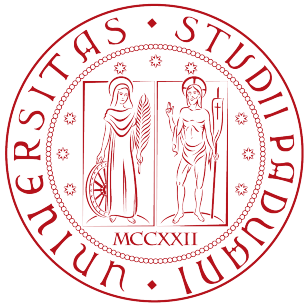


Università degli Studi di Padova

Dipartimento di Fisica ed Astronomia "Galileo Galilei"

Corso di Laurea Magistrale in Fisica

Tesi di Laurea

Exploiting foreground non-Gaussianity in component separation for Cosmic Microwave Background analysis

Candidato: Omar Darwish
Relatore: Prof. Michele Liguori

Anno Accademico 2016/2017

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

1

INTRODUCTION

Today, our cosmological theory is mainly explained by the $\Lambda - CDM$ model plus the theory of Inflation, which predicts an accelerated expansion of the Universe $10^{-35}s$ after the Big Bang and the origin of the rich structure of our Universe. Furthermore, this theory predicts also the generation of a stochastic background of primordial gravitational waves.

One of the main success of our cosmological theory is the prediction of a snapshot of the primordial Universe given by the photons of the Cosmic Microwave Background radiation, one of the best probes of our Universe and its history. Cosmology has been revolutionized by the precision measurements of the CMB, since its first detection in 1965.

The first big studies of the CMB were focused on its temperature anisotropies. Only recently, after 2000's, physicist managed to detect also the CMB polarizations anisotropies. There are two sources of polarization anisotropies; one of them is sourced by gravitational waves from Inflation. A detection of this type of polarization, called $B-$mode, would be a strong confirmation of Inflation.

The still undetected $B-$mode is expected to have an amplitude $\sim$ 100, or more, times smaller than that of the temperature anisotropies.

In addition to this, in the Universe there is a variety of contaminating signals, the foregrounds, that make the extraction of the primordial $B-$mode signal harder. Indeed, it is expected that the $B-$mode signal is $\leq 1\%$ of the foreground emission. The largest contamination comes from emissions in our galaxy, while a small fraction comes from extragalactic sources. Therefore, the development of foreground substraction methods has become increasingly more important, and many methods have been proposed so far, based on analysis of data at different frequencies and on the specific frequency dependence of the astrophyisical emission laws. Indeed foreground emission is typically characterized by power law frequency distribution opposed to the black body emission of the CMB.

Furthermore, the foreground spatial distribution is clearly non-Gaussian, contrary to the CMB which has been measured to be Gaussian to a $< 0.1\%$ level. While the fequency differences are typically exploited foreground cleaning algorithms statistical differences are general not taken into account. The foreground are usually modeled with Gaussian distributions.

In this work our aim will be to use the foregrounds non Gaussian information to extract the CMB signal from data, using a Bayesian approach in attempt to improve over current methods.

In Chapter 2 we introduce some statistics necessary to understand the main results of the CMB. Then there is also some discussion on Gaussian and Non-Gaussian distributions.

In Chapter 3 we briefly review the physics behind the CMB temperature and polarizations anisotropies as well the foregrounds.

Chapter 4 is devoted to the component analysis in signal processing problems and then we present the theoretical core of this work, with all the calculations and results.

Chapter 5 is more practical: we do a simulation with a toy model to see if our method has some improvement with respect to other existing ones.

Finally Chapter 6 is dedicated to our conclusions.

2

A STATISTICAL APPROACH IN COSMOLOGY

According to the theory of Inflation, primordial quantum fluctuations left imprints in the structure of the Cosmic Microwave Background and in the Distribution of Galaxies. Due to the stochastic nature of quantum processes, our Universe is just one of all the possible Universes that could have arisen from Inflation.

Cosmological theories allow us to make statistical predictions about the cosmological (random) fields of the Universe, but not about the details of our particular realization[22]. For example we don't know that a given galaxy will form at a specific 'point' in space and time or that a specific patch of the Cosmic Microwave Background will have some fixed temperature[26]. But we can predict average statistical properties of the cosmological fields, observing only a particular realization of them.

In the next sections we will acquire some tools and language to treat properly cosmological problems.

2.1 PILLS OF PROBABILITY

There are two school of thoughts that define probability:

- Frequentists: probability is the limiting value of the number of successes in a sequence of trials
- Bayesians: probability is a measure of the degree of belief in a proposition based on our available knowledge

We will take the second as our definition of probability, trying to quantify our uncertainty about the world.

Let $A, B, C, ...$ denote propositions(e.g. that it will be sunny tomorrow). Let Ω be a sample space of the possible outcomes of an experiment(e.g. it will be sunny, it will rain, etc...).

Some notation :

- $\bar{A}$ denotes the proposition that A is false
- I is the relevant background information at hand, i.e. any relevant information that is assumed to be true
- $prob(A|B, I)$ is the probability of A happening given that B(and I) has happened
- $prob(A, B|I) = prob(B, A|I)$ is the (joint) probability that both A and B are true (given I).

The two basic rules of probability theory are:

The sum rule:

$$prob(A|I) + prob(\bar{A}|I) = 1$$

and the product rule:

$$prob(A, B|I) = prob(A|B, I) \times prob(B|I) \tag{2.1}$$

From the product rule it is possible to derive Bayes' theorem

$$prob(A|B, I) = \frac{prob(B|A, I) \times prob(A|I)}{prob(B|I)} \tag{2.2}$$

that will be fundamental to the estimation of signals treated in the next chapters.

Finally we introduce the concept of marginalization. The probability that A is true, irrespective of the results of B, is given by[1]

$$prob(A|I) = \int p(A, B)dB \tag{2.3}$$

2.2 RANDOM FIELDS

2.2.1 *Random variables*

Let Ω be a sample space of the possible outcomes of an experiment. A (continuous) random variable $X : \Omega \to \mathbb{R}$ is a function that assigns a number for each element of Ω. We will assign a probability density function (pdf) to X, defined over the space of real numbers:

$$p_X(X = x)dx \equiv p(x)dx \tag{2.4}$$

with dx the Lebesgue measure on the space $\mathbb{R}$[2].

A good probability density has the two following conditions:

- $p(x) \geq 0$
- $\int p(x)dx = 1$

The pdf gives the probability $p(x)\delta x$ to find $x < X \leq x + \delta x$ as $\delta x \to 0$.

1 There is also a discrete version of it. But because we will deal only with continuous variables the one with the integration is given.

2 In general, $\mathbb{R}$ could be replaced by a measurable set of values $\mathcal{A}$(e.g. $\mathbb{R}^n$).

2.2.1.1 *Some definition*

Any expectation value of some function $f(X)$ defined over $\mathcal{A} = \mathbb{R}$ can be obtained directly from the pdf

$$\langle f(X) \rangle = \mathbb{E}[f(X)] = \int f(x)p(x)dx \ . \tag{2.5}$$

Let's take two random variables X_1 and X_2 , with set values $\mathcal{A}_1$ and $\mathcal{A}_2$ respectively. We can think of (X_1, X_2) as a random variable taking values in (a subset of) $\mathcal{A}_1 \times \mathcal{A}_2$. The distribution of (X_1, X_2) is called the joint probability distribution $p_{X_1X_2}(x_1, x_2)$, while the distributions of X_1 and of X_2 are the marginal distributions. For example, the marginal distribution of X_2 is

$$p_{X_2}(x_2) = \int p_{X_1X_2}(x_1, x_2)dx_1 \ . \tag{2.6}$$

These definitions can be generalized to several random variables.

We also introduce the conditional probability density for X_1, which is computed by considering only the cases where X_2 takes the fixed value x_2,

$$p_{X_1|X_2}(x_1|x_2) = \frac{p_{X_1X_2}(x_1, x_2)}{\int p_{X_1X_2}(x_1, x_2)dx_1} \ . \tag{2.7}$$

We say that $X_1$and X_2 are statistically independent if

$$p_{X_1|X_2}(x_1|x_2) = p_{X_1}(x_1) \ . \tag{2.8}$$

And their joint pdf is simply $p_{X_1X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$.

2.2.1.2 *Moments and Cumulants*

The characteristic function of the random variable X is defined as

$$\chi_X(u) = \langle e^{iuX} \rangle = \int p_X(x)e^{iux}dx \tag{2.9}$$

which is simply the Fourier transform of the pdf. From this, we can define the $n-$th moment μ_n of X, when it exists, by

$$\mu_n = \langle X^n \rangle = \frac{1}{i^n} \frac{d^n \chi_X(u)}{du^n}|_{u=0} \tag{2.10}$$

The $n-$th cumulants of X are instead

$$\kappa_n = \frac{1}{i^n} \frac{d^n ln \chi_X(u)}{du^n}|_{u=0} \tag{2.11}$$

If the $n-$th moments and $n-$th cumulants are defined for each n then it is possible to perform a McLaurin expansion of χ_X and $ln\chi_X$

$$\chi_X(u) = 1 + \sum_{n=1}^{\infty} \frac{(iu)^n}{n!}\mu_n \tag{2.12}$$

$$ln\chi_X(u) = \sum_{n=1}^{\infty} \frac{(iu)^n}{n!}\kappa_n \tag{2.13}$$

The first n cumulants can be related to the first n moments (and viceversa). Here there are the relations up to order 6[18]:

$$\kappa_1 = \mu_1 \ , \ \mu_1 = \kappa_1 \tag{2.14}$$

$$\kappa_2 = \mu_2 \ , \ \mu_2 = \kappa_2 \tag{2.15}$$

$$\kappa_3 = \mu_3 \ , \ \mu_3 = \kappa_3 \tag{2.16}$$

$$\kappa_4 = \mu_4 - 3\mu_2^2 \ , \ \mu_4 = \kappa_4 + 3\kappa_2^2 \tag{2.17}$$

$$\kappa_5 = \mu_5 - 10\mu_3\mu_2 \ , \ \mu_5 = \kappa_5 + 10\kappa_3\kappa_2 \tag{2.18}$$

$$\kappa_6 = \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3 \ , \ \mu_6 = \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3 \tag{2.19}$$

μ_1 is the mean of X. Usually κ_2 is denoted as σ^2, the variance. On the other hand, the skewness is defined as $\kappa_3/\kappa_2^{3/2} = \kappa_3/\sigma^3$ and the (excess) kurtosis is κ_4/σ^4.

The probability density function can be obtained from the characteristic function with an inverse Fourier transform. The characteristic function can be determined completely by its cumulants or moments. So the same it is true for the pdf. In cosmology, when we try to study a certain distribution, in practice we measure only a finite number of moments and truncate the expansion. This is an important fact to be kept in mind for section 1.5.

EXAMPLE: THE 1-D GAUSSIAN DISTRIBUTION This is

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{2.20}$$

where μ and σ^2 are two parameters. Their meaning can be deduced by calculating the cumulants:

$$\kappa_1 = \mu \tag{2.21}$$

$$\kappa_2 = \sigma^2 \tag{2.22}$$

$$\kappa_n = 0 \ \forall n \geq 3 \tag{2.23}$$

Equation (1.23) is really important. It tells us that if a random variable distribution is Gaussian then all cumulants of higher order than three are zero. If this doesn't happen we have a non-Gaussian pdf.

On the other hand, the moments of higher order are non-zero for both Gaussian and non-Gaussian distributions, but for the former case they are specified by the first two cumulants ((see (1.14) and following)).

2.2.2 *Random Fields*

A $n-$dimensional random field $q(\vec{x})$ is a collection of random variables, indexed by a continuous (real) space $\mathcal{M}$ of dimension n, characterised by a probability functional $\mathcal{P}[\hat{q}(\vec{x})]$, linked to the probability for the occurrence of a particular field configuration of the field (realization).

A realization of the field, $\hat{q}(\vec{x})$ is a deterministic function of the position $\vec{x}$[3], that represents one of the possible outcomes of the random field[22].

We will take the q's from the space of square integrable functions defined at each point of $\mathcal{M}$, with a scalar product and a norm.

The probability density functional has the following three properties[Sabino Matarrese-unpublished notes]:

1. it is semi-positive definite
2. it can be normalized to unity[4]
3. it goes to zero as $q \to \pm\infty$

3 And eventually of a time t that we omit for clarity because we will not need it here.

4 $\int D[q(\vec{x})]\mathcal{P}[\hat{q}(\vec{x})] = 1$, where the integral is a path integral over field configurations.

The statistical (ensemble) average[5] of some functional $\mathcal{F}[q]$ can be obtained as[6] [Sabino Matarrese-unpublished notes]:

$$\langle \mathcal{F}[q] \rangle \equiv \int \mathcal{D}[q] \mathcal{F}[q] \mathcal{P}[q] \tag{2.24}$$

The meaning of the ensemble average is that we consider several, many realizations (i.e. similarly prepared systems) beloning to a large set, the ensemble. We measure what it is inside the bracket for each realization(e.g.Universe) and then averge it over many realizations(Universes).

2.2.2.1 *The n−point correlation functions*

Given the three conditions above, we can introduce a generating functional[7] (also known as partition function)

$$\mathcal{Z}[J(\vec{x})] = \int \mathcal{D}[q] \mathcal{P}[q] exp[i \int d\vec{x} J(\vec{x}) q(\vec{x})] \tag{2.25}$$

and define the $N-$point disconnected correlation functions $D^{(N)}$ as the coefficients of its McLaurin expansion [23]:

$$\mathcal{Z}[J(\vec{x})] = 1 + \sum_{N=1}^{\infty} \frac{i^N}{N!} \int d\vec{x}_1 ... \int d\vec{x}_N D^{(N)}(\vec{x}_1, .., \vec{x}_N) J(\vec{x}_1) ... J(\vec{x}_N) \quad . \tag{2.26}$$

Or in other words by functional differentiation

$$D^{(N)}(\vec{x}_1, .., \vec{x}_n) = (-i)^N \frac{\delta^N \mathcal{Z}[J]}{\delta J(\vec{x}_1) ... \delta J(\vec{x}_N)} |_{J=0}$$

But this is

$$\int D[q] \mathcal{P}[q] q(\vec{x}_1) ... q(\vec{x}_n) = (-i)^N \frac{\delta^N \mathcal{Z}[J]}{\delta J(\vec{x}_1) ... \delta J(\vec{x}_N)} |_{J=0}$$

In the end

$$D^{(N)}(\vec{x}_1, .., \vec{x}_n) = \int D[q] \mathcal{P}[q] q(\vec{x}_1) ... q(\vec{x}_n) \equiv \langle q(\vec{x}_1) ... q(\vec{x}_n) \rangle \tag{2.27}$$

5 Denoted by the bracket.

6 Formally inside each argument of the integral it should be present $\hat{q}(\vec{x})$. We omit for brevity.

7 See analogy with the characteristic function of random variables

Similarly, we can define the connected correlation functions as the coefficients of the generating functional of (connected) correlation functions

$$ln\mathcal{Z}[J(\vec{x})] = \sum_{n=1}^{\infty} \frac{i^n}{n!} \int d\vec{x}_1 ... \int d\vec{x}_n C^{(n)}(\vec{x}_1, .., \vec{x}_n) J(\vec{x}_1) ... J(\vec{x}_n) \quad (2.28)$$

denoted also as[8]

$$\langle q(\vec{x}_1), .., q(\vec{x}_n) \rangle_c = C^{(n)}(\vec{x}_1, .., \vec{x}_n) \quad (2.29)$$

As a pdf of a random variable is fully specified by its moments or cumulants, a random field with a probability distribution is specified by its correlation functions(connected or disconnected).

2.2.3 *Statistical Homogeneity and Isotropy in Cosmology*

We have developed some theoretical tools to compute quantities averaged over the ensemble of all possible Universes of our 'true' Universe. Practically these averages are impossible to do since our observations can't probe entirely the single realization in which we live but only a part of it. But we would like to have information about the whole Universe and it seems that this is basically impossible. Then, how can we link data to the underlying theory?

A fundamental assumption in Cosmology is the statistical cosmological principle: on large scales[9] (statistically speaking) the Universe is homogeneous and isotropic. What does it mean?

A random field $q(\vec{x})$ is statistically homogeneous if $\forall n$ the joint multipoint probability distribution functions $P(q_1, q_2, ..., q_n)$[10] are invariant under translation (by the same vector) of $\vec{x}_1, \vec{x}_2, ..., \vec{x}_n$[3]. For example this implies that $\xi = \langle q(\vec{x}_1) q(\vec{x}_2) \rangle$ depends only on the relative positions.

From this assumption follows an important fact. The connected correlation functions $\langle q(\vec{x}_1), .., q(\vec{x}_n) \rangle_c$ vanish if at least two points belong to casually disconnected regions of the Universe[11][22]. Therefore, if we take widely separated parts of the Universe, each one has its own set of connected correlation functions. Assuming statistical homogeneity these will coincide with those of any other region. This means that different regions of our Universe has the same statistical properties (e.g. same averages). This is closely related to the fair sample hypothesis: well separated regions of the Universe are indepen-

8 Note the commas to indicate that we are not doing averages.

9 For us those are the scales much greater than the size of gravitationally collapsed structures[http://www.astro.ufl.edu/~vicki/galaxies/cosmologynotes.pdf]

10 Shorthand notation $q_i = q(\vec{x}_i)$

11 That exist due to the finite speed of light.

dent realizations of the same physical process (they are realizations inside the realization) and in the observable part of the Universe there are enough independent samples to be representative of the statistical ensemble(of all possible Universes)[26]. Then, taking spatial averages over many regions is the same as taking expectations over the ensemble.

The random field $q(\vec{x})$ is statistically isotropic if there is invariance of $P(q_1, q_2, ..., q_n)$ under global spatial rotations($\vec{x} \to R\vec{x}$). If we add this to the previous example this means that ξ will depend only on the modulus of the relative positions[12].

2.2.3.1 *A simplification*

When we do practical observations, we produce pixelised maps with a finite number of pixels. So that the index space $\mathcal{M}$ gives us a finite number of indices. From now on, for cosmological applications we will consider a random field as a (large) random vector $\vec{X} = (X_1, ..., X_D)^T$ with dimension D.

2.3 RANDOM FIELDS ON A SPHERE

Our focus will be on the random field $f = T$ of the Cosmic Microwave Background temperature. When we take observations of it from the sky, we only sample a projection on the celestial sphere at the time of last scattering of the photons. In other words we deal with 2 dimensional maps. In cosmology, it is quite popular a convenient spherical decomposition to study a 2-D random field that simplifies a lot calculations and the interpretation of results.

The idea is to build our theory on the spherical analog of the Fourier transform, the spherical harmonics.

As said, we will consider the space of square integrable functions. The spherical harmonics form a basis for this space and allow us to express f through a modal expansion (the index is $\hat{n} = (\theta, \phi)$)

$$f(\theta, \phi) = f(\hat{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=+l} a_{lm} Y_{lm}(\theta, \phi) \quad . \tag{2.30}$$

This resembles the $2D$ Fourier transform

$$f(x) = \sum_{n_x=0}^{\infty} \sum_{n_y=0}^{\infty} A_{n_x n_y} e^{i n_x k x} e^{i n_y k y}$$

12 A random field which is homegenous and isotropic is generally called stationary[Cosmological Perturbations].

The Y_{lm}'s are complete and as the Fourier transform, respect an orthonormality relation on the sphere

$$\int_{-\pi}^{\pi}\int_{0}^{\pi} Y_{lm}(\theta,\phi)Y_{l'm'}(\theta,\phi)sin\theta d\theta d\phi = \int d\Omega Y_{lm}(\hat{n})Y_{l'm'}(\hat{n}) = \delta_{ll'}\delta_{mm'} \tag{2.31}$$

From here, and from completeness, the (complex) coefficients of the expansion (27) are obtained by

$$a_{lm} = \int d\Omega f(\hat{n})Y^{*}_{l'm'}(\hat{n}) = \int_{-\pi}^{\pi}\int_{0}^{\pi} d\theta d\phi sin\theta f(\theta,\phi)Y_{l'm'}(\theta,\phi) \tag{2.32}$$

2.3.1 *Meaning of l and m*

The index l is an integer associated with the number of spatial oscillations (nodes) in the θ direction in spherical coordinates, and the integer m in the ϕ direction ($|m| \leq l$).

2.3.1.1 *Example*

- $l = 0$ has zero oscillation in the θ direction, it's a monopole, constant over the whole sphere.
- $l = 1$ has one full oscillation over the sphere in the θ direction, it's a dipole.
- Higher l values gives us more spatial oscillations over the sphere surface and therefore smaller wavelength(or higher frequency).

For more insights see [15].

