. *Modern cosmology*. Academic Press, 2003.

[2] A. R. Liddle. *An Introduction to Modern Cosmology*. Wiley, 2003.

[3] A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *ApJ*, 142:419P, 1965.

[4] Øyvind Grøn. *Lecture Notes on the General Theory of Relativity*. Springer Science+Business Media, 2009.

[5] A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, January 1981.

[6] M. Tegmark. Doppler Peaks and all that: CMB Anisotropies and what they can tell us. In S. Bonometto, J. R. Primack, & A. Provenzale, editor, *Dark Matter in the Universe*, pages 379–+, 1996.

[7] P. Callin. How to calculate the CMB spectrum. *ArXiv Astrophysics e-prints*, June 2006.

[8] D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters. *The Astrophysical Journal Supplement Series*, 192:16–+, February 2011.

[9] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Górski, and P. B. Lilje. Asymmetries in the Cosmic Microwave Background Anisotropy Field. *Astrophysical Journal*, 605:14–20, April 2004.

[10] F. K. Hansen, A. J. Banday, K. M. Górski, H. K. Eriksen, and P. B. Lilje. Power Asymmetry in Cosmic Microwave Background Fluctuations from Full Sky to Sub-Degree Scales: Is the Universe Isotropic? *Astrophysical Journal*, 704:1448–1458, October 2009.

[11] H. K. Eriksen, A. J. Banday, K. M. Górski, F. K. Hansen, and P. B. Lilje. Hemispherical Power Asymmetry in the Third-Year Wilkinson Microwave Anisotropy Probe Sky Maps. *The Astrophysical Journal Letters*, 660:L81–L84, May 2007.

[12] J. Hoftuft, H. K. Eriksen, A. J. Banday, K. M. Górski, F. K. Hansen, and P. B. Lilje. Increasing Evidence for Hemispherical Power Asymmetry in the Five-Year WMAP Data. *Astrophysical Journal*, 699:985–989, July 2009.

[13] N. E. Groeneboom, M. Axelsson, D. F. Mota, and T. Koivisto. Imprints of a hemispherical power asymmetry in the seven-year WMAP data due to non-commutativity of space-time. *ArXiv e-prints*, November 2010.

[14] F. Paci, A. Gruppuso, F. Finelli, P. Cabella, A. de Rosa, N. Mandolesi, and P. Natoli. Power asymmetries in the cosmic microwave background temperature and polarization patterns. *Monthly Notices of the Royal Astronomical Society*, 407:399–404, September 2010.

[15] A. Bernui. Anomalous CMB north-south asymmetry. *Phys. Rev. D*, 78(6):063531–+, September 2008.

[16] D. Pietrobon, A. Amblard, A. Balbi, P. Cabella, A. Cooray, and D. Marinucci. Needlet detection of features in the WMAP CMB sky and the impact on anisotropies and hemispherical asymmetries. *Phys. Rev. D*, 78(10):103504–+, November 2008.

[17] C. L. Bennett, R. S. Hill, G. Hinshaw, D. Larson, K. M. Smith, J. Dunkley, B. Gold, M. Halpern, N. Jarosik, A. Kogut, E. Komatsu, M. Limon, S. S. Meyer, M. R. Nolta, N. Odegard, L. Page, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Are There Cosmic Microwave Background Anomalies? *The Astrophysical Journal Supplement Series*, 192:17–+, February 2011.

[18] D. Hanson and A. Lewis. Estimators for CMB statistical anisotropy. *Phys. Rev. D*, 80(6):063004, Sep 2009.

[19] L. Ackerman, S. M. Carroll, and M. B. Wise. Imprints of a primordial preferred direction on the microwave background. *Phys. Rev. D*, 75(8):083502–+, April 2007.

[20] A. L. Erickcek, M. Kamionkowski, and S. M. Carroll. A hemispherical power asymmetry from inflation. *Phys. Rev. D*, 78(12):123520–+, December 2008.

[21] A. L. Erickcek, C. M. Hirata, and M. Kamionkowski. A scale-dependent power asymmetry from isocurvature perturbations. *Phys. Rev. D*, 80(8):083507–+, October 2009.

[22] T. Koivisto and D. F. Mota. Accelerating Cosmologies with an Anisotropic Equation of State. *Astrophysical Journal*, 679:1–5, May 2008.

[23] C. L. Bennett, M. Bay, M. Halpern, G. Hinshaw, C. Jackson, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, D. T. Wilkinson, E. Wollack, and E. L. Wright. The Microwave Anisotropy Probe Mission. *Astrophysical Journal*, 583:1–23, January 2003.

[24] B. Gold, N. Odegard, J. L. Weiland, R. S. Hill, A. Kogut, C. L. Bennett, G. Hinshaw, X. Chen, J. Dunkley, M. Halpern, N. Jarosik, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Galactic Foreground Emission. *The Astrophysical Journal Supplement Series*, 192:15–+, February 2011.

[25] M. Limon et al. Wilkonson Microwave Anisotropy Probe (WMAP): Seven-Year Explanatory Supplement, April 2010.

[26] Jay L. Devore and Kenneth N. Berk. *Modern Mathematical Statistics with Applications*. Thomson Brooks/Cole, 2007.

[27] Jonathan R. Shewchuk. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1994.

[28] H. K. Eriksen, I. J. O'Dwyer, J. B. Jewell, B. D. Wandelt, D. L. Larson, K. M. Górski, S. Levin, A. J. Banday, and P. B. Lilje. Power Spectrum Estimation from High-Resolution Maps by Gibbs Sampling. *The Astrophysical Journal Supplement Series*, 155:227–241, December 2004.

[29] K. M. Smith, O. Zahn, and O. Doré. Detection of gravitational lensing in the cosmic microwave background. *Phys. Rev. D*, 76(4):043510–+, August 2007.

[30] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen. MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets. *Astrophysical Journal*, 567:2–17, March 2002.

[31] David A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer, 1997.

[32] Dag Sverre Seljebotn. Hemispherical Power Asymmetry in the Cosmic Microwave Background by Gibbs Sampling. Master's thesis, the University of Oslo, 2010.

[33] K. M. Górski et al. The HEALPix primer. *astro-ph/9905275*, 2000.

[34] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *Astrophysical Journal*, 622:759–771, April 2005.

[35] D.A. Varshalovich, A.N. Moskalev, and V.K. Khersonskiĭ. *Quantum theory of angular momentum: irreducible tensors, spherical harmonics, vector coupling coefficients, 3nj symbols*. World Scientific Pub., 1988.