2.3.2 *Angular Correlation Functions*

The angular $n-$point correlation function is

$$\langle f(\hat{n}_1)f(\hat{n}_2)...f(\hat{n}_n)\rangle \tag{2.33}$$

used to characterize a cluestering pattern of fluctations on the sky, $f(\hat{n})$[16]. One disadvantage of the angular correlation function is that data points of the correlation function at different angular scales are generally not independent of each other, but correlated[16]. Thus, it is difficult to interperate the data and do a statistical analysis. Using spherical harmonics it is possible to simplify more the problem and what we have to consider is the angular $n-$point harmonic spectrum,

$$\langle a_{l_1m_1}a_{l_2m_2}...a_{l_nm_n}\rangle \tag{2.34}$$

In cosmology the most used harmonic spectra are the $2-$, $3-$ and $4-$ point ones, called the power spectrum, the bispectrum, and trispectrum, respectively.

For the most important distribution used in Cosmology, the Gaussian one, the angular spectra at different angular scales, or at different $l's$, are uncorrelated[PhD Komatsu]. Moreover, orthogonality of the spherical harmonics for different ls, highlights characteristic sturctures on the sky at a given l.

As said, in reality, we cannot measure ensamble averages, but only single realizations. So we measure a single $a_{l_1m_1}a_{l_2m_2...}a_{l_nm_n}$, which is noisy and we want to average it somehow to reduce the noise. By statistical isotropy we are justified to average (with weights) the spectrum over m_i, reducing the statistical error of the measured harmonic spectra. It is possible to find the appropriate weights for any harmonic spectrum of a given order[16]. In the next lines we will see only the expression for the power spectrum and bispectrum for relevant cases.

2.3.2.1 *Power Spectrum*

The angular power spectrum measures how much fluctuations exist on a given angular scale. For example, the variance of a_{lm} for $l \geq 1$, $\langle |a_{alm}|^2 \rangle$, measure the amplitude of fluctuations at a given l.

In general, the covariance matrix

$$C_{lml'm'} = \langle a_{lm} a_{l'm'} \rangle \tag{2.35}$$

is not necessarily diagonal (fact that simplifies a lot calculations). But, diagonality is achieved when we assum full sky coverage and rotational invariance of (35). It is possible then to show[PhD Komatsu] that

$$C_{lml'm'} = \langle a_{lm} a_{l'm'} \rangle = \langle C_l \rangle \delta_{ll'} \delta_{mm'} \tag{2.36}$$

2.3.2.2 *Bispectrum*

The bispectrum

$$B^{l_1l_2l_3}_{m_1m_2m_3} \equiv \langle a_{l_1m_1} a_{l_2m_2} a_{l_3m_3} \rangle \quad . \tag{2.37}$$

has the following form when we assum statistical isotropy

$$B^{l_1l_2l_3}_{m_1m_2m_3} = \mathcal{G}^{l_1l_2l_3}_{m_1m_2m_3} b_{l_1l_2l_3} \tag{2.38}$$

where $b_{l_1l_2l_3}$ is a real symmetric function of l_1, l_2, l_3 called the reduced bispectrum. We used also the Gaunt integral

$$\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} \equiv \int d\Omega Y_{l_1 m_1}(\vec{e}_1) Y_{l_2 m_2}(\vec{e}_2) Y_{l_3 m_3}(\vec{e}_3) =$$

$$= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{2.39}$$

where $\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ is the $3j$ Wigner symbol. The Gaunt integral must obey three conditions to be non null:

- $m_1 + m_2 + m_3 = 0$(triangular closure condition)
- $l_1 + l_2 + l_3 = 2n,\ n \in \mathbb{N}$
- $|l_i - l_j| \leq l_k \leq l_i + l_j$ for every cyclic permutation of

In practice, beacuse we have only one realization of the Universe, the bispectrum is not observable. An estimator of the bispectrum is

$$\hat{B}_{l_1 l_2 l_3} = \sum_{m_1, m_2, m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B^{l_1 l_2 l_3}_{m_1 m_2 m_3} =$$

$$= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} b_{l_1 l_2 l_3} \tag{2.40}$$

where for each m_i the sum runs from $-l_i$ to $+l_i$.

In important theoretical works it was possible to obtain analytically the form of the bispectrum with a parameter model dependence, making possible a fit approach for testing them [20].

2.4 GAUSSIAN DISTRIBUTION

The Gaussian distribution, also know as Normal distribution, for a real random vector $\vec{X}$ is

$$\mathcal{N}(\vec{X}|\vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp[-\frac{1}{2}(\vec{X} - \vec{\mu})^T \Sigma^{-1} (\vec{X} - \vec{\mu})] \tag{2.41}$$

where Σ is a $D \times D$ covariance matrix, $|\Sigma|$ its determinant, $\vec{\mu}$ is a $D-$dimensional vector (mean) and ^{T} is the operation of transposition.

Without loss of generality we can take the covariance matrix symmetric, because any asymmetric component would disappear from the exponential[13].

Defined in this way, the Normal distribution respects the two requirements for good probability distributions:

- $\mathcal{N}(\vec{X}|\vec{\mu},\Sigma) \geq 0$
- $\int \mathcal{N}(\vec{X}|\vec{\mu},\Sigma)d\vec{X} = 1$

Other properties of the Gaussian distribution:

- $\mathbb{E}[\vec{X}] = \vec{\mu}$, so $\vec{\mu}$ is the mean of the distribution
- $\mathbb{E}[\vec{X}\vec{X}^T] = \vec{\mu}\vec{\mu}^T + \Sigma$ which groups all the second order moments
- $cov[\vec{X}] = \mathbb{E}[(\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})^T] = \Sigma$
- If the covariance matrix is diagonal, the components of $\vec{X}$ are independent. In other words, in the Gaussian case uncorrelation means independence[14]. A particular case of this is when $\Sigma = \sigma\mathbb{I}_{D\times D}$, known as isotropic covariance. If the surfaces of constant density were plotted we would obtain spherical surfaces.

There is a very remarkable result that makes Gaussian distributions very special.

Wick's/Isserlis' Theorem : For a Gaussian random field q the following relation holds ($q_i = q(\vec{x}_i)$)

$$\langle q_{i_1} q_{i_2} .... q_{i_n} \rangle = \sum_I \langle q_{j_1} q_{k_1} \rangle .... \langle q_{j_{n/2}} q_{k_{n/2}} \rangle \tag{2.42}$$

where the sum is over all possible pairings of $i_1, ..., i_n$into pairs $(j_1, k_1), ..., (j_{n/2}, k_{n/2})$ (n is even)[15].

So, the odd $N-$point (disconnected) correlation functions of a Gaussian are zero while the even ones are all derivable from the two-point one, in the sense that correlation functions are decomposable as products of the $2-$point correlation function, thus determining the statistical properties of the random field[16].

The Gaussian distribution arises in many physical applications. One of the most important is when we consider the sum of multiple identically distributed and independent random fields(each one with its

13 Furthermore, for the Gaussian distribution to be well defined, it is necessary that Σ is also positive definite[Bishop].

14 Remember that independence means uncorrelation, but the converse is not true for general non-normal distributions .

15 For a proof see The Cosmic Microwave Background by Ruth Durrer

16 One of the most important examples is the Cosmic Microwave Background two-point function.

distribution). In this case the central limit theorem applys: as the number of fields increases, their sum, which itself is a random field, has a distribution that becomes increasingly Gaussian[4]. This is connected for example with cosmology: in linear perturbation theory where we have a large number of cosmological fluctuations evolving independently, we can expect, based on the central limit theorem, that the Universe will obey a Gaussian distribution[1]. In the next chapter we will see also the CMB example.

2.5 NON-GAUSSIANITY

If the fluctuation is Gaussian, then the two-point correlation function is what we all need to specifiy the Gaussian distribution. On the other hand, if we need higher-order correlation functions to determine the statistical properties of our distribution we are dealing with a Non-Gaussian pdf.

How do we describe a non-Gaussianity? There is only one way for Gaussianity but an infinite number of ways of being non-Gaussian.

Any deviation from Gaussianity can be quantified using the three-point and higher order correlation functions. In Cosmology the most popular approach is to use the harmonic tranform of the three-point function, the bispectrum. This is because usually we will deal with distributions that are close to a Normal one.

2.5.1 *Nearly Gaussian Distributions*

As said, in cosmology it is common to deal with fields that depart only weakly from a Normal distribution [3]. In many applications, to extract useful information on the underlying physical processes, it is more interesting to measure the deviations of a probability density function (pdf) from the Normal distribution than to prove that it is close to the Gaussian one. For example, the three-point function, also known as the bispectrum, is a sensitive test for a non-Gaussian contribution to the fluctuation spectrum since it is precisely zero in the Gaussian limit. It is possible to find a number of approximations to express the local nearly gaussian pdf in another ways[5]. Here we will deal with one approximation of quasi-Gaussian distributions using the EdgeWorth Expansion.

The EdgeWorth expansion expresses one probability density in terms of another by the use of higher-order statistics.

THE 1-PDF EDGEWORTH EXPANSION Le'ts consider a random variable X with a probability density function $f(X = x) = f(x)$ and $Z(x)$ a Gaussian density function, both with the same mean and variance. If the cumulants κ_i^f and κ_i^Z, of $f(x)$ and $Z(x)$ respectively, become

'close' sufficiently quickly as i increases [17]then we say that $f(x)$ is close to Gaussian. It seems natural then to use the expansion

$$f(x) = \sum_{n=0}^{\infty} c_n \frac{d^n Z(x)}{dx^n} \tag{2.43}$$

We now investigate a family of orthogonal polynomials which form a natural basis for the expansion of $f(x)$. We say that two polynomials $P_n(x)$ and $Q_n(x)$ of degress $n \neq m$ are orthogonal on the real axis with respect to a weight function $w(x)$ if

$$\int_{-\infty}^{+\infty} P_n(x) Q_m(x) w(x) dx = 0, \quad n \neq m$$

For example a weight function could be $w(x) = w_{CH}(x) = exp(-x^2/2)$ proportional to the Gaussian function $Z(x) = \frac{exp(-x^2/2)}{\sqrt{2\pi}}$. In this case the associated polynomials are given by the Rodrigues' formula

$$He_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \tag{2.44}$$

These are called Chebyshev-Hermite polynomials. Examples are

$$He_0(x) = 1$$

$$He_1(x) = x$$

$$He_2(x) = x^2 - 1$$

$$He_3(x) = x^3 - 3x$$

Instead of w_{CH} we could use $w_H(x) = exp\{-x^2\}$ proportional to $Z^2(x)$. In this case we obtain the Hermite polynomials[18]

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \tag{2.45}$$

From the definition of $Z(x)$ and $He_n(x)$ we have that

$$\frac{d^n Z(x)}{dx^n} = (-1)^n He_n(x) Z(x) \tag{2.46}$$

17 http://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat05844/abstract

18 Chebyshev-Hermite polynomials are also called the probabilists' Hermite functions, while Hermite polynomials the physicists' Hermite function. The two are linked by a rescaling. See Wikipedia.

Plugging this last equation in the expansion of $f(x)$ we obtain that

$$f(x) = (-1)^n \sum_{n=0}^{\infty} c_n He_n(x) Z(x) \tag{2.47}$$

To obtain the coefficients we simply use the definition of orthogonal polynomials[19], obtaining

$$c_n = \frac{(-1)^n}{n!} \int_{-\infty}^{+\infty} f(t) He_n(t) dt \tag{2.48}$$

With this coefficients we are left with what it is called the Gram-Charlier expansion of type A [5]. Expressing the polynomial $He_n(t)$ as $\sum_{k=0}^{n} a_k t^k$ stands out the fact that the coefficients c_n are a linear combination of the moments α_k of the random variable X. To find the full formula it is possible to show[5]that

$$He_n(x) = n! \sum_{n=0}^{[n/2]} \frac{(-1)^k x^{n-2k}}{k!(n-2k)!2^k}$$

The problem of the Gram-Charlier series is that it has poor convergence properties. For realistic cases, very often it diverges. This is linked to the sensitivity of the Gram-Charlier to the behaviour of $f(x)$ at infinity: the latter must fall to zero faster than $exp(-x^2/4)$ for the series to converge. See again[5]also for a real example.

We can normalize our random variable X to unit variance by diving by its standard deviation σ.Let's consider as expansion parameter

$$S_n = \frac{\kappa_n}{\sigma^{2n-2}} \tag{2.49}$$

We denote with k_m the set of all non-negative integers satisfing $k_1 + 2k_2 + ... + nk_n = n$ and $r = k_1 + k_2 + ... + k_n$. The Edgeworth expansion at arbitrary order of a probability density $g(x)$ with respect to the normalized unit variance probability Gaussian function $\phi(x)$ is obtained by this

Result: Given a nearly Gaussian pdf $f(x)$ this can be written as a perturbation of a Gaussian by the asymptotic Edgeworth expansion for arbitrary order s

$$g(x) = \sigma f(\sigma x) = Z(x)\{1 + \sum_{s=1}^{\infty} \sigma^s \times \sum_{\{k_m\}} He_{s+2r}(x) \prod_{m=1}^{s} \frac{1}{k_m!} (\frac{S_{m+2}}{(m+2)!})^{k_m}\} \tag{2.50}$$

19 Remeber that for a polynomial of order n, its $n-$th derivative is proportional to $n!$

where by asymptotic we mean that if the first N terms are retained in the sum over s, then $g(x)$ minus the partial sum is of a lower order than the $N-th$ term in the sum. See [Nearly] for a derivation and for how to do the sum. In the standard literature we will find only the few first terms written. In this thesis for example we will only need the expansion to S_3. For future reference we write here some terms of the expansion[20]:

$$f(y) = \frac{1}{\sigma} Z(\frac{y}{\sigma})[1 + \sigma \frac{S_3}{6} He_3(\frac{y}{\sigma}) + \sigma^2 \frac{S_4}{24} He_4(\frac{y}{\sigma}) + ...] =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} [1 + \sigma \frac{S_3}{6} He_3(\frac{y}{\sigma}) + \sigma^2 \frac{S_4}{24} He_4(\frac{y}{\sigma}) + ...] \tag{2.51}$$

For a simple application of this, see [17].

ATTENTION: A well-known problem with truncating the moments(or cumulants) expansion, is that the resulting distribution is not a well-defined probability distribution function. Hence The Edgeworth series, although properly normalised, can produce negative values if we use big values of the expansion parameters(e.g. skewness). This is a breakdown of the series approximation, and one should take care not to force the PDF into this regime. As a rule of thumb we will take $S_n << 1$.

THE MULTIVARIATE EDGEWORTH EXPANSION We rewrite here the form for a Gaussian discretized random field $\Phi(\vec{x})$ [21]with zero mean and covariance matrix Σ

$$p(\Phi|\Sigma) = \frac{exp[-\Phi^T \Sigma^{-1} \Phi/2]}{\sqrt{(2\pi)^D |\Sigma|}} \tag{2.52}$$

where $\Phi^T \Sigma^{-1} \Phi = \sum_{ij} \Phi(x_i)(\Sigma)_{ij} \Phi(x_j)$.

An example could be the temperature anisotropy for a given direction in the sky, $\Phi(\vec{x}) = \delta T(\hat{n})$[22]. In this case it is more convenient to use the spherical harmonic transform

$$\Phi(\vec{x}) = \delta T(\hat{n}) = \sum_{lm} a_{lm} Y_{lm}(\hat{n})$$

20 Note we do the change of variable $y = \sigma x$

21 Whose components, $\Phi_i = \Phi(\vec{x}_i)$, may represent, for example, an array of pixel values or the amplitude of a set of harmonic modes

22 In general we should also consider the observer position $\delta T(\vec{x}, \hat{n})$.

or[23]

$$a_{lm} = \int d\Omega \delta T(\hat{n}) Y^*_{lm}(\hat{n})$$

Then we could write the Gaussian pdf as ($C = \Sigma$ and $C^{-}1$ its inverse)

$$p(\vec{\Phi}) = p(\{a_{lm}\}) = \frac{1}{(2\pi)^D |C|^{1/2}} exp[-\frac{1}{2}\sum_{lm} a^*_{lm}(C^{-1})_{lm,l'm'}a_{l'm'}] \quad (2.53)$$

with $D = N_{harm}$ the number of harmonics and $C_{lm,l'm'} = \langle a^*_{lm} a_{l'm'} \rangle$. In general C have a non diagonal form[24].

What happens if we want to consider Non-Gaussianities for a multivariate case (in the space of harmonics)? We could again use, as with the $1-D$ case, the fact that measuring higher order (than two) correlation functions allows us to retrace a pdf. The problem is that in general practically we are only able to measure three or four points correlation functions (angular bispectrum $\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle$ and trispectrum $\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} a_{l_4 m_4} \rangle$). But what can we do is to suppose to have a weak non-gaussianity and so we are justified to use a multidimensional Edgeworth expansion.

We can write an expression for the Edgeworth expansion of a multivariate PDF in the harmonic coefficients[27]

$$p(\{a_{lm}\}) = p(\vec{a}) = \frac{1}{(2\pi)^D |C|^{1/2}} exp[-\frac{1}{2}\sum_{lm} a^*_{lm}(C^{-1})_{lm,l'm'}a_{l'm'}] \times$$

$$\times \{1 + \frac{1}{6}\sum_{all\ lm} \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle [(C^{-1}a)_{l_1 m_1} (C^{-1}a)_{l_2 m_2} (C^{-1}a)_{l_3 m_3}] +$$

$$-\frac{1}{6}\sum_{all\ lm} \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle [3(C^{-1})_{l_1 m_1, l_2 m_2} (C^{-1}a)_{l_3 m_3}] + ...\} \quad (2.54)$$

But if we consider only Cosmic Microwave Background temperature anisotropy(as later will be the case), then we can apply rotational invariance to obtain:

$$C_{lm,l'm'} = C_l \delta_{ll'} \delta_{mm'} \equiv C_l \quad (2.55)$$

23 $d\Omega$ is the volume element $d\hat{n}$.

24 For example if we measure a mixture of emissions of astrophysical components.

and the Edgeworth expansion is [2]

$$p(\vec{a}) = [1 - \sum_{all\ lm} \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle \frac{\partial}{\partial a_{l_1 m_1}} \frac{\partial}{\partial a_{l_2 m_2}} \frac{\partial}{\partial a_{l_3 m_3}}] \prod_{lm} \frac{e^{-\frac{a^*_{lm} a_{lm}}{2C_l}}}{\sqrt{2\pi C_l}} \quad (2.56)$$

3

THE MICROWAVE EMISSIONS IN THE SKY

In Figure 1.1 there is a temperature fluctuations map for the Cosmic Microwave Background. This is used for example to obtain the cosmological parameters necessary to test our standard model of cosmology (ΛCDM). Ideally, this map includes only the CMB signal. But in the reality, between the Last Scattering Surface and us (the observer) there are other sources of mircowave radiation that contaminate the CMB signal with what is called foreground signal. If we imagine an observer in the outer space, these contaminations will be mainly from our own Galaxy (over a variety of angular scales) and other extragalactic sources (on small scales).

There is a long map producing process that includes removal of foregrounds. It is clear that if our map is not reliable, i.e. it includes a proper cleaning from the foregrounds, we will obtain wrong estimations of the cosmological parameters, being their calculation our final goal (to check our cosmological model). Therefore, being interested in the CMB signal , it is necessary to understand and remove foreground signals from our maps for a correct interpretation of CMB data.

In this chapter we will get to know better our signal of interest and the sources of the foregrounds. In the next one we will dive in the cleaning techniques world.

3.1 THE COSMIC MICROWAVE BACKGROUND

The Cosmic Microwave Background (CMB) is a signal that provides a snapshot of our Universe when it was 380000 years old ($z \sim 1100$) . The Universe was a hot plasma of electrons, protons, photons, neutrinos in thermal equilibrium. With the expansion the Universe cooled and eventually protons and electrons formed mainly neutral hydrogen atoms(recombination), thus marking the decoulping of photons from matter(last scattering). From that time, photons have 'free streamed' across the Universe, eventually reaching the Earth.

There are important facts about the CMB [12]:

- It has a near perfect blackbody spectrum. This tells us that in the Early Universe the plasma and the photons were in thermal equilibrium, giving a blackbody nature to the CMB. Then adiabatic cooling from the expansion presereved the thermal

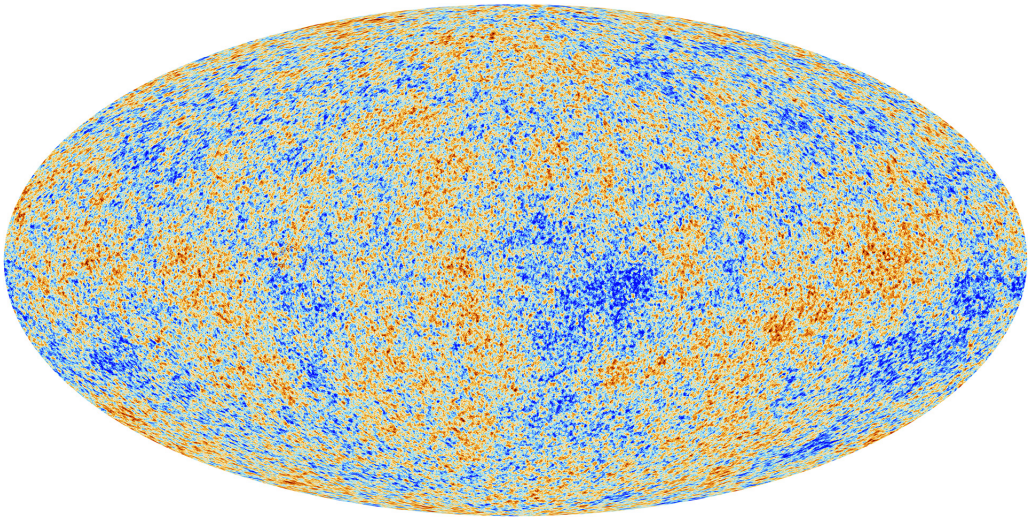


Figure 3.1: CMB Temperature Anisotropies Map (Source:Planck CMB, ESA and the Planck Collaboration)

spectrum. Today we see a thermal blackbody spectrum with parameter $T = 2.72548 \pm 0.00057K$.[1].

- It's uniform and isotropic one part in 10^5 . These tiny anisotropies bear the imprint, filtered through the dynamics and geometry of the expanding universe, of the seeds of large scale structure of the Universe. CMB anisotropies can therefore shed light on not only the mysteries of structure formation but also such fundamental quantities as the expansion rate, matter content and geometry of the universe. These perturbations were the seeds of large scale structure that gave rise to the galaxies[2].

3.1.1 *CMB Temperature Anisotropies*

3.1.1.1 *Description*

The Cosmic Microwave Background (CMB) temperature can be represented as a random field $T(\theta, \phi)$ on a unit sphere S^2, that is, for each $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$, $T(\theta, \phi)$ is a real random variable [20].

$$T(\theta, \phi) = \bar{T} + \delta T(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta, \phi) \tag{3.1}$$

The value $l = 0$ corresponds to the monopole, or the mean value. Because we are interested in anisotropies we omit it. Then there is $l = 1$,the dipole, with an amplitude of order $10^{-3}K$. This is thought to be due to the peculiar motion of the Earth with respect to the CMB

1 To not be interpreted in the thermodynamical sense because we haven't anymore a thermal bath of particles.

2 In according to our theory of structure formation by gravitational instability.

frame. This is the frame in which the dipole is zero(indeed it's a frame dependent quantity [24]).

The part of the temperature fluctuations given by higher multipoles ($l \geq 2$) is classified in primary anisotropies, during the epoch of recombination, and secondary anisotropies, due to the interaction of CMB photons with the rest of the Universe during their travel from the last scattering surface to us.

The primordial, primary, anisotropies were imprinted in CMB photons with the tiny fluctuations, $O(10^{-5})$, in the primordial plasma via gravitational interaction. Because matter and radiation were tightly coupled, this increased the pressure of the photons, that acted as a restoring force producing acoustic oscillations in the plasma. When photons decoupled from matter at recombination, the phases of the oscillations were frozen in, so the more compressed (diluted) regions appear as a hotter (colder) deviation from the mean CMB temperature[28].

Because anisotropies carry so much detail we will focus on $\delta T(\theta, \phi)$, which we will take with zero mean and its covariance will be invariant with respect to the group of rotations(isotropy). Using the spherical harmonic expansion,

$$\delta T(\theta, \phi) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta, \phi) \tag{3.2}$$

where the a_{lm}s ($l = 1, 2, .., ,\ m = -l, ..., 0, ..., +l$) are random coefficients that can be obtained by

$$a_{lm}^T \equiv a_{lm} = \int_{-\pi}^{\pi} \int_{0}^{\pi} \delta T(\theta, \phi) Y_{lm}^*(\theta, \phi) sin\theta d\theta d\phi \quad . \tag{3.3}$$

The index l runs from 1 to infinity. However, in any realistic experiment there is an upper limit L on the multipoles we may observe, depending on the resolution of the experiment and the presence of noise. For example L is on the order of 500/800 for WMAP and 1500/2000 for Planck[20]. The multipole l goes roughly as $l \sim \pi/\theta$ where θ is the angular scale. From this we see that Planck has a much better resolution than WMAP.

If the temperature fluctuations are Gaussian and isotropic[3], then all the statistical information of the anisotropies is contained in the CMB temperature power spectrum[4],

$$C_l^{TT} = \langle a_{lm}^{T*} a_{l'm'}^T \rangle = C_l^{TT} \delta_{mm'} \delta_{ll'} \tag{3.4}$$

3 Many inflation models suggest a Normal distribution for temperature anisotropies and this is in accordance with experimental data.

4 And the multipoles are independent from each other.

where T indicates the temperature. The sequence $\{C_l^{TT}\}$ denotes the angular power spectrum for TT.

Obviously, because we cannot perform ensemble averages this is only a theoretical equation for the power spectrum. Practically, what we use is the unbiased estimator (i.e. as $l \to \infty$ it goes to the theoretical value)

$$\hat{C}_l^{TT} = \frac{1}{2l+1} \sum_{m=-l}^{+l} \langle a_{lm}^{T*} a_{l'm'}^{T} \rangle \tag{3.5}$$

These Cls are distributed with a χ distribution with $2l+1$ degrees of freedom, mean values equal to C_l^{TT}(the theoretical value) and a variance of $2C_l^{TT}/(2l+1)$. So the variance of the variance is

$$\Delta C_l = \sqrt{\frac{2C_l^{TT}}{2l+1}} \tag{3.6}$$

This is a fundamental limit of the determination of C_l^{TT} which cannot be overcomed by instruments[5]. This is called cosmic variance and it is related to the fact that we have only one realization of our true Universe and only $2l+1$ modes to obtain our estimation at the multipole l.

3.1.1.2 *Description*

We understood that we only need the power spectrum for the description of CMB anisotropies. Anisotropies depend on the observed angular scale, both because they stem from different mechanisms and because the details of the same phenomenon can change with scale[6].

In Figure 2.2 we see that CMB power spectrum is characterized by four distinct regions in l.

The horizon scale at last scattering is given by $l \sim 100$. So, anisotropies at larger scales have not evolved that much (for causality), and hence they directly reflect the initial perturbations. At these scales anisotropies are produced by the so called ordinary Sachs-Wolfe effect, due to the effect of the primordial gravitational potential at the surface of last scattering, and the integrated Sachs-Wolfe effect, due to the changes in the gravitational potential that redshift the CMB photons travelling from the last scattering surface to us.

For a scale invariant primordial power spectrum we have a plateau, $l(l+1)C_l \sim const$ at low ls.

The mechanism generating primordial perturbations could generate scalar, vector and tensor modes[11]. As now we described the

5 Furthermore, this uncertainty on the power spectrum in a given multipole becomes worse when we include instrumental noise, finite beam resolution and observation over a finite fraction of the sky[Arxiv 0803.0834.pdf].

effect of the scalar ones. On the other hand, vector modes decay with the expansion of the Universe. Tensor modes have the effect of rising the plateu at low $l \leq 100$. At small scales the tensor spectrum decays by cosmic expansion [6].. When we measure CMB temperature-temperature spectrum we can't discriminate between the contributions of tensor and scalar modes. Nevertheless, the tensor can be detected using polarization information(see Polarization).

At angular scales corresponding to $100 \leq l \leq 1000$ the physical processes that gives temperature anisotropies are causally connected and affect the primordial plasma before recombination. As previously said, in the primordial Universe, we had matter and radiation tightly coupled, behaving as a single fluid. The richness in the anisotropy spectrum is given by the gravity-driven acoustic oscillations occurring before the atoms in the Universe became neutral. At the time of recombination, the modes of acoustic oscillations are frozen at different phases of oscillations, and consequently at different wavelengths, giving peaks in the power spectrum fluctuations. The peaks represent the scales of maximum compression and rarefaction of the plasma. The valleys in between the peaks are non-zero due to Doppler shifting of the emission, as they correspond to the maximum in the velocity of the plasma[28].

The peaks depend on the cosmological parameters used for our model of the Universe, which gives CMB anisotropies their great constraining power[11]. Indeed it is possible to extract the parameters describing comsological models of our Universe. See [8] for an example with Planck experiment.

At $l \geq 1000$ the acoustic oscillations decrease exponentially. We should keep in mind that the coupling between photons and baryons is not perfect. Photons have a non zero mean free path. Furthermore, the recombination process is not instantaneous, which gives a thickness to the last scattering surface. Because of this, we have diffusion of photons which washes out the anisotropies on scales smaller than their mean free path. This decreasing effect, is called Silk damping.

CMB have also secondary anisotropies, anisotropies generated during the photons propagation from the last scattering surface to us (the observer). The physical processes which induce them are also responsible of the structure formation and the evolution of our Universe. These processes include reoinization, graviational lensing, Sunyaev-Zel'dovich effect and others. What we have to bear in mind is that those secondary anisotropies have an effect on the CMB temperature power spectrum and thus if we are interested only on the primary fluctuations these should be removed by some cleaning technique.

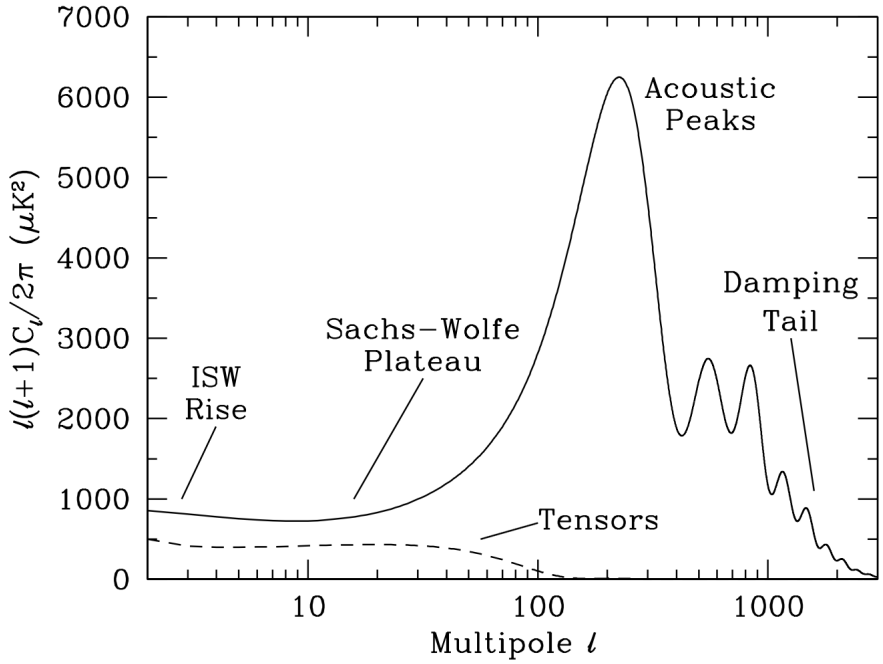


Figure 3.2: Scalar and tensor temperature-temperature power spectra for a standard $\Lambda - CDM$ model[https://arxiv.org/pdf/astro-ph/0406567.pdf].

3.1.2 *CMB Polarization*

We have another probe of the early Universe. CMB carries additional information about the Universe that is encoded in its polarization. A light wave traveling in some $\hat{k}$ direction corresponds to electric and magnetic fields oscillating in the plane perpendicular to the direction of propagation, given by two directions $\hat{x}_1$ and $\hat{x}_2$. If the intensity of the two transverse directions is equal, then the light is unpolarized. Otherwise, the light radiation is polarized[11].

CMB photons are polarized at $\leq 10\%$ level by the Thomson scattering at recombination and reionization epochs[**Ichiki:aa**].

3.1.2.1 *Origin*

Thomson scattering between an electron and unpolarized light with a quadrupolar component ($l = 2$) in its intensity produces a linear polarization in the scattered field. In the primodial epoch this could had happened only a little before recombination[6]. Indeed, at earlier times electrons and photons are tightly coupled to each other, implying a very weak quadrupole. We expect primordial polarization signal to be much smaller than the temperature one.

6 Because we need scattering between light and electrons. After recombination scatterings are very rare(until reionization epoch that enhances the polarization signal at large scales)[Challinor, CMB review].

The CMB temperature anisotropies have three geometrical distinct sources, namely scalar(compressional), vector and tensor modes.

Scalar fluctuations generate velocity gradients in the plasma, as the matter falls and bounces back from the potential wells at the last scattering surface. The velocities of the plasma are out of phase with the density distribution. As we saw the scalar fluctuations are also the origin of most of the temperature anisotropy. Vector fluctuations correspond to vorticities in the plasma at the last scattering surface. But they get damped by expansion(so we won't consider them). Tensor fluctuations compress and stretch the space in perpendicular directions, compressing and relaxing the plasma and generating a quadrupole pattern in the temperature. They are expected to be fluctuations from the primordial gravitational waves predicted by Inflation[28].

3.1.2.2 *Patterns of polarization*

We saw how to describe CMB temperature anisotropies. Now it's time for the CMB polarization anisotropies.

Lets take a general electromagnetic field $\vec{e}$, orthogonal to its direction of propagation $\vec{k}$. We take two basis vectors $\hat{x}$ and $\hat{y}$ orthogonal to $\vec{k}$. All the statistical information is encoded in the 'coherence matrix' C[21]:

$$C = \begin{pmatrix} \langle|\vec{e}_x|^2\rangle & \langle\vec{e}_x\vec{e}_y^*\rangle \\ \langle\vec{e}_y\vec{e}_x^*\rangle & \langle\vec{e}_y|^2\rangle \end{pmatrix} = \frac{1}{2}\begin{pmatrix} I+Q & U-iV \\ U+iV & I-Q \end{pmatrix} \tag{3.7}$$

where the quantities I, Q, U, V are averages over time and they are called Stokes parameters. I corresponds to the total intensity of the field, that for a blackbody emission is equivalent to the temperature of the radiation. Q and U represent the linear polarization. V is the circular polarization. Thomson scattering doesn't lead to circular polarization, so in CMB studies $V = 0$. Any $V \neq 0$ could be used to detect foreground(see later) contamination or some error in the systematics. These four parameters span the space of unpolarized, partially polarized and fully polarized light. For example, unpolarized radiation is $Q = U = V = 0$. While, for linear polarization, Q and U allow to measure the polarisation amplitude P and the polarisation angle χ(with $\hat{x}$)

$$\vec{P} = \sqrt{Q^2+U^2}\hat{\chi} \quad \hat{\chi} \parallel \vec{e} \quad \chi = \frac{1}{2}arctan(\frac{U}{Q})$$

The parameters I and V are physical observables independent of the coordinate system, while Q and U depend on the orientation of $\hat{x}$ and $\hat{y}$. Indeed the latter transform like a spin-2 object, i.e. if the

coordinate system is rotated by an angle α around an axis $\hat{n}$, then Q and U rotate to Q' and U' by an angle 2α[21]

$$(Q \pm iU)'(\hat{n}) = e^{\mp 2i\alpha}(Q \pm iU)(\hat{n}) \tag{3.8}$$

We can combine Q and U in one single complex spin-2 object representing the polarization in the direction $\hat{n} = (\theta, \phi)$ on the sky

$$P(\hat{n}) = (Q \pm iU)(\hat{n}) \tag{3.9}$$

The problem with this view is that Q and U depend on the, arbitrary, coordinate system that we take. Instead of continuing on this path, from now on we will rely on an easier geometrical view of the polarization of an electromagnetic wave which describes polarization with respect to the wave itself. Lets introduce a curl-free 'electric' $E-$mode and a divergence-free 'magnetic' $B-$mode. These two quantities are connected to (9), see for example[28], and as with the temperature anisotropies they can be decomposed using spherical harmonics:

$$E(\theta, \phi) = \sum_{lm} a^E_{lm} Y_{lm}(\theta\phi) \qquad B(\theta, \phi) = \sum_{lm} a^B_{lm} Y_{lm}(\theta\phi)$$

Different sources of temperature anisotropies give different patterns in polarization. $E-$ modes can be generated by scalar and tensor fluctuations. On the other hand, $B-$modes only by tensor perturbations. Thus, only primordial gravitational waves could leave an imprint on the $B-$mode spectrum[7][28].

3.1.2.3 *Description*

Temperature and polarization anisotropies are both generated by the same primordial fluctuations which we think to be Gaussian[8]. Linear theory gives an appropriate description of the evolution of these perturbations, at least until last scattering. Therefore, also primordial temperature and polarization anisotropies in the CMB are expected to follow a Gaussian statistics. The expansion coefficients a^X_{lm}, $X = T, B, E$ are expected to follow a Gaussian distribution and are independent of each other. To describe fully a Gaussian statistics we need the mean and the two point correlation function. The means are

$$\langle T_{lm} \rangle = 0 \quad \langle E_{lm} \rangle = 0 \quad \langle B_{lm} \rangle = 0$$

7 This is true at linear perturbation order. Graviational lensing can generate non-primordial $B-$modes from E-modes at second order.

8 Or very near to a normal distribution

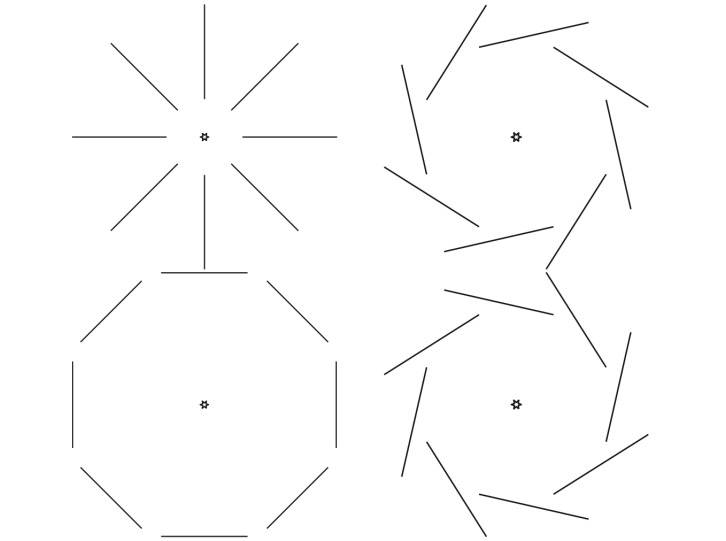


Figure 3.3: Typical E or B type polarization patterns. Image from [Ruth Durrer The Cosmic Microwave Background]

Adding isotropy, we are left with the power spectra of the perturbations.The mixed power spectra are:

$$\langle a^{X*}_{lm} a^{Y}_{l'm'} \rangle = C^{XY}_{l} \delta_{ll'} \delta_{mm'} \quad , \tag{3.10}$$

where X, Y could be T, E, B. If we require parity invariance in the physics responsible for anisotropies and polarization then it is possible to show that [(Graduate Texts in Physics) Gianluca Calcagni (auth.)-Classical and Quantum Cosmology-Springer International Publishing (2017).pdf]:

$$C^{TB}_{l} = 0 = C^{EB}_{l} \tag{3.11}$$

3.1.3 *CMB Histroy*

- 1965, Penzias and Wilson found a 'noisy' uniform signal in the microwave band. First official discovery of CMB. Its spectrum fitted that of a blackbody at temperature around $3K$. Modern Cosmology was born.
- 1989, COBE (Cosmic Background Explorer) satellite mission by NASA. COBE came with three instruments: DMR (Differential Microwave Radiometer) that would map anisotropies in the CMB(with angular resolution of 7°), FIRAS (Far Infrared Absolute Spec-

trophotometer) that would measure the spectrum of CMB and DIRBE (Diffuse Infrared Background Experiment) that would map dust emission[19]. Confirmation that CMB has a perfect blackbody temperature (parameter) $T = (2.735 \pm 0.060)K$ and most importantly with statistical spatial temperature anisotropies at a level of a part in one hundred thousands. These tiny variations in the intensity of the CMB over the sky show how matter and energy was distributed when the Universe was still very young[9]. We believe that they are the seeds for early structure formation that gave birth to galaxies and clusters of galaxies.

- 2001-2010, WMAP (Wilkinson Microwave Anisotropy Probe) satellite mission by NASA. WMAP had a sensitivity (i.e. how faint a signal each pixel can detect) 45 times that of COBE and an angular resolution of $22'$. In addition, it was also able to detect CMB polarization. Observing in five frequency bands between 23 and 94GHz, it was able to measure the acoustic peaks up to $l \sim 800$. WMAP CMB maps being highly accurate, precise and reliable placed tighter constraints on the cosmological parameters. Confirmation of the Standard Model of Cosmology.We had also the first detection of polarization. Cosmology has become a precision science.

- 2009-2013, Planck satellite mission by ESA. Angular resolution three times that of WMAP and sensitivity more than five times that of WMAP. Planck contained two instruments: LFI, Low Frequency Instrument, and HFI, High Frequency Instrument. In intensity Planck had nine bands between 30 and $857GHz$, while in polarization seven bands between 30 and $353GHz$. It improved the precision of the parameters of the standard model of comsmology . Still not detected the primordial gravitational waves via $B-$mode.

- March 2014: BICEP2 (Background Imaging of Cosmic Extragalactic Polarization) claimed the detection of large scale $B-$modes. In September of the same year they claim was withdrawn because it was found that the signal was largerly due to thermal dust emission, a type of foreground emission. This is a problem, because there is signal that is contaminating our signal of interest and we cannot extract it reliably. For this, in the next section we will give a look at those contaminating signals.

9 https://lambda.gsfc.nasa.gov/product/cobe/dmr_overview.cfm

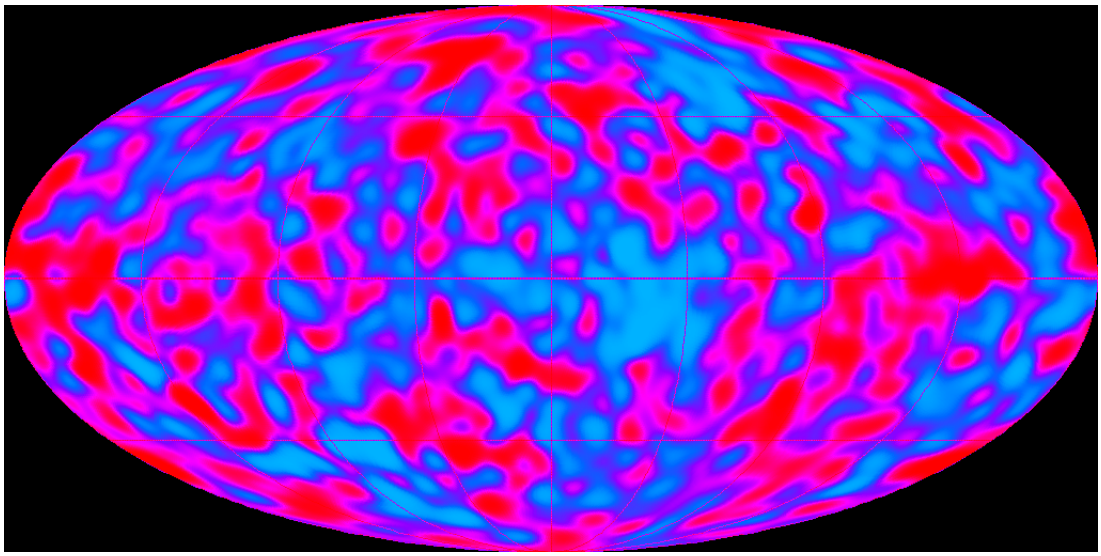


Figure 3.4: CMB temperature anisotropies map from COBE, using the DMR. Credits: NASA

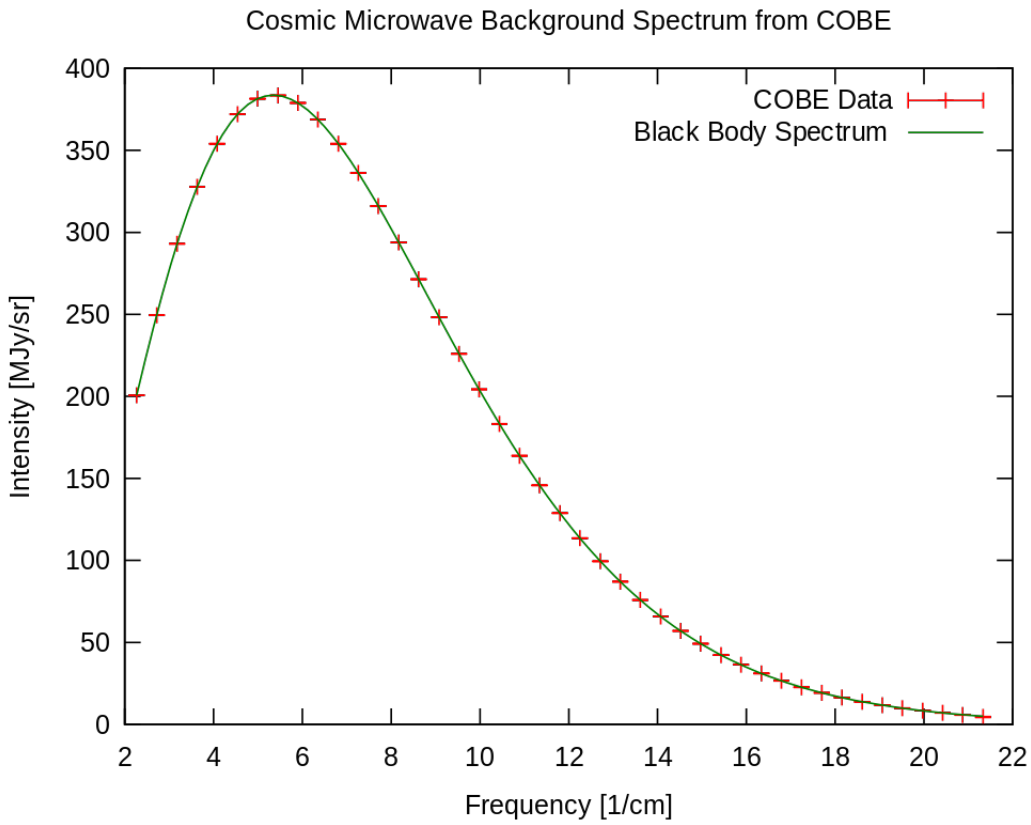


Figure 3.5: CMB perfect fit between the theoretical blackbody curve predicted by big bang theory and that observed in the microwave background by COBE with FIRAS. Credits: NASA

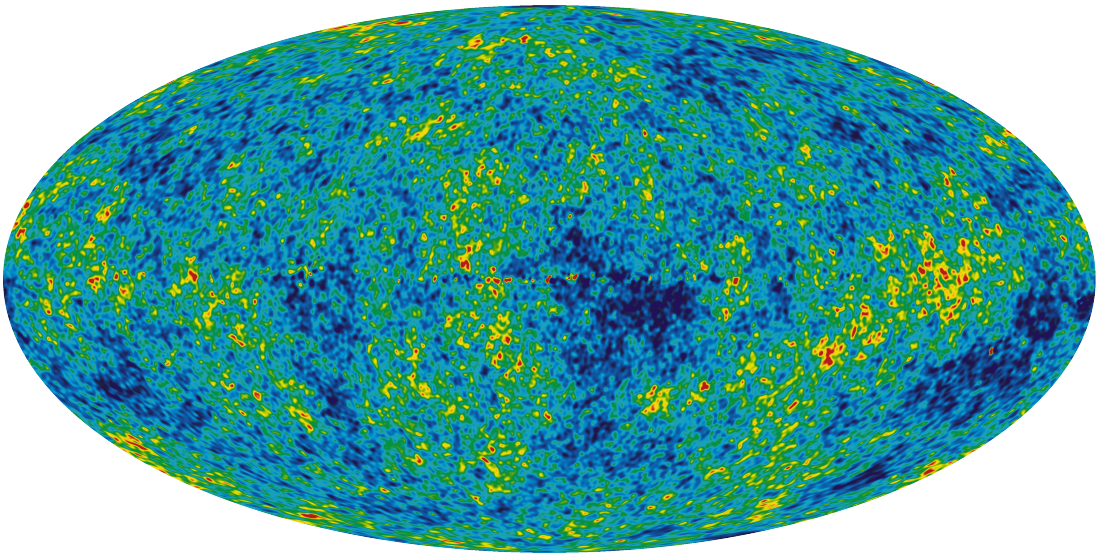
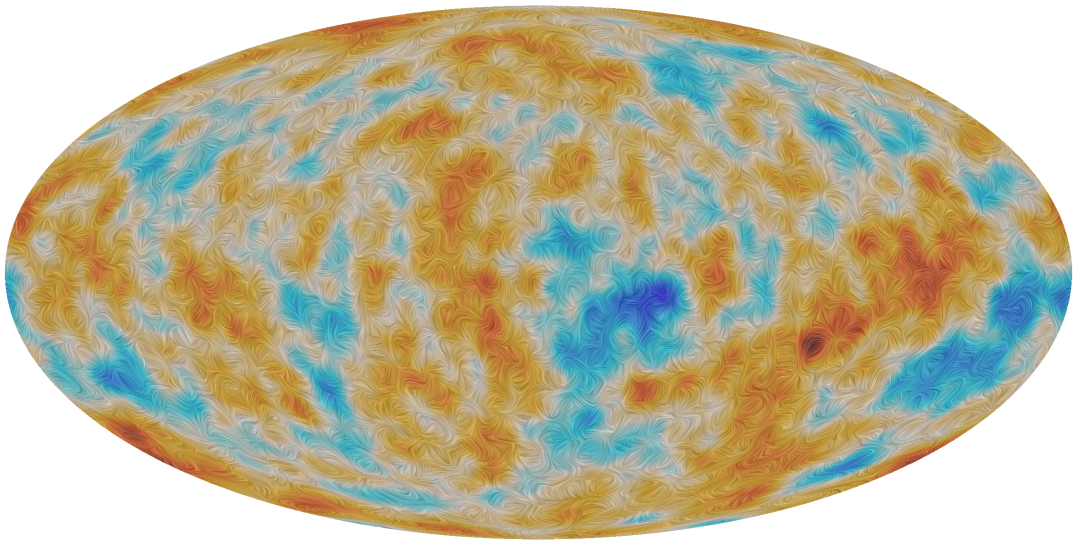


Figure 3.6: A five year foreground cleaned WMAP map. Colors represent the tiny temperature fluctuations, as in a weather map. Red regions are warmer and blue regions are colder by about 0.0002 degree. Credits: NASA

Figure 3.7: Planck Polarization Map. Credits: ESA and the Planck Collaboration



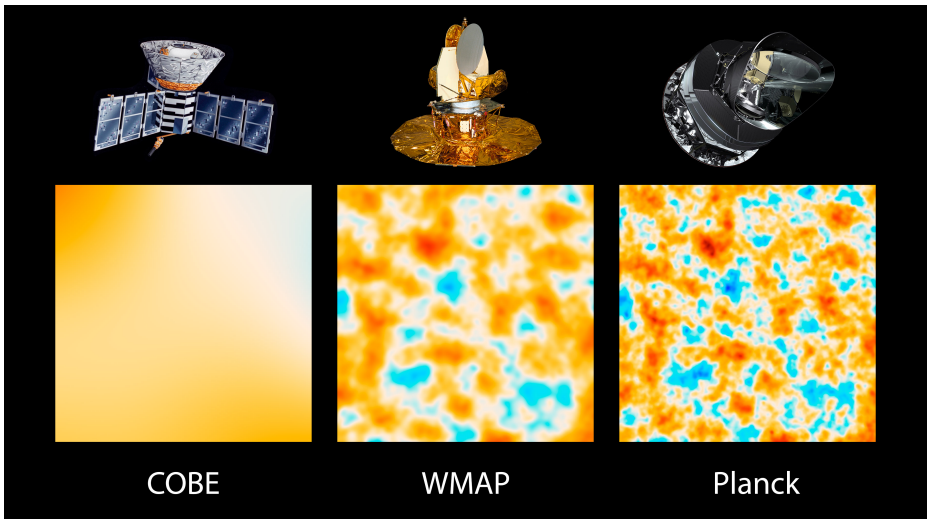


Figure 3.8: The Universe Comes into Sharper Focus. Image credit: NASA/JPL-Caltech/ESA

3.2 FOREGROUNDS

When we see the sky at the microwave band, our signal will be generally composed by: CMB signal + Foreground Signals. If we imagine to perform experiments in the outer space, the latter will be due to astrophysical components. Foregrounds will be characterized by their physical properties, which include their spatial morphology, their localisation, and their frequency scaling based on the physical understanding of their emission mechanisms[9].

Today we arrived at a level in which instruments has such sensitibity and angular resolution, that the main source of uncertainty when performing CMB experiments is the contamination by foregrounds and not instrumental noise.

There are three classes of foregrounds that can be found: solar system emission which comes from the planets and bodies in our solar system. Extragalactic emissions arising from a large background of resolved and unresolved radio and infrared galaxies, as well as clusters of galaxies. And finally diffuse galactic emissions which will be our focus in this section being the most annyoing source of contamination (as now)[10].

Diffuse galactic emissions are created by the local interstellar medium (ISM) that fills the space between the stars of our own galaxy, the Milky Way. The ISM is composed by three different media, all concentrated in the galactic plane[11]:

- a hot ionized medium (with low density $n \sim 10^3 m^{-3}$ and high temperature $T \sim 10^6 K$), presumably formed by Supernovae explosions

10 WHY??

11 http://www.star.ucl.ac.uk/~msw/teaching/PHAS2521/notes_1.pdf

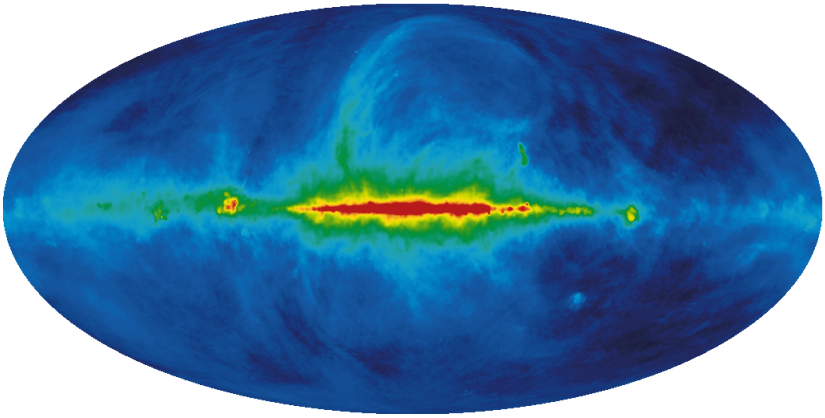


Figure 3.9: Synchrotron all sky map at $408MHz$ by [Haslam et al.] . This should be only used as a visual impression and not for analysis .

- A cold neutral medium (density $n \sim 10^7 - 10^9 m^{-3}$ and $T \sim 30 - 100K$) of molecular or atomic gas
- an interface region partly ionised, composed by:
  - a warm ionized medium ($n \sim 1.5 \times 10^5 m^{-3}$ and $T \sim 8000K$)
  - a warm neutral medium ($n \sim 1.5 \times 10^5 m^{-3}$ and $T \sim 8000K$)

The intensity of corresponding emissions decreases with galactic latitude, following a cosecant law(of the latitude)[9].

There are four main types of galactic emission. We will takle each one singularly.

3.2.1 *Synchrotron*

Synchrotron emission arises from relativistic electrons, produced by supernovae explosions, accelerated by the magnetic fields of the Milky Way. The magnetic fields extend outside the galactic plane, making possible for the energetic electrons to depart from it, generating emission at higher latitudes. Thus, with respect to other types of foreground emissions, synchrotron emission is less concentrated in the galactic plane.

The intensity of this radiation is $I_\nu \propto \nu^{-\alpha}$ where α is called the spectral index. The number density distribution is a function of the electron energy E following also a power law $N(E) \propto E^{-\gamma}$, with $(\gamma - 1)/2 = \alpha$. Finally, the Rayleigh-Jeans temperature is $\Delta T \propto \nu^{-\beta}$[12] with β taking values from 2.5 to 3.1, depending on the place in the sky.

The synchrotron emission can be highly polarized, up to $50 - 70\%$[Diffuse source separation].

12 Valid for low frequencies, $h\nu << k_B T$

3.2.2 *Free-free*

Interaction of free electrons with ions in ionised media makes the electrons lose energy by emitting photons. It's possible to calculate the brightness temperature at frequency ν for free-free emission,

$$T_{ff} \approx 0.08235 T_e^{-0.35} \nu^{-\beta} \int_{l.o.s.} N_e N_i dl \tag{3.12}$$

with spectral index β from 2.1 to 2.15 with errors ± 0.03[9].

Free-free radiation is not polarized because it is an incoherent emission from individual electrons scattered by nuclei[13] in a partially ionised medium.

3.2.3 *Thermal dust*

Dust is composed by small particles of several materials of various sizes and shapes, in amorphous or crystalline form, sometimes in aggregates or composites. Dust grains have sizes that span from few nanometers to micrometers[9].

In the microwave sky at frequencies $\geq 70GHz$, thermal dust emission dominates the foregrounds. There are several mechanisms of radiation emission from interstellar dust. The most important one, concerning CMB observations, is the grey body emission in the far infrared(wavelengths range: few μm to mm), due to the absorption of UV and optical photons coming from the interestallar radiation field[Bottino_Maria]. The intensity of the radiation depends on the composition, size and shape of the dust grain. If T_{dust} is the single component dust temperature, β the emissivity index , with values in the range $1-2$[10], and ν the fequency of observation, dust emissivity can be approximated by a modified black body law:

$$I_\nu \propto \nu^\beta B_\nu(T_{dust}) \tag{3.13}$$

where $B_\nu(T)$ is the usual black body radiation

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{exp(h\nu/kT) - 1} \quad , \tag{3.14}$$

T is the temperature of the body, c velocity of light, h Planck constant and k Boltzmann constant.

Dust is mainly concentrated along the Galactic plane and near the stellar formation inner regions, although being diffusely present in all the Galaxy. Its temperature will depend on the local radiation field, as well as on the efficience of emitting far-infrared light, ranging from about $5K$ to more than $30K$(due to the strong radiation produced by nearby stars). It is important to note that while seeing the sky in

13 In an isotropic and random way.

some direction, there will be dust clouds along the line of sight with different materials and compositions(in size, shapes and materials) as well different spectral indices β's, leading to different emissions. Therefore the formula of the greybody should be taken only as an approximation for a single component type of dust.

Typical thermal dust emission polarization levels are of the order of few per cent, up to $15 - 20\%$ in certain regions[9].

3.2.4 *Spinning dust*

It is possible to have non-thermal emissions from galactic dust at frequencies below $30GHz$. One mechanism of such non-thermal radiation is spinning dust, emitted by the smallest (orders of nm) interstellar dust grains and molecules, which can rotate at GHz frequencies.

The polarization amplitde of spinning dust emission is expected to be small, especially at frequencies above a few GHz[10].

3.2.5 *Foreground Removal*

Obviously if we would want to remove foregrounds from CMB data one idea is to use prior information we have about them. Information that includes what we've just described in the previous sections. For example we could select the region of observation so that we reduce contamination by foregrounds to the minimum(e.g. the galactic plane). The problem with this approach is that we couldn't estimate very well the level of contamination.

Another strategy consists in masking those regions in which we believe there is a significant foreground emission, deriving the CMB properties in the clean regions. The drawback of this strategy is that sky maps are incomplete.

Each foreground has a spectral law different from the others and from the CMB itself. Therefore, contaminations will be not the same in each frequency. The selection of the frequency of observation to minimize the overall foreground contamination is another option. From Figures 10 and 11 we see that there are intervals in which contamination is minimal. For example, for the CMB temperature anisotropies a good frequency window to observe the sky is around $70 - 100GHz$ where CMB dominates. On the other hand, from Figure 11 in polarization CMB signal is fainter than the foregrounds. This is a problem, because it is a wall against our knowledge of the CMB.

Finally, there are techniques for foreground removal from CMB data. Examples will be shown in Chapter 4.

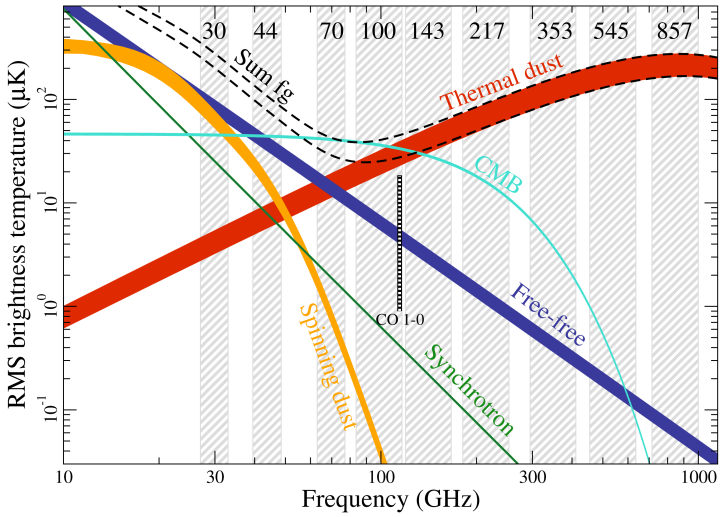


Figure 3.10: Temperature Foreground Spectra, from Planck 2015 Results X. Credits: ESA/Planck Collaboration

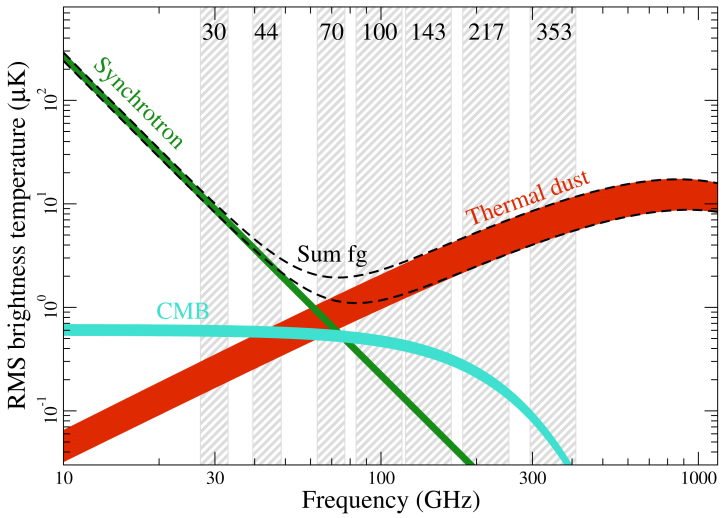


Figure 3.11: Polarized Foreground Spectra, from Planck 2015 Results X. Polarization anisotropy for free-free emission and dust emission is negligible. Credits: ESA/Planck Collaboration

3.3 MEASURING B-MODES

The CMB is linearly polarized at the 10% level. Also if we didn't talk about it, a key motivation for polarization measurements is the theory of inflation. One prediction of inflation is that it would produce a background of gravitational waves[14] accompanying the primordial density fluctuation. The amplitude of primordial gravitational waves is parameterized by the tensor-to-scalar ratio r, which could be determined by the unknown energy scale of inflationE_{inf} by

$$r = 8 \times 10^{-3}(E_{inf}/10^{16}GeV)^4 \tag{3.15}$$

Gravitational waves bump away due to the expansion of the universe when their wavelength is smaller than Hubble radius[7]. But we have the possibility to detect them via the imprint that they left on the CMB anisotropies. As said, temperature ones don't allow us to detect them. But, polarization yes. From the amplitude of the $B-$modes it is possible to measure r. So, a major current goal is to constrain, and if possible detect, large-scale $B-$modes. The detection of primordial $B-$modes on scales $l < 100(> 2\ deg)$ would be then a smoking-gun signature of inflation, determining the energy scale of inflation.

The problem is that as we saw foregrounds are expected to emit polarised light, with a polarisation fraction typically comparable, or larger, than that of the CMB. In particular, for $B-$modes of CMB polarisation, will be, if measurable, sub-dominant at every scale and every frequency[15]. This is an obstacle for component separation while we try to disentangle the CMB signal from the foregrounds.

Furethermore, foregrounds are a source to Non-Gaussianity to CMB maps, introducing a non zero bispectrum that adds information to our maps. So an idea would be to use this information to remove the foregrounds and extract the CMB. We will see in the next chapter a way to include this information, with the idea of estimating the CMB signal.

14 https://www.cfa.harvard.edu/~cbischoff/cmb/
15 https://arxiv.org/pdf/0901.1056.pdf

4

COMPONENTS STATISTICAL ANALYSIS

4.1 GENERAL SETTING OF THE PROBLEM

We imagine a room full of people. Among them one is saying something important that we want to listen to. But when we try to listen, what we obtain is something that is far from a clean conversation. We can suppose that the signal we are receiving from some point of the room is a linear combination of the various conversations around the room. This will obviously depend also on where we are receiving. So, the nearer we are to the people of interest the better is the conversation we want to listen to.

We could have been more general: imagining that our observations are some function of the sources. But we will consider linear functions only: these are more simpler to compute and to interprete. What do we mean with 'interprete'? In general we will have o observations that belong to some $o-$dimensional space. We want to transform our variables in some representation space of dimension r that can give us some information hidden in the data set of observations[13]. The transformed variables should be the underlying physical components that generated the data by some process.

Let's formalize the problem: consider a case where we have a number N_c of observable field signals $s_1, s_2, ..., s_{N_c}$ emitted by some physical source at some point $\vec{x}$. In cosmology, we could have the Galactic synchrotron radiation, Galactic thermal dust radiation, the atmosphere or the CMB itself, all of those emitting electromagnetic signals. A receiver hasn't the possibility to know the source of the hitting photon(e.g. is the photon a CMB one or a dust one or etc...?): it can only capture them (intensity or polarization) and convert in some electric signal.

Suppose we have a number N_f of detectors, each one looking for a certain point, with the ability to scan one and only one frequence, giving each one an observation value d_i at some point $\vec{x}$. i is the frequency of each detector($1 \leq i \leq N_f$). Observations are often discretized/pixelized, so we have only a countable number of points to check(on some map).

In the previous chapter we tried to grasp some understanding about the CMB signal and the foregrounds. The latter are not a big issue for the observations of CMB temperature anisotropies because they don't dominate at high Galactic latitudes and we can choose a frequency window to minimize their contamination. Similarly, for $E-$mode observations we haven't great problems. On the other hand,

primordial $B-$mode signal is expected to be too faint with respect to foregrounds, making its extraction a challenge. Its detection would constitute strong evidence of the existence of primordial gravitational waves.

From here there is the importance of finding methods of good component separation/foreground removal that could perform better than the existing ones.

4.2 FOREGROUND CLEANING METHODS

The forground emissions strongly contaminate CMB observations, making a challenge to use the latter for cosmological studies. There are several methods that were used to obtain CMB maps. Some of them attempt to estimate the CMB signal as clean as possible from any contamination, without any concern about the foreground emission physical laws. Others, tackle the problem by cleaning CMB data using (the little and inaccurate) information about the foregrounds. But all of them make use of multi-frequency input data to be separated into several physical (independent) components, among which there is the CMB. And all are based on the idea that the observed signal is the result of the superposition of independent sources: this means that in regions where foregrounds are strongly correlated and mixed those methods deteriorate.

Here we will see three examples only to have an idea on how some of the methods work. Good overviews are [9] and [14].

4.2.1 *Template Fitting*

The method is based on the idea that the observed signal in some direction in the sky at a microwave frequency ν is

$$T(\hat{n},\nu) = \sum_i T_i(\hat{n},\nu) + n(\hat{n},\nu) = \sum_i a_i(\nu) X_i(\hat{n}) + n(\hat{n},\nu) \quad . \quad (4.1)$$

We separated each signal $T_i(\hat{n},\nu)$ as $T_i(\hat{n},\nu) = a_i(\nu)X_i(\hat{n})$, where $X_i(\hat{n})$ is the template and $a_i(\nu)$ are the template coefficients. We have the data over a number of pixels and templates for the signals in which we are uninterested(constructed from observations where the non important component is dominant). Then we do a fit and find the template of the signal of interest(i.e. the CMB).

This method was applied by WMAP [14] using template maps for synchrotron, free-free and dust emissions.

A problem of the method is that the template coefficients are supposed to be independent of the position, fact that is not satisfied in the reality. On the other hand, the statistical properties of the noise are unaffected by the procedure[14]. The error in the cleaned map should

increase as we include a larger number of parameters. However, usually we use a much larger number of pixels $O(10^5)$ that can be used for the fit than the number of parameters $O(10)$, and therefore the impact of the noise amplitude after cleaning is negligible.

4.2.2 *ICA and FastICA*

This is a Blind Source Separation component method, i.e. it doesn't need any prior assumption about the distribution of the foreground components and their frequency dependence. Only statistical independence of the components is assumed.

To show the main idea lets take the simplest case of zero noise and the number of components equal to that of frequencies. Thus the model is

$$\vec{d}(\hat{n}) = A\vec{s}(\hat{n}) \tag{4.2}$$

where A is the unkown mixing matrix. We are looking for a separating matrix W such that $W\vec{d}(\hat{n})$ has independent entries. We know that if $W = A^{-1}$ then we have independent entries. Otherwise, we have to look for W that makes independence possible. Thus the blind source separation principle is: to separate components, make them independent. Then, we will obtain the components and the mixing matrix.

This blind method solution is not unique: indeed we can rescale or permute the components without touching independence. So, if we would like to fix this indeterminacies, we have to add constrains.

Note that in general the various components will be not Gaussian (especially if we consider foregrounds). On the other hand, the combination of independent foregrounds will tend toward a Gaussian distribution(Central Limit Theorem). So, we have 'more' Gaussianity in the data than in the single components.

In the FastICA method we assume that the signals are indepedent random processes on the map domain and that all the signals, but at most one, have non-Gaussian distributions. Then, independent components are extracted by maximizing some measure of non-Gaussianity[1].

4.2.3 *ILC*

The ILC (foreground substraction) method was applied succesfully by the WMAP team to obtain CMB anisotropies maps, so we will depict at least the idea behind it.

1 https://arxiv.org/pdf/astro-ph/0507267.pdf

The aim is to obtain a CMB cleaned map with poor information about foregrounds. To apply it we use the observed maps (at different frequencies), the fact that the foregrounds and the CMB component of interest are uncorrelated, and that the latter has a blackbody spectrum.

We model the data as

$$d^j_{lm} = a^j s_{lm} + f^j_{lm} + n^j_{lm}$$

where $d^j(\hat{n}_i)$ is the datum taken at frequency j and point $\hat{n}_i$, $f^j(\hat{n}_i)$ and $n^j(\hat{n}_i)$the foreground and noise contributions at frequency channel j, respectively, s the CMB spectrum (because of its blackbody nature is independent from the frequency) and a^j coefficients that take into account the possibility that observations would not be calibrated with respect to the CMB.

The idea is to obtain an estimate $\hat{s}$ of the true CMB $\vec{s}$ by performing a weighted average of the observed data over the frequency bands:

$$\hat{s}_{lm} = \vec{w}_l \cdot \vec{d}_{lm} \tag{4.3}$$

where the entries of the weight vector are chosen to maximize some criterion about the reconstructed estimate $\hat{s}$ of s, while keeping the component of interest untouched. This requires that the sum of the entries of $\vec{w}$ is 1.

As criterion we could choose the minimization of the variance σ^2 of the map $\hat{s}_{lm}$ with weight independent of $\vec{w}_l$.

First of all (3) is

$$\hat{s}_{lm} = \sum_i w^i_l d^i_{lm} = s_{lm} + \sum_i w^i_l f^i_{lm} + \sum_i w^i_l n^i_{lm} \tag{4.4}$$

So, if we assume decorrelation between s_{lm} and all foregrounds, and s_{lm} and all noises, the variance of the error is minimum when the variance of the ILC map itself is minimum.

The variance is $\langle |\hat{s}_{lm}|^2 \rangle$. Under the constraint $\sum_i w^i_l = 1$, minimizing we obtain

$$\vec{w}_l = \frac{\vec{a}^T C_l^{-1}}{\vec{a}^T C_l^{-1} \vec{a}} \tag{4.5}$$

where C_l is the covariance matrix of observations $C_l = \frac{1}{2l+1} \sum_m \vec{d}_{lm} \vec{d}^T_{lm}$.

The ILC method is based on the fact that the CMB signal and foregrounds are independent. Consequently, we cannot use it to estimate foregrounds components separately because they are correlated with each other.

4.3 ESTIMATION BY A BAYESIAN APPROACH

All the precedent methods have in common the facts that they choose some basis in which to express the data(e.g. spherical harmonic space) and a parametrisation of the data[[25].

One step forward is to choose a generic statistical model of the components based on generic assumptions to allow a full Bayesian exploration of the posterior density[25], with the aim of creating a CMB map.

4.3.0.1 *Bayesian Approach*

We collect some data and we want to interpret them via a model. This model contains signals that we think determine our data. Our main goal is to provide an estimation of this signals(or at least one of them). More specifically we would like to obtain a full probability distribution of the signals of interest given the data d, or $p(s|d)$(called also the posterior probability). Then we could take the mean or the maximum of this pdf to obtain our desired estimation.

Practically what we do is tackling the problem from another route: we first obtain the distribution of data given the signal and relate it to the posterior. We do it by means of Bayes theorem,

$$p(s|d) = \frac{p(d,s)}{p(d)} = \frac{p(d|s)}{p(d)} p(s)$$

where $p(d)$ is called the evidence $p(d) = \int ds p(d|s)p(s)$. The function $p(s)$ is the prior distribution, expressing our knowledge about the parameters. In general we will not have exact knowledge about them. So, it is reasonable to expect that our estimations will be biased.

Once done that we marginalize over any signal we are not interested in. If $s = (s_{in}, s_{other})$, where s_{in} is the signal of interest, then what we do is $p_{final}(s_{in}) = \int ds_{other} p(s|d)$. Then $p_{final}(s_{in})$ is all what we need for studying our signal of interest.

4.3.1 *Modeling the data*

Our goal is to produce a CMB map, given the data and some other information(covariance matrices, statistical hypothesis on the components signal, etc...). The data model is a signal plus noise. The signal is assumed to be a linear mixture of several emissions(e.g. CMB, free-free, dust, etc...). Data at fixed spherical harmonic coefficient (l, m)

of the observation map at some frequency ν_i(that it will be indicated simply as i) will be modeled as:

$$d_{ilm} = \sum_{k=1}^{N_c} A_{ik} s_{klm} + n_{ilm} \tag{4.6}$$

where i runs over 1 to N_f(the total number of frequencies), k from 1 to N_c (total number of signal components at fixed (l, m)) . A_{ik} is the amount of component k at frequency i $s_k = \{s_{klm}; l = l_{min}, ..., l_{max}, m = -l, ..., l\}$ is the spherical harmonic transform of the $k-$th component map and n_{ilm} is how much of the instrumental noise is present in d_{ilm}. The formula can be recast in a matrix form

$$\vec{d}_{lm} = A\vec{s}_{lm} + \vec{n}_{lm} \tag{4.7}$$

The matrix A, whose components are A_{ik}, is called mixing matrix. We will now assume Gaussian noise with zero mean, isotropy and diagonal covariance matrix N_{lm}, for a fixed (l, m), whose components are $\langle n_{ilm} n_{i'l'm'} \rangle = \nu_{il}^2 \delta_{ii'} \delta_{mm'}$. Because of this, $N_{lm} \equiv N_l$ will be our $N_f \times N_f$ diagonal covariance matrix, with components ν_{il}^2 for a fixed l. This means that we assume uncorrelation(and for gaussian case means independence) between different frequencies and different multipoles (l, m).

The power spectra of the components are encapsulated in C_{klm}. If we assume statistical isotropy we will have $C_{klm} = C_{kl}$, where $C_{kl}\delta_{kk'}\delta_{ll'}\delta_{mm'} = {}_{klm}\, s_{k'l'm'}\rangle$.

Finally, we are also assuming statistical independence between the components(also the noise is independent).

4.3.2 *Notation*

We can now impose the problem notation:

- N_l diagonal (channels are independent from each other)

  $N_l = diag(\nu_{1,l}^2, ..., \nu_{N_f,l}^2)$

- The N_c component vector is splitted in $1 + (N_c - 1)$ components

  $$\vec{s}_{lm} = \begin{pmatrix} s_{cmb,lm} \\ \vec{s}_{f,lm} \end{pmatrix} = \begin{pmatrix} s_{cmb,lm} \\ \vec{0} \end{pmatrix} + \begin{pmatrix} 0 \\ \vec{s}_{f,lm} \end{pmatrix} \equiv \vec{s}_{c,lm} + \vec{s}_{f,lm}$$

  where $s_{cmb,lm}$ is the CMB signal, while $s_{f,lm}$ is the foregrounds signal.

- The diagonal matrix C_l is written with a 1×1 and a $N_c - 1 \times N_c - 1$ matrices

  $$C_l = \begin{pmatrix} C_{cmb,l} & 0 \\ 0 & C_{f,l} \end{pmatrix}$$

with

$$C_l^{-1} = \begin{pmatrix} C_{cmb,l}^{-1} & 0 \\ 0 & C_{f,l}^{-1} \end{pmatrix}$$

because C_l is diagonal.
And we can write $C_l = C_{c,l} + C_{f,l}$, where

$$C_{c,l} = \begin{pmatrix} C_{cmb,l} & \vec{0}^T \\ \vec{0} & \mathbb{O} \end{pmatrix} \quad C_{f,l} = \begin{pmatrix} 0 & \vec{0}^T \\ \vec{0} & C_{f,l} \end{pmatrix}$$

- Finally[2]

$$A = \begin{pmatrix} 1 & \\ . & \\ . & A_f \\ . & \\ 1 & \end{pmatrix} = \begin{pmatrix} 1 & \\ . & \\ . & \mathbb{O}_{N_f \times (N_c - 1)} \\ . & \\ 1 & \end{pmatrix} + \begin{pmatrix} 0 & \\ . & \\ . & A_f \\ . & \\ 0 & \end{pmatrix} \equiv A_c + A_f$$

(don't confuse the two A_f's)
With this notation

$$A\vec{s}_{lm} = A_c\vec{s}_{c,lm} + A_f\vec{s}_{f,lm}$$

And

$$C_l\vec{s}_{f,lm} = C_{f,l}\vec{s}_{f,lm} \qquad C_l\vec{s}_{c,lm} = C_{c,l}\vec{s}_{c,lm}$$

4.4 GAUSSIAN CASE

Our starting point is

$$P(A, \vec{s}, C|\vec{d}) \propto \mathcal{L}(\vec{d}|A, \vec{s})P(A, \vec{s}, C) = \mathcal{L}(\vec{d}|A, \vec{s})P(A|\vec{s}, C)P(\vec{s}, C) =_{Independence}$$

$$= \mathcal{L}(\vec{d}|A, \vec{s})P(A)P(\vec{s}, C) = \mathcal{L}(\vec{d}|A, \vec{s})P(A)P(\vec{s}|C)P(C)$$

We also assume:

2 We also define $\vec{e} = \begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix}$

- Gaussian Prior for the components $\vec{s}$ (with zero mean)
- Flat Prior for A and C, $P(A) \propto c_A$ and $P(C) \propto c_C$, where c_A and c_C are costants.

Then for our patch of the sky (rember prior probability on data as a known costant)

$$P(A,\vec{s},C|\vec{d}) \propto \mathcal{L}(\vec{d}|A,\vec{s})P(\vec{s}) = \prod_{lm} \mathcal{N}(\vec{d}_{lm}; A\vec{s}_{lm}, N_l)\mathcal{N}(\vec{s}_{lm};\vec{0},C_l) =$$

$$\prod_{lm} \mathcal{N}(\vec{d}_{lm}; A\vec{s}_{lm}, N_l)\mathcal{N}(\vec{s}_{lm};\vec{0},C_l) = \prod_{lm} (2\pi)^{-N_f/2}|N_l|^{-1/2}\times$$

$$\times exp[-\frac{1}{2}(\vec{d}_{lm} - A\vec{s}_{lm})^T N_l^{-1}(\vec{d}_{lm} - A\vec{s}_{lm})] \times (2\pi)^{-N_c/2}|C_l|^{-1/2} exp[-\frac{1}{2}\vec{s}_{lm}^T C_l^{-1}\vec{s}_{lm}] \tag{4.8}$$

Our aim is to obtain a distribution that depends only on $\vec{s}_{c,lm}$. Because for every lm the distribution is the same in form (in the product (1), and as we said A is the same) we fix lm. From the exp in (1)

$$exp\{-\frac{1}{2}[\vec{d_{lm}}^T N_l^{-1}\vec{d_{lm}} - \vec{d}_{lm}^T N_l^{-1} A\vec{s}_{lm}\}\times$$

$$\times exp\{-(A\vec{s}_{lm})^T N_l^{-1}\vec{d}_{lm} + (A\vec{s}_{lm})^T N_l^{-1} A\vec{s}_{lm} + \vec{s}_{lm}^T C_l^{-1}\vec{s}_{lm}]\}$$

Using $(N_l^{-1})^T = N_l^{-1}$ because N_l is diagonal and the fact that $exp(a^T) = exp(a)$ if $a = \vec{A}^T\vec{B}$ is a number.

$$= exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm} + (A\vec{s}_{lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A\vec{s}_{lm})^T N_l^{-1} A\vec{s}_{lm} - \frac{1}{2}\vec{s}_{lm}^T C_l^{-1}\vec{s}_{lm}\} =$$

We want to marginalize with respect to $\vec{s}_{f,lm}$ so that we can obtain a distribution for $\vec{s_{c,lm}}$. So firstly we write the above expression so that it will be simple to make the integral of the marginalization. It is useful to write the last expression more explicitly with the developed notation:

$$= exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm}\}\times$$

$$\times exp\{(A_f\vec{s}_{f,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_f\vec{s}_{f,lm})^T N_l^{-1} A_f\vec{s}_{f,lm} - (A_c\vec{s}_{c,lm})^T N_l^{-1} A_f\vec{s}_{f,lm}\}\times$$

$$\times exp\{-\frac{1}{2}\vec{s}_{f,lm}^{T}C_l^{-1}\vec{s}_{f,lm} - \vec{s}_{c,lm}^{T}C_l^{-1}\vec{s}_{f,lm}\}\times$$

$$\times exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1}A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^{T}C_l^{-1}\vec{s}_{c,lm}\}$$

We will now focus on the second and third row[3].

$$exp\{-\frac{1}{2}\vec{s}_{f,lm}{}^T(A_f^T N_l^{-1}A_f + C_l^{-1})\vec{s}_{f,lm} + (A_f\vec{s}_{f,lm})^T N_l^{-1}\vec{d}_{lm}\}\times$$

$$\times exp\{-(A_c\vec{s}_{c,lm})^T N_l^{-1}A_f\vec{s}_{f,lm} - \vec{s}_{c,lm}^{T}C_l^{-1}\vec{s}_{f,lm}\}$$

We note that $\vec{s}_{c,lm}^{T}C_l^{-1}\vec{s}_{f,lm} = 0$ so that in the end

$$exp\{-\frac{1}{2}\vec{s}_{f,lm}{}^T(A_f^T N_l^{-1}A_f + C_l^{-1})\vec{s}_{f,lm} + (A_f\vec{s}_{f,lm})^T N_l^{-1}\vec{d}_{lm} - (A_c\vec{s}_{c,lm})^T N_l^{-1}A_f\vec{s}_{f,lm}\} =$$

$$= exp\{-\frac{1}{2}\vec{s}_{f,lm}{}^T(A_f^T N_l^{-1}A_f + C_l^{-1})\vec{s}_{f,lm} + \vec{s}_{f,lm}{}^T(A_f^T N_l^{-1}\vec{d}_{lm} - A_f^T N_l^{-1}A_c\vec{s}_{c,lm})\} \equiv$$

$$\equiv exp\{-\frac{1}{2}\vec{s}_{f,lm}{}^T\Lambda_l\vec{s}_{f,lm} + \vec{s}_{f,lm}{}^T\vec{a}_{lm}\}$$

The argument of the exp function can be rewritten

$$-\frac{1}{2}\vec{s}_{f,lm}{}^T\Lambda_l\vec{s}_{f,lm} + \vec{s}_{f,lm}{}^T\vec{a}_{lm} = -\frac{1}{2}(\vec{s}_{f,lm} - \Lambda_l^{-1}\vec{a}_{lm})^T\Lambda_l(\vec{s}_{f,lm} - \Lambda_l^{-1}\vec{a}_{lm}) - \frac{1}{2}\vec{a}_{lm}{}^T\Lambda_l^{-1}\vec{a}_{lm}$$

Now making the marginalization with respect to $\vec{s}_{f,lm}$ means that we have to make this integral

$$\int d\vec{s}_{f,lm} exp\{-\frac{1}{2}(\vec{s}_{f,lm} - \Lambda_l^{-1}\vec{a}_{lm})^T\Lambda_l(\vec{s}_{f,lm} - \Lambda_l^{-1}\vec{a}_{lm})\} \tag{4.9}$$

By calling where Γ_l the matrix Λ_l without the first row and the first column and $\vec{a}_{lm} = (0, \vec{r}_{lm})^T$, (now the integral is $N_c - 1$-dimensional)

$$\int d\vec{s}_{f,lm} exp\{-\frac{1}{2}(\vec{s}_{f,lm} - \Gamma_l^{-1}\vec{r}_{lm})^T\Lambda_l(\vec{s}_{f,lm} - \Gamma_l^{-1}\vec{r}_{lm})\} \tag{4.10}$$

3 We don't need the first because there is no dependence on the components(once data are given).

This is the integral of an unnormalized multivariate gaussian with a mean and a covariance matrix. We recall that in general a normalized gaussian (of dimension N) is

$$\mathcal{N}(\vec{x};\vec{\mu},\Sigma) = (2\pi)^{-N/2}|\Sigma|^{-1/2}exp\{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\}$$

with the two conditions

- $\mathcal{N}(\vec{x};\vec{\mu},\Sigma) \geq 0$
- $\int_{domain} d\vec{x}\mathcal{N}(\vec{x};\vec{\mu},\Sigma) = 1$

Then (2) depends only on the determinant of the covariance matrix Γ_l . This gives a number that we 'put' in the proportionality simbol.

In the end we have

$$exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1}A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^T C_l^{-1}\vec{s}_{c,lm}\}\times$$

$$\times exp\{-\frac{1}{2}\vec{a}_{lm}{}^T\Lambda_l^{-1}\vec{a}_{lm}\} =$$

$$= exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1}A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^T C_l^{-1}\vec{s}_{c,lm}+$$

$$+\frac{1}{2}[A_f^T N_l^{-1}(\vec{d}_{lm} - A_c\vec{s}_{c,lm})]^T\Lambda_l^{-1}[A_f^T N_l^{-1}(\vec{d}_{lm} - A_c\vec{s}_{c,lm})]\} =$$

$$= exp\{-\frac{1}{2}\vec{s}_{c,lm}^T(A_c^T N_l^{-1}A_c + C_l^{-1} - A_c^T N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1}A_c)\vec{s}_{c,lm}+$$

$$+\vec{s}_{c,lm}^T(A_c^T N_l^{-1} - A_c^T N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1})\vec{d}_{lm}\} \times no\vec{s}_{c,lm}$$

where $no\vec{s}_{c,lm}$ indicates the part with no dependence on $\vec{s}_{c,lm}$. Neglecting this, we focus on

$$exp\{-\frac{1}{2}\vec{s}_{c,lm}^T(A_c^T N_l^{-1}A_c + C_l^{-1} - A_c^T N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1}A_c)\vec{s}_{c,lm}+$$

$$+\vec{s}_{c,lm}^T(A_c^T N_l^{-1} - A_c^T N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1})\vec{d}_{lm}\} \tag{4.11}$$

Rewritten as

$$exp\{-\frac{1}{2}(\vec{s}_{c,lm} - K_l^{-1}\vec{b}_{lm})^T K_l(\vec{s}_{c,lm} - K_l^{-1}\vec{b}_{lm})\} \times exp\{-\frac{1}{2}\vec{b}_{lm}^T K_l^{-1}\vec{b}_{lm}\}$$

(4.12)

with $\vec{b}_{lm}$ independent from $\vec{s}_{c,lm}$. As before with $\vec{s}_{f,lm}$, we can see that we have an unnormalized multivariate Gaussian function of $\vec{s}_{c,lm}$.

We will take as estimator of the true CMB signal the maximum of the posterior (gaussian) distribution[4].

We see that maximizing (4) is the same as maximizing (3). So we will maximize (3) and will find the value of $s_{cmb,lm}$.

Now we explicate the terms in (3):

$$(A_c\vec{s}_{c,lm})^T N_l^{-1}(A_c\vec{s}_{c,lm}) = s_{cmb,lm}^2 \vec{e}^T N_l^{-1}\vec{e}$$

$$\vec{s}_{c,lm}{}^T C_l^{-1}\vec{s}_{c,lm} = s_{cmb}^2 C_{cmb,l}^{-1}$$

Note now that if $B_l = (N_l + A_f C_l A_f^T)$ then

$$N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1} = N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1}B_l B_l^{-1} = (N_l^{-1}A_f\Lambda_l^{-1})(A_f^T N_l^{-1}B_l)B_l^{-1} =$$

$$= (N_l^{-1}A_f\Lambda_l^{-1})(\Lambda_l C_l A_f^T)B_l^{-1} = N_l^{-1}A_f C_l A_f^T B_l^{-1} = N_l^{-1}(B_l - N_l)B_l^{-1}$$

The remaining terms are

1. $(A_c\vec{s}_{c,lm})^T N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1}(A_c\vec{s}_{c,lm}) = s_{cmb,lm}^2\vec{e}^T N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1}\vec{e} =$
   $= s_{cmb,lm}^2\vec{e}^T N_l^{-1}(B_l - N_l)B_l^{-1}\vec{e}$
1. $\vec{s}_{c,lm}^T(A_c^T N_l^{-1} - A_c^T N_l^{-1}A_f\Lambda_l^{-1}A_f^T N_l^{-1})\vec{d}_{lm} =$
   $= s_{cmb,lm}\vec{e}^T(N_l^{-1} - N_l^{-1}(B_l - N_l)B_l^{-1})\vec{d}_{lm} = s_{cmb,lm}\vec{e}^T B_l^{-1}\vec{d}_{lm}$

Adding all terms we obtain in the exponential

$$exp\{-\frac{1}{2}s_{cmb,lm}^2\vec{e}^T N_l^{-1}\vec{e} - \frac{1}{2}s_{cmb}^2 C_{cmb,l}^{-1} + \frac{1}{2}s_{cmb,lm}^2\vec{e}^T N_l^{-1}(B_l - N_l)B_l^{-1}\vec{e} + s_{cmb,lm}\vec{e}^T B_l^{-1}\vec{d}_{lm}\}$$

The first derivative is

$$exp\{-\frac{1}{2}s_{cmb,lm}^2\vec{e}^T N_l^{-1}\vec{e} - \frac{1}{2}s_{cmb}^2 C_{cmb,l}^{-1} + \frac{1}{2}s_{cmb,lm}^2\vec{e}^T N_l^{-1}(B_l - N_l)B_l^{-1}\vec{e} + s_{cmb,lm}\vec{e}^T B_l^{-1}\vec{d}_{lm}\}\times$$

$$(-s_{cmb,lm}\vec{e}^T N_l^{-1}\vec{e} - s_{cmb}C_{cmb,l}^{-1} + s_{cmb,lm}\vec{e}^T N_l^{-1}(B_l - N_l)B_l^{-1}\vec{e} + \vec{e}^T B_l^{-1}\vec{d}_{lm})$$

4 We also note that for a gaussian distribution the mean maximizes it. So, in principle we could calculate it to obtain the maximum.

Focusing on the second factor

$$(-s_{cmb,lm}C^{-1}_{cmb,l} - s_{cmb,lm}\vec{e}^T B_l^{-1}\vec{e} + \vec{e}^T B_l^{-1}\vec{d}_{lm}) = 0$$

In the end

$$\hat{s}_{cmb,lm} = \frac{\vec{e}^T B_l^{-1}\vec{d}_{lm}}{C^{-1}_{cmb,l} + \vec{e}^T B_l^{-1}\vec{e}} = \frac{\vec{e}^T (N_l + A_f C_{f,l} A_f^T)^{-1}\vec{d}_{lm}}{C^{-1}_{cmb,l} + \vec{e}^T (N_l + A_f C_{f,l} A_f^T)^{-1}\vec{e}} \quad (4.13)$$

where here A_f is $N_f \times (N_c - 1)$ and $C_{f,l}$ is $(N_c - 1) \times (N_c - 1)$. This is the result of equation (14) of [25]. As pointed out in this paper, if CMB fluctuations are neglected in then standard ILC formula and if we the prior variance of CMB is infinite in the last formula, then we obtain the standard ILC result of the previous section.

4.5 NEARLY GAUSSIAN FORGROUNDS COMPONENTS CASE

In the vector $\vec{s}_{lm}$ the first component, for our convention, is the component of interest, while the other components are the ones to be marginalized. In the previous section we saw how to estimate the component of interest, s_{cmb}, supposing $\vec{s}_{lm}$, and the noise, to follow a multivariate gaussian distribution. Now we want to relax a bit this condition, by imposing a gaussian distribution for the component of interest(the CMB), and a non-gaussian one (in particular a nearly gaussian) for the other components (we saw in the previous chapter why we could do this choice). The noise will still follow a multivariate gaussian.

Now, the initial probability density is

$$K(2\pi)^{-N_f/2}|N_l|^{-1/2} exp[-\frac{1}{2}(\vec{d}_{lm} - A\vec{s}_{lm})^T N_l^{-1}(\vec{d}_{lm} - A\vec{s}_{lm})] \times$$

$$\times (2\pi)^{-N_c/2}|C_l|^{-1/2} exp[-\frac{1}{2}\vec{s}^T_{lm} C_l^{-1}\vec{s}_{lm}] \times$$

$$\times \prod_{i=1}^{N_c-1}[1 + \sum_{k=1}^{\infty} a_{ik} H_k(\frac{s_{i,lm}}{\sigma_i})] \quad (4.14)$$

Where the coefficients a_{il} are calculated by means of the formula (43) of [Expansions for nearly Gaussian distribution]. For reference we report the coefficients in the Appendix. (nota però che la gaussiana $Z(x)$ lì usata è a varianza unitaria).

We should note that the series in this formula is not supposed to converge going to infinity. Practically we will always truncate this series, so that all higher coefficients will be set to zero. Valid edgeworth expansion. So we suppose the series to converge.

We proceed as in the previous section. In the end we will obtain something like this

$$K(2\pi)^{-N_f/2}|N_l|^{-1/2}(2\pi)^{-N_c/2}|C_l|^{-1/2} \times exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm}\}\times$$

$$\times exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1} A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^T C_l^{-1}\vec{s}_{c,lm}\}\times$$

$$\times[exp\{-\frac{1}{2}\vec{a}_{lm}{}^T\Lambda_l^{-1}\vec{a}_{lm}\}(2\pi)^{N_c/2}|\Lambda_l|^{-1/2}+$$

$$+\int d\vec{s}_{f,lm} exp\{-\frac{1}{2}\vec{s}_{f,lm}{}^T\Lambda_l\vec{s}_{f,lm} + \vec{s}_{f,lm}{}^T\vec{a}_{lm}\}(\prod_{i=1}^{N_c-1}[1+\sum_{k=1}^{\infty}a_{ik}H_k(s_i)]-1) \tag{4.15}$$

where we recall that $\vec{s}_{f,lm} = (0,\vec{s})^T$ with $\vec{s} = (s_1,...,s_{N_c-1})^T$.

$$exp\{-\frac{1}{2}\vec{s}_{f,lm}{}^T\Lambda_l\vec{s}_{f,lm} + \vec{s}_{f,lm}{}^T\vec{a}_{lm}\}$$

is

$$exp\{-\frac{1}{2}\vec{s}^T R_l\vec{s} + \vec{s}^T\vec{b}_{lm}\}$$

where R_l is the matrix Λ_l without the first row and the first column[5] and $\vec{a}_{lm} = (0,\vec{b}_{lm})^T$.

$exp\{-\frac{1}{2}\vec{s}^T R_l\vec{s} + \vec{s}^T\vec{b}_{lm}\}$ can be rewritten as

$$exp\{-\frac{1}{2}(\vec{s}_{lm} - R_l^{-1}\vec{b}_{lm})^T R_l(\vec{s}_{lm} - R_l^{-1}\vec{b}_{lm}) - \frac{1}{2}\vec{b}_{lm}{}^T R_l^{-1}\vec{b}_{lm}\}$$

We obtain

$$K(2\pi)^{-N_f/2}|N_l|^{-1/2}(2\pi)^{-N_c/2}|C_l|^{-1/2} \times exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm}\}\times$$

5 It was Γ_l in the previous chapter

$$\times exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1} A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^T C_l^{-1}\vec{s}_{c,lm}\}\times$$

$$\times[exp\{-\frac{1}{2}\vec{a}_{lm}{}^T\Lambda_l^{-1}\vec{a}_{lm}\} \times (2\pi)^{N_c/2}|\Lambda_l|^{-1/2}+$$

$$+exp\{-\frac{1}{2}\vec{b}_{lm}{}^T R_l^{-1}\vec{b}_{lm}\}\times$$

$$\times \int d\vec{s}_{lm} exp\{-\frac{1}{2}(\vec{s}_{lm} - R_l^{-1}\vec{b}_{lm})^T R_l(\vec{s}_{lm} - R_l^{-1}\vec{b}_{lm})\}(\prod_{i=1}^{N_c-1}[1+\sum_{k=1}^{\infty} a_{ik}H_k(s_{i,lm})] - 1) \tag{4.16}$$

Now we have 'only' to perform the integral

$$\int d\vec{s}_{lm} exp\{-\frac{1}{2}(\vec{s}_{lm} - R_l^{-1}\vec{b}_{lm})^T R_l(\vec{s}_{lm} - R_l^{-1}\vec{b}_{lm})\}(\prod_{i=1}^{N_c-1}[1+\sum_{k=1}^{\infty} a_{ik}H_k(s_i)] - 1)$$

We define $f(\vec{s}_{lm}) = \prod_{i=1}^{N_c-1}[1+\sum_{l=k}^{\infty} a_{ik}H_k(s_i)]$. Let's change the integration variable:

$$\vec{t}_{lm} = \vec{s}_{lm} - R_l^{-1}\vec{b}_{lm}$$

e $g(\vec{t}_{lm}) := f(\vec{t}_{lm} + R_l^{-1}\vec{b}_{lm})$.
So that

$$\int d\vec{t}_{lm} exp\{-\frac{1}{2}\vec{t}_{lm}{}^T R_l\vec{t}_{lm}\}g(\vec{t}_{lm})$$

Let's note that $g(\vec{t}_{lm})$ is analytic (is a product of polynomials, so a sum of polynomials and we supposed that it is convergent).
Now, let's use the 'master' formula[6]

$$\int f(\vec{x})e^{(-\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}x_i x_j)}d^n x = \sqrt{\frac{(2\pi)^n}{detA}}e^{(\frac{1}{2}\sum_{i,j=1}^{n}(A^{-1})_{ij}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j})}f(\vec{x})|_{\vec{x}=0} \tag{4.17}$$

.

6 https://en.wikipedia.org/wiki/Gaussian_integral#n-dimensional_and_functional_generalization . But there are other ways to compute the integral

So that

$$\int d\vec{t}_{lm} exp\{-\frac{1}{2}\vec{t}_{lm}{}^T R_l \vec{t}_{lm}\} g(\vec{t}_{lm}) =$$

$$= \sqrt{\frac{(2\pi)^{N_c-1}}{detR_l}} exp\{\frac{1}{2}\sum_{i,j=1}^{N_c-1}(R_l^{-1})_{ij}\frac{\partial}{\partial t_{lm,i}}\frac{\partial}{\partial t_{lm,j}}\} g(\vec{t}_{lm})|_{\vec{t}_{lm}=0} =$$

$$= \sqrt{\frac{(2\pi)^{N_c-1}}{detR_l}} exp\{\frac{1}{2}\sum_{i,j=1}^{N_c-1}(R_l^{-1})_{ij}\frac{\partial}{\partial s_{lm,i}}\frac{\partial}{\partial s_{lm.j}}\} f(\vec{s}_{lm})|_{\vec{s}_{lm}=R_l^{-1}\vec{b}_{lm}} =$$

$$\sqrt{\frac{(2\pi)^{N_c-1}}{detR_l}} exp\{\frac{1}{2}\sum_{i,j=1}^{N_c-1}(R_l^{-1})_{ij}\frac{\partial}{\partial s_{lm,i}}\frac{\partial}{\partial s_{lm.j}}\}(\prod_{i=1}^{N_c-1}[1+\sum_{k=1}^{\infty}a_{ik}H_k(s_i)]-1)|_{\vec{s}_{lm}=R_l^{-1}\vec{b}_{lm}} \tag{4.18}$$

To apply the derivatives in the exponential we see it through its series expansion.

The final formula is:

$$K(2\pi)^{-N_f/2}|N_l|^{-1/2}(2\pi)^{-N_c/2}|C_l|^{-1/2} \times exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm}\}\times$$

$$\times exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1} A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^T C_l^{-1}\vec{s}_{c,lm}\}\times$$

$$\times[exp\{-\frac{1}{2}\vec{a}_{lm}{}^T\Lambda_l^{-1}\vec{a}_{lm}\} \times (2\pi)^{(N_c-1)/2}|R_l|^{-1/2} +$$

$$+exp\{-\frac{1}{2}\vec{b}_{lm}{}^T R_l^{-1}\vec{b}_{lm}\}\sqrt{\frac{(2\pi)^{N_c-1}}{detR_l}}\times$$

$$\times exp\{\frac{1}{2}\sum_{i,j=1}^{N_c-1}(R_l^{-1})_{ij}\frac{\partial}{\partial s_{lm,i}}\frac{\partial}{\partial s_{lm.j}}\}(\prod_{i=1}^{N_c-1}[1+\sum_{k=1}^{\infty}a_{ik}H_k(s_i)]-1)|_{\vec{s}_{lm}=R_l^{-1}\vec{b}_{lm}}] \tag{4.19}$$

If we define $B_l = N_l + A_f C_l A_f^T$, then the second and third line of exponentials of this last formula can be written as

$$exp\{-\frac{1}{2}s_{c,lm}^2 a_l + s_{c,lm} b_{lm}\}$$

where

$$a_l = C_{cmb,l}^{-1} + \vec{e}^T B_l^{-1} \vec{e} \qquad b_{lm} = \vec{e}^T B_l^{-1} \vec{d}_{lm}$$

The fraction b_{lm}/a_l is nothing else but the bayesian ILC estimation. So that we simply have (note the normalization constant changed)

$$K'(2\pi)^{-N_f/2}|N_l|^{-1/2}(2\pi)^{-N_c/2}|C_l|^{-1/2} \times exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1} \vec{d}_{lm}\} \times$$

$$\times exp\{-\frac{1}{2}s_{c,lm}^2 (C_{cmb,l}^{-1} + \vec{e}^T B_l^{-1} \vec{e}) + s_{c,lm} \vec{e}^T B_l^{-1} \vec{d}_{lm}\} \times$$

$$\times [(2\pi)^{(N_c-1)/2}|R_l|^{-1/2} +$$

$$+ \sqrt{\frac{(2\pi)^{N_c-1}}{det R_l}} \times$$

$$\times exp\{\frac{1}{2}\sum_{i,j=1}^{N_c-1}(R_l^{-1})_{ij}\frac{\partial}{\partial s_{lm,i}}\frac{\partial}{\partial s_{lm.j}}\}(\prod_{i=1}^{N_c-1}[1+\sum_{k=1}^{\infty}a_{ik}H_k(s_i)]-1)|_{\vec{s}_{lm}=R_l^{-1}\vec{b}_{lm}}] \tag{4.20}$$

4.6 MULTIDIMENSIONAL NEARLY GAUSSIAN

We relax the hypothesis of isotropy for the components. There's still independence bewteen frequencies and between the various components. We should point out the fact that this type of problems of component separation also depends on how we represent and manipulate our variables. So, when trying to face them it's a good idea to find a suitable representation of multivariate data.

4.6.1 *Representation of variables*

We will suppose to have an experiment with $l_{min} \leq l \leq l_{max}$(based on the resolution of the experiment).

The 'noise' covariance matrix is

$$\mathcal{N} = \begin{pmatrix} . & & & \\ & N_l & & \\ & & . & \\ & & & . \\ & & & & . \end{pmatrix} \tag{4.21}$$

Each N_l is repeated $2l+1$ times. The total number of matrices in $\mathcal{N}$ is $\sum_{l=l_{min}}^{l_{max}}(2l+1) = k_{tot}$.
Because N_l are diagonal also this is diagonal, so that

$$\mathcal{N}^{-1} = \begin{pmatrix} . & & & \\ & N_l^{-1} & & \\ & & . & \\ & & & . \\ & & & & . \end{pmatrix} \tag{4.22}$$

Representation of the components of the signal

$$\vec{s} = \vec{s}_c + \vec{s}_f \tag{4.23}$$

This is

$$\vec{s} = \begin{pmatrix} . \\ \vec{s}_{lm} \\ . \\ . \\ . \end{pmatrix} = \begin{pmatrix} . \\ \vec{s}_{cmb,lm} + \vec{s}_{f,lm} \\ . \\ . \\ . \end{pmatrix}$$

For each $\vec{s}_{lm}$ we will have a mixing matrix A. We will suppose A to be uniform on the entire celestial sphere[7]. So that, based on the previous sections we obtain the following vector of dimension N_f

$$\vec{d}_{lm} = A\vec{s}_{lm} + \vec{n}_{lm}$$

Or ($\mathbb{I}_{N_f}$ is the $N_f \times N_f$ matrix identity)

$$\vec{n}_{lm} = \mathbb{I}_{N_f}\vec{d}_{lm} - A\vec{s}_{lm} \tag{4.24}$$

Now if we want to create a maxi vector that contains all the data

7 (and constant in time)

$$\begin{pmatrix} \mathbb{I}_{N_f} & & & & \\ & . & & & \\ & & \mathbb{I}_{N_f} & & \\ & & & . & \\ & & & & \mathbb{I}_{N_f} \end{pmatrix} \begin{pmatrix} . \\ \vec{d}_{lm} \\ . \\ . \\ . \end{pmatrix} - \begin{pmatrix} A & & & & \\ & . & & & \\ & & A & & \\ & & & . & \\ & & & & A \end{pmatrix} \begin{pmatrix} . \\ \vec{s}_{lm} \\ . \\ . \\ . \end{pmatrix} =$$

$$= \vec{n} = \vec{d} - \mathcal{A}\vec{s} \tag{4.25}$$

We are simply creating a mapping between a sphere with a grid and a linear vector space. Note that the order of the components in the vectors is not important. But this one has a simple representation of the matrices.

$\mathcal{A}$ is $(kN_f) \times (kN_c)$. $\vec{s}$ is a kN_c vector.

Now let's see what are our priors.

Prior distribution for $\vec{n}$, $g(\vec{n})$ which gives us $g(\vec{d} - \mathcal{A}\vec{s})$ is

$$\frac{1}{(2\pi)^{(kN_f)/2}} \frac{1}{|\mathcal{N}|} exp\{-\frac{1}{2}(\vec{d} - \mathcal{A}\vec{s})^T \mathcal{N}^{-1}(\vec{d} - \mathcal{A}\vec{s})\} \tag{4.26}$$

We would want to do the same for the components. In the previous sections we focused on an (l, m) so that we had a covariance matrix at fixed l, say C_l. Now relaxing the hypothesis of statistical isotropy, we have a non diagonal covariance matrix $C_{lm,l'm'}$.

Taking a single component and suppose we have a Gaussian Distribution for it:

$$\frac{1}{(2\pi)^{k/2}} \frac{1}{|C_i|} exp\{-\frac{1}{2}\vec{s_i}^T C_i^{-1} \vec{s_i}\} =$$

$$= \frac{1}{(2\pi)^{k/2}} \frac{1}{|C_i|} exp\{-\frac{1}{2} \sum_{\{lm\}} s_{i,lm} C_{i,lm,l'm'} s_{i,l'm'}\}$$

where $\vec{s}_i$ is a vector of dimension k which contains all the lms for a single component.

If we want to create a Gaussian of statistically independent components each one with its Gaussian(for N_c components)

$$\prod_i \frac{1}{(2\pi)^{k/2}} \frac{1}{|C_i|} exp\{-\frac{1}{2}\vec{s_i}^T C_i^{-1} \vec{s_i}\} =$$

$$= \prod_i \frac{1}{(2\pi)^{k/2}} \frac{1}{|C_i|} exp\{-\frac{1}{2} \sum_{\{lm\}} s_{i,lm} C_{i,lm,l'm'} s_{i,l'm'}\}$$

Or

$$exp\{-\sum_i \frac{1}{2} \vec{s_i}^T C_i^{-1} \vec{s_i}\} \prod_i \frac{1}{(2\pi)^{k/2}} \frac{1}{|C_i|}$$

We now define $\vec{s}'$ as the vector whose components are

$$\vec{s}' = \begin{pmatrix} . \\ \vec{s}_{i,k} \\ . \\ . \\ . \end{pmatrix} = \begin{pmatrix} \vec{s}_{cmb,k} \\ . \\ \vec{s}_{j,k} \\ . \\ . \end{pmatrix} = \vec{s}'_{cmb} + \vec{s}'_f$$

where $\vec{s}_{i,k}$ is the vector of dimension k which contains all the lms for the single component i (now with a pedix k for clarity)[8].

The block 'covariance matrix' for the vector $\vec{s}'$ is

$$\mathcal{C} = \begin{pmatrix} C_{cmb} & & & & \\ & . & & & \\ & & C_j & & \\ & & & . & \\ & & & & . \end{pmatrix} \tag{4.27}$$

And for the properties of block diagonal matrices (each one of dimension k)

$$\mathcal{C}^{-1} = \begin{pmatrix} C_{cmb}^{-1} & & & & \\ & . & & & \\ & & C_j^{-1} & & \\ & & & . & \\ & & & & . \end{pmatrix} =$$

8 While before we had $\vec{s}_{lm}$ containing all components at the single lm.

$$= \begin{pmatrix} C_{cmb}^{-1} & & & \\ & \mathbb{O}_k & & \\ & & . & \\ & & & \mathbb{O}_k \\ & & & & . \end{pmatrix} + \begin{pmatrix} \mathbb{O}_k & & & \\ & . & & \\ & & C_j^{-1} & \\ & & & . \\ & & & & . \end{pmatrix} \equiv \mathcal{D}_c + \mathcal{D}_f \tag{4.28}$$

So that we obtain as Gaussian distribution

$$\frac{1}{(2\pi)^{(N_c k)/2}} \frac{1}{|\mathcal{C}|} exp\{-\frac{1}{2}\vec{s}'^T \mathcal{C}^{-1} \vec{s}'\} \tag{4.29}$$

If we wanted a nearly Gaussian distribution this would be

$$q(\vec{s}') = \frac{1}{(2\pi)^{(kN_c)/2}} \frac{1}{|\mathcal{C}|} exp\{-\frac{1}{2}\vec{s}'^T \mathcal{C}^{-1} \vec{s}'\} \times f(\vec{s}'_f) \tag{4.30}$$

where the first factor is a multivariate guassian while the second belongs to some non gaussian correction. This could be for example from the EdgeWorth Expansion.

Now we create a link between $\vec{s}$ and $\vec{s'}$ by a linear transformation

$$\mathcal{Q} : \vec{s'} \to \vec{s} \tag{4.31}$$

or $\vec{s} = \mathcal{Q}\vec{s'}$, so that we can mix the various components using the mixing matrix.

Then (30) is the same, while (26) is now

$$\frac{1}{(2\pi)^{(kN_f)/2}} \frac{1}{|\mathcal{N}|} exp\{-\frac{1}{2}(\vec{d} - \mathcal{A}\mathcal{Q}\vec{s'})^T \mathcal{N}^{-1} (\vec{d} - \mathcal{A}\mathcal{Q}\vec{s'})\} \tag{4.32}$$

We rename $\vec{s'}$ as $\vec{s}$and $\mathcal{A}\mathcal{Q} = \mathcal{A}'$:

4.6.2 *Calculation*

The full pdf is very similar to the single (l, m) case:

$$\frac{1}{(2\pi)^{(kN_f)/2}} \frac{1}{|\mathcal{N}|} exp\{-\frac{1}{2}(\vec{d} - \mathcal{A}'\vec{s})^T \mathcal{N}^{-1} (\vec{d} - \mathcal{A}'\vec{s})\} \frac{1}{(2\pi)^{(kN_c)/2}} \frac{1}{|\mathcal{C}|} exp\{-\frac{1}{2}\vec{s}^T \mathcal{C}^{-1} \vec{s}\} \times f(\vec{s}_f) \tag{4.33}$$

Or

$$\frac{1}{(2\pi)^{(kN_f)/2}}\frac{1}{|\mathcal{N}|}exp\{-\frac{1}{2}\vec{d}^T\mathcal{N}^{-1}\vec{d}+\vec{s}^T\mathcal{A}'^T\mathcal{N}^{-1}\vec{d}-\frac{1}{2}\vec{s}^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}\}\times$$

$$\times\frac{1}{(2\pi)^{(kN_c)/2}}\frac{1}{|\mathcal{C}|}exp\{-\frac{1}{2}\vec{s}_c^T\mathcal{D}_c\vec{s}_c\}\times exp\{-\frac{1}{2}\vec{s}_f\mathcal{D}_f\vec{s}_f\}\times f(\vec{s}_f) \quad (4.34)$$

Now let's expand the argument of the exponential on the first line of (34).

$$\vec{s}_c^T\mathcal{A}'^T\mathcal{N}^{-1}\vec{d}+\vec{s}_f^T\mathcal{A}'^T\mathcal{N}^{-1}\vec{d}-\frac{1}{2}\vec{s}_c^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_c-\frac{1}{2}\vec{s}_f^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f-\vec{s}_c^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f$$

So that we now have

$$\frac{1}{(2\pi)^{(kN_f)/2}}\frac{1}{|\mathcal{N}|}\times\frac{1}{(2\pi)^{(kN_c)/2}}\frac{1}{|\mathcal{C}|}\times exp\{-\frac{1}{2}\vec{d}^T\mathcal{N}^{-1}\vec{d}\}\times$$

$$\times exp\{-\frac{1}{2}\vec{s}_c^T\mathcal{D}_c\vec{s}_c\}\times exp\{\vec{s}_c^T\mathcal{A}'^T\mathcal{N}^{-1}\vec{d}\}\times exp\{-\frac{1}{2}\vec{s}_c^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_c\}\times$$

$$\times exp\{-\frac{1}{2}\vec{s}_f^T\mathcal{D}_f\vec{s}_f+\vec{d}^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f-\frac{1}{2}\vec{s}_f^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f-\vec{s}_c^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f\}\times f(\vec{s}_f) \quad (4.35)$$

Now we want to marginalize on the variables on which we are not interested. In other words we will focus on the last line of the last formula and marginalize over the foregrounds

$$exp\{-\frac{1}{2}\vec{s}_f^T\mathcal{D}_f\vec{s}_f+\vec{d}^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f-\frac{1}{2}\vec{s}_f^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f-\vec{s}_c^T\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}'\vec{s}_f\}\times f(\vec{s}_f)$$

First let's remember that $\vec{s}_f$ contains the foreground components without the CMB. So, the first k columns are all zero. Then this vector can be connected to a vector of dimension $(N_c-1)\times k$ that contains all the foregrounds signals. Let's call this vector $\vec{f}_f$. Now in the last formula line we can recast all the addends in this form

$$exp\{-\frac{1}{2}\vec{s}_f^T\mathcal{V}\vec{s}_f+\vec{v}^T\vec{s}_f\}$$

Where

$$\mathcal{V}=(\mathcal{D}_f+\mathcal{A}'^T\mathcal{N}^{-1}\mathcal{A}')$$

and

$$\vec{v} = \mathcal{A}'^T \mathcal{N}^{-1} \vec{d}^T - \mathcal{A}'^T \mathcal{N}^{-1} \mathcal{A}' \vec{s}_c$$

But for what we've just said $\vec{s}_f^T \mathcal{V} \vec{s}_f = \vec{f}_f^T \mathcal{M} \vec{f}_f$ for some submatrix $\mathcal{M}$ of $\mathcal{V}$.

Now we name the last $\mathcal{Q} = (N_c - 1) \times k$ entries of $\vec{v}$ with $\vec{a}$.

Before this we rewrite it in a another form(assume $\mathcal{M}$ invertible)

$$exp\{-\frac{1}{2}\vec{f}_f^T \mathcal{M} \vec{f}_f + \vec{a}^T \vec{f}_f\} = exp\{-\frac{1}{2}\vec{a}^T \mathcal{M}^{-1} \vec{a}\} \times exp\{-\frac{1}{2}(\vec{s}_f - \mathcal{M}^{-1}\vec{a})^T \mathcal{M} (\vec{s}_f - \mathcal{M}^{-1}\vec{a})\}$$

We have to marginalize[9]

$$\int d\vec{f}_f exp\{-\frac{1}{2}(\vec{s}_f - \mathcal{M}^{-1}\vec{a})^T \mathcal{M} (\vec{s}_f - \mathcal{M}^{-1}\vec{a})\} f(\vec{f}_f) \tag{4.36}$$

With a shift $\vec{t}_f = \vec{f}_f - \mathcal{M}^{-1}\vec{a}$ this transforms as

$$\int d\vec{t}_f exp\{-\frac{1}{2}\vec{t}_f^T \mathcal{M} \vec{t}_f\} \times g(\vec{t}_f) \tag{4.37}$$

where $g(\vec{t}_f) = f(\vec{t}_f + \mathcal{M}^{-1}\vec{a})$. Then we use again the 'master' formula to obtain

$$\sqrt{\frac{(2\pi)^{\mathcal{Q}}}{det\mathcal{M}}} exp\{\frac{1}{2}\sum_{i,j=1}^{N_c-1} (\mathcal{M}^{-1})_{ij} \frac{\partial}{\partial t_{f,i}} \frac{\partial}{\partial t_{f,j}}\} g(\vec{t}_f)|_{\vec{t}_{lm}=0} \tag{4.38}$$

The final posterior distribution function the CMB signal component is

$$\frac{1}{(2\pi)^{(kN_f)/2}} \frac{1}{|\mathcal{N}|} \times \frac{1}{(2\pi)^{(kN_c)/2}} \frac{1}{|\mathcal{C}|} \times exp\{-\frac{1}{2}\vec{d}^T \mathcal{N}^{-1} \vec{d}\} \times \sqrt{\frac{(2\pi)^{\mathcal{Q}}}{det\mathcal{M}}} \times$$

$$\times exp\{\frac{1}{2}\sum_{i,j=1}^{N_c-1} (\mathcal{M}^{-1})_{ij} \frac{\partial}{\partial t_{f,i}} \frac{\partial}{\partial t_{f,j}}\} g(\vec{t}_f)|_{\vec{t}_{lm}=0} \times$$

$$\times exp\{-\frac{1}{2}\vec{s}_c^T \mathcal{D}_c \vec{s}_c\} \times exp\{\vec{s}_c^T \mathcal{A}'^T \mathcal{N}^{-1} \vec{d}\} \times exp\{-\frac{1}{2}\vec{s}_c^T \mathcal{A}'^T \mathcal{N}^{-1} \mathcal{A}' \vec{s}_c\} \quad , \tag{4.39}$$

what all we need for a Bayesian treatment.

9 $f(\vec{f}_f) = f(\vec{s}_f)$

5

LINES FOR THE SIMULATION

Our main aim is to create a complete CMB cleaned from the foregrounds contamination. In our way we have to calculate the final pdf for the CMB signal, seen in the last part of the previous chapter, and do a full Bayesian analysis. Also trying to include only the three point function, we would require long calculations. This requires performing long calculations. So what we can do as a prestep is to reduce the problem to a 1 dimensional one and see if our method works. So we will imagine to have only one dimensional signal of interest and one contaminating signal. Then we try to see if our method performs better than bayesian ILC.

5.1 1 CMB(COMPONENT OF INTEREST)+1 FOREGROUND GIVEN N_f FREQUENCIES

Let's suppose for a moment to have $N_c - 1$ foregrounds, where N_c is the total number of components. Because in cosmology usually we deal with the skewness for non Gaussian distributions we will start from the following expression in which the edgeworth expansion includes only the skewness

$$K(2\pi)^{-N_f/2}|N_l|^{-1/2}exp[-\frac{1}{2}(\vec{d}_{lm} - A\vec{s}_{lm})^T N_l^{-1}(\vec{d}_{lm} - A\vec{s}_{lm})]\times$$

$$\times(2\pi)^{-N_c/2}|C_l|^{-1/2}exp[-\frac{1}{2}\vec{s}_{lm}^T C_l^{-1}\vec{s}_{lm}]\times$$

$$\times\prod_{i=1}^{N_c-1}(1+\gamma_i H_3(\frac{s_{i,lm}}{\sigma_i})) \tag{5.1}$$

$\gamma_i = \frac{\sigma_i S_{i,3}}{6}$ with $S_{i,3} = \frac{\kappa_{i,3}}{\sigma_i^4}$, so that $\gamma_i = \frac{\kappa_{i,3}}{6\sigma_i^3}$. We suppose $S_{i,3} << 1$, then $\gamma_i << 1$. Then we can replace the product in (1) with a sum

$$K(2\pi)^{-N_f/2}|N_l|^{-1/2}exp[-\frac{1}{2}(\vec{d}_{lm} - A\vec{s}_{lm})^T N_l^{-1}(\vec{d}_{lm} - A\vec{s}_{lm})]\times$$

$$\times(2\pi)^{-N_c/2}|C_l|^{-1/2}exp[-\frac{1}{2}\vec{s}_{lm}^T C_l^{-1}\vec{s}_{lm}]\times$$

$$(1 + \sum_{i=1}^{N_c-1} \gamma_i H_3(\frac{s_{i,lm}}{\sigma_i})) \tag{5.2}$$

$$K(2\pi)^{-N_f/2}|N_l|^{-1/2}(2\pi)^{-N_c/2}|C_l|^{-1/2} \times exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm}\}\times$$

$$\times exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1} A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^T C_l^{-1}\vec{s}_{c,lm}\}\times$$

$$\times[exp\{-\frac{1}{2}\vec{b}_{lm}{}^T R_l^{-1}\vec{b}_{lm}\}(2\pi)^{(N_c-1)/2}|R_l|^{1/2}+$$

$$+ \sum_{j=1}^{N_c-1}\int d\vec{s}_{f,lm}\prod_{i=1}^{N_c-1} exp\{-\frac{1}{2}s_{i,lm}^2(R_l)_{i,i} + s_{i,lm}c_i\}\gamma_j H_3(s_{j,lm})] \tag{5.3}$$

For $N_c = 2$, (3) is

$$[-\frac{1}{2}s_{1,lm}^2(\Lambda_l)_{2,2} + s_{1,lm}(\vec{a}_{lm})_2]$$

with

$$\Lambda_l = \begin{pmatrix} C_{cmb,l}^{-1} & 0 \\ 0 & \sum_{k=1}^{N_f} a_{k1}^2\nu_{k,l}^{-2} + C_{1,l}^{-2} \end{pmatrix} = \begin{pmatrix} C_{cmb,l}^{-1} & 0 \\ 0 & \lambda_1 \end{pmatrix}$$

The marginalization is

$$\int d(\begin{matrix} 0 \\ s_{1,lm} \end{matrix}) exp\{-\frac{1}{2}s_{1,lm}^2(\Lambda_l)_{2,2} + s_{1,lm}(\vec{a}_{lm})_2\}\gamma_1 H_3(\frac{s_{1,lm}}{\sigma_1})$$

Written in another way is (we use the expression for the hermite polynomial of order 3)

$$\int ds\ exp\{-\frac{1}{2}s^2 a + sb\}\gamma(\frac{s^3}{\sigma_1^3} - 3\frac{s}{\sigma_1}) =$$

$$= \int ds\ exp\{-\frac{1}{2}s^2 a + sb\}\gamma H_3(s/\sigma) =$$

$$= \int \sigma ds'\ exp\{-\frac{1}{2}\sigma^2 s'^2 a + s'\sigma b\}\gamma H_3(s')$$

$$= \frac{\gamma'\sqrt{2\pi}b'e^{b'^2/(2a')}}{a'^{7/2}}(3a' + b'^2 - 3a'^2) =$$

$$= \frac{\gamma\sqrt{2\pi}be^{b^2/(2a)}}{\sigma^3 a^{7/2}}(3a + b^2 - 3\sigma^2 a^2)$$

where

$$\gamma' = \gamma\sigma \;\; a' = a\sigma^2 \;\; b' = b\sigma$$

With $a = (\Lambda_l)_{2,2} = \sum_{k=1}^{N_f} a_{k1}^2 \nu_{k,l}^{-2} + C_{1,l}^{-2}$, $b = (\vec{a}_{lm})_2 = \sum_{k=1}^{N_f} a_{k1}\nu_{k,l}^{-2} y_{lm,k}$, where $y_{lm,k} = (\vec{d}_{lm})_k - s_{cmb,l}$ and $\gamma = \gamma_1$. This is because

$$\vec{a}_{lm} = A_f^T N_l^{-1}(\vec{d}_{lm} - A_c\vec{s}_{c,lm}) = A_f^T N_l^{-1}(\vec{d}_{lm} - \vec{e}s_{cmb,lm})$$

And

$$A_f = \begin{pmatrix} 0 & a_{11} \\ . & . \\ . & . \\ . & . \\ 0 & a_{N_f 1} \end{pmatrix} = \begin{pmatrix} 0 & \\ . & \\ . & \vec{a} \\ . & \\ 0 & \end{pmatrix}$$

So that

$$\vec{a}_{lm} = \begin{pmatrix} 0 \\ \sum_{k=1}^{N_f} a_{k1}\nu_{k,l}^{-2} y_{lm,k} \end{pmatrix} = \begin{pmatrix} 0 \\ \vec{a}^T N_l^{-1}\vec{y}_{lm} \end{pmatrix} = \begin{pmatrix} 0 \\ \vec{a}^T N_l^{-1}\vec{d}_{lm} - \vec{a}^T N_l^{-1}\vec{e}s_{cmb,lm} \end{pmatrix} = \begin{pmatrix} 0 \\ a_{1,lm} \end{pmatrix}$$

$a = \lambda_1$, $b = a_{1,lm} = a_1 = d - es_{cmb}$

$$p(s_c) = K(2\pi)^{-N_f/2}|N_l|^{-1/2}(2\pi)^{-N_c/2}|C_l|^{-1/2} \times exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm}\}\times$$

$$\times exp\{(A_c\vec{s}_{c,lm})^T N_l^{-1}\vec{d}_{lm} - \frac{1}{2}(A_c\vec{s}_{c,lm})^T N_l^{-1}A_c\vec{s}_{c,lm} - \frac{1}{2}\vec{s}_{c,lm}^T C_l^{-1}\vec{s}_{c,lm}\}\times$$

$$\times exp\{-\frac{1}{2}a_{1,lm}^2\lambda_1^{-1}\}[(2\pi)^{1/2}|\lambda_l|^{-1/2}+$$

$$+\frac{\gamma\sqrt{2\pi}a_1}{\sigma^3\lambda_1^{7/2}}(3\lambda_1 + {a_1}^2 - 3\sigma^2{\lambda_1}^2)] \qquad (5.4)$$

$$p(s_c) = K'(2\pi)^{-N_f/2}|N_l|^{-1/2}(2\pi)^{-N_c/2}|C_l|^{-1/2} \times exp\{-\frac{1}{2}\vec{d}_{lm}^T N_l^{-1}\vec{d}_{lm}\}\times$$

$$\times exp\{-\frac{1}{2}s_{c,lm}^2(C_{cmb,l}^{-1} + \vec{e}^T B_l^{-1}\vec{e}) + s_{c,lm}\vec{e}^T B_l^{-1}\vec{d}_{lm}\}\times$$

$$\times(2\pi)^{1/2}|\lambda_l|^{-1/2}[1+$$

$$+\frac{\gamma a_1}{\sigma^3\lambda_1^3}(3\lambda_1 + {a_1}^2 - 3\sigma^2{\lambda_1}^2)] \quad , \tag{5.5}$$

where $B_l = N_l + \vec{a}C_l\vec{a}^T$.

This is our final probability distribution that we will use for estimating our signal. As criterion of estimation we will maximize $p(s_c)$. This is the same criterion used for Bayesian ILC.

5.1.1 *The Simulation*

We want to test our method of using EdgeWorth expansion. The idea is that we have some foreground signal generated by a nearly Gaussian pdf, with known skewness. Then we have our Gaussian pdf and a Gaussian noise.

So, the idea is to generate a realization of the signal, say composed by N samples, imagining to take measure at 'time' t_i with $i = 1, .., N$. For each one of this N samples we make an estimation of the true signal, once with Bayesian ILC and the other one with Non Gaussian Bayesian ILC (ILCNG).

5.1.1.1 *Evaluating Performance*

Now, we would want to check if with NILC we could improve ILC. To do so we will run a test based on the sum of the squared differences of the estimations minus the true signal. In other words, for a realization we will calculate the sum $S = \sum_i(\hat{s}_i - s_i)^2$ where s_i is the true value for the $i-$th sample and $\hat{s}_i$ its estimation, $i = 1, .., N$. Then we will say that the method that performs better is that with the smaller S.

5.1.1.2 *In Practice*

To perform the simulation we wrote a program in Python, using mainly the numpy library that allows to handle quickly array of data. Here the main steps are listed:

Creation of a 'realization'(suppose a signal in which we take samples at times $t_1, t_2, ..., t_N$):

1. Generate a 'CMB' signal $s(t_1), ..., s(t_n)$ from a $\mathcal{N}(0, \sigma_c)$, one 'foreground' signal $f(t_1), ..., f(t_N)$ from $\mathcal{NG}(0, \sigma_f, skew_f)$.
2. Random noise $n(t_1), ..., n(t_n)$ from $\mathcal{N}(0, N)$(where N is the covariance for the noise) for N_f channels.
3. Calculatedata $d = A(s, f) + n$
4. Use d for:
   a) Bayesian ILC
   b) Numerical Maximum for ILCNG
5. Repeat 2-4 for several times, taking 1 fixed[1].
6. Check how many times ILCNG performs better with the test $\sum_i (\hat{s}_{i,x} - s_{i,t})^2$, where $ILC, ILCNG$
7. Do steps 1-5 several times with different parameters and see how the results change.

5.1.2 *The Proper Simulation*

As our non Gaussian function we took a Chi-Squared distribution with k degrees of freedom

$$\chi(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \tag{5.6}$$

where $\Gamma(n)$ is the Euler gamma function. This pdf has the following values for mean, variance, and skewness respectively

$$k, \quad 2k, \quad \sqrt{\frac{8}{k}} \quad .$$

The good thing about it is the fact that it is possible to use only one parameter that describes this 3 cumulants. Furethermore, it has the property that for $k >> 1$ it tends to a Gaussian.

The Edgeworth function with only the Hermite polynomial of order three is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2}\frac{x^2}{\sigma^2})(1 + \gamma H_3(\frac{x}{\sigma})) \tag{5.7}$$

1 For this it would be better to use another term intead of realization, because usually this is fixed by s and f.

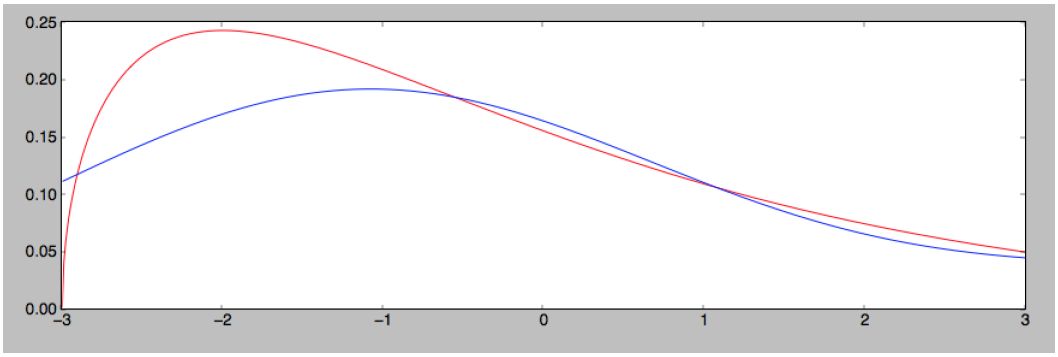


Figure 5.1: $k = 3$. Red is the Chi distribution, blue the Edgeworth one.

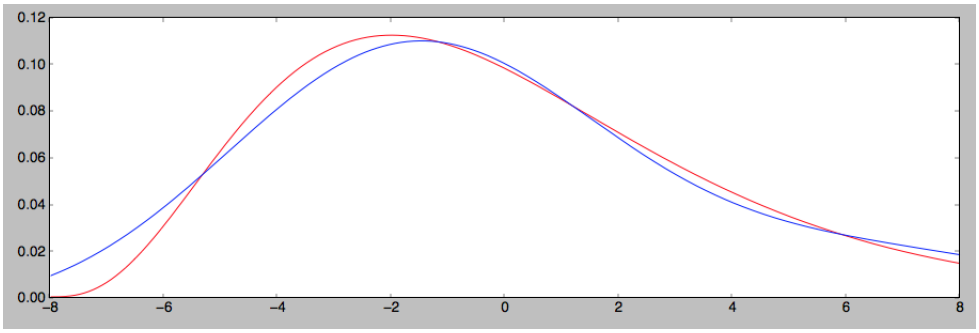


Figure 5.2: $k = 8$

where $\gamma = Skewness/6$. Figures 1-4 have some examples of Chi-Squared distribution and the EdgeWorth that tries to approximate it.

Once we had all the parameter at hand we did the simulation. In Figures 5 and 6 there are the posterior marginalized probability distributions for the case $k = 8$.

We performed 1000 realizations, for a fixed value of k, each one with 1000 samples with 4 channels. We found systematically that for smaller ks, approximately smaller than 10, ILCNG performes better than Bayesian ILC. For example for $k = 8$ Figure 7 shows great improvement.

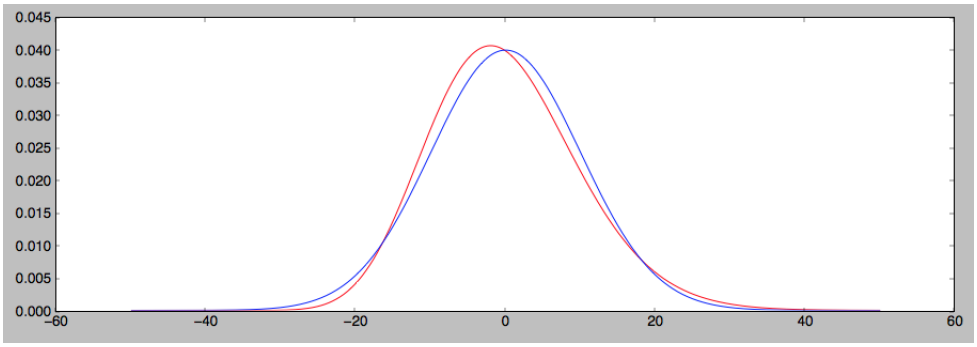


Figure 5.3: $k = 50$

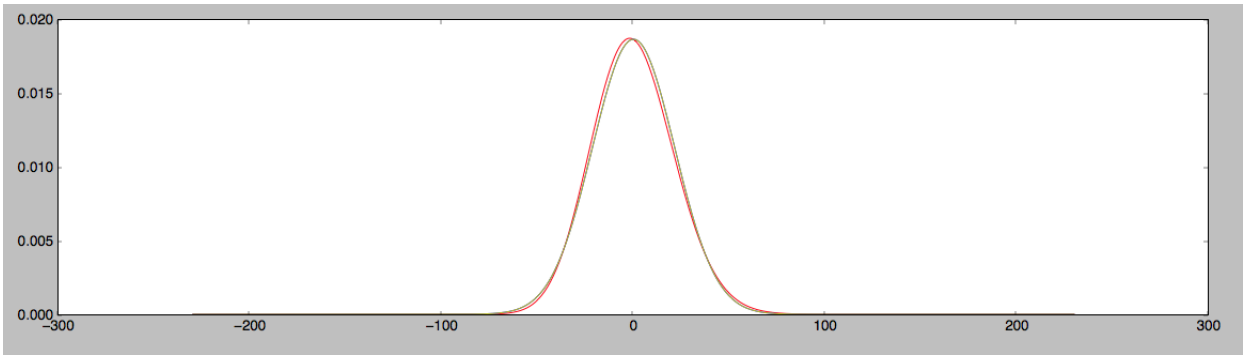


Figure 5.4: $k = 230$. Now there is in yellow a Gaussian with zero mean and same variance of the Chi.

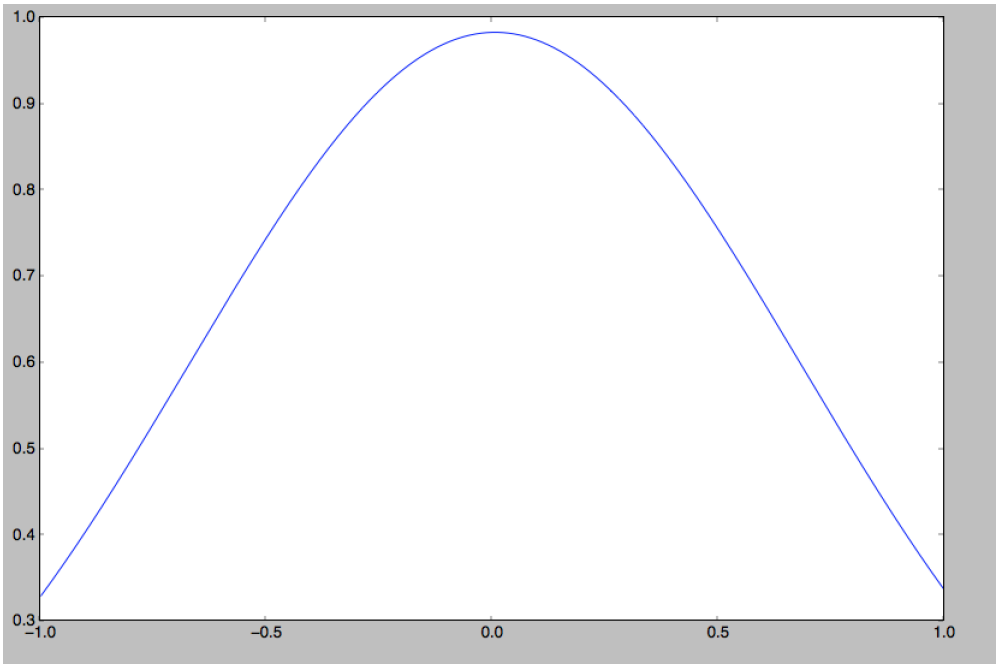


Figure 5.5: Example of the (unnormalized) posterior $p(s_c)$ with parameters taken from a simulation with a Chi-Square distribution with 8 degrees of freedom.

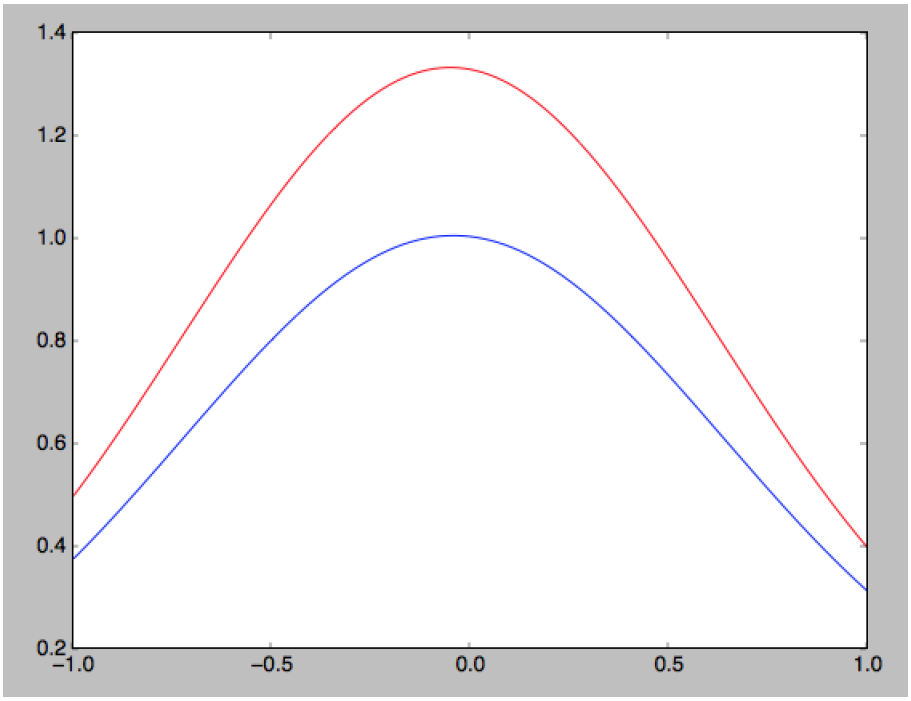


Figure 5.6: Unnormalized posterior $p(s_c)$ in red and unnormalized posterior of Bayesian ILC (Gaussian) in blue for the same Chi-Square of Figure 1.

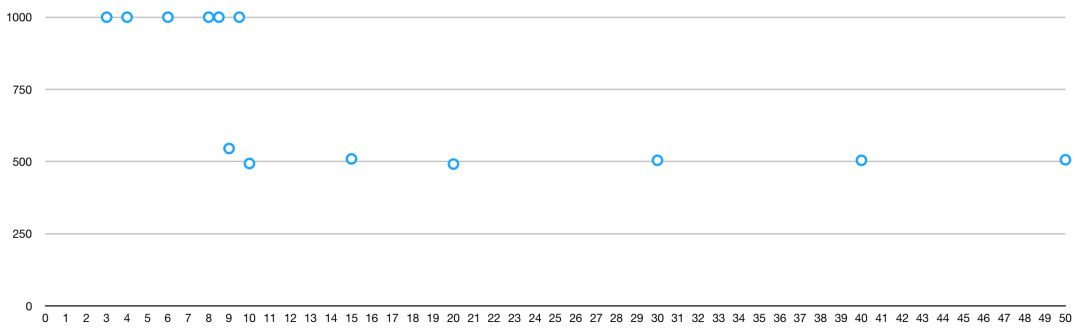


Figure 5.7: Number of times ILC leads to an S greater than ILCNG in function of k. ILCNG performes better for $k < 10$. The two methods are very similar for the other cases.

6 CONCLUSION

In this work we presented a new calculation for foreground cleaning, expanding what was did in [25], where a Bayesian Approach was followed to do the estimation of the CMB signal. What we did was to modify the Bayesian ILC assumptions where all the prior signals are supposed to be Gaussian. But in reality, apart from the CMB and in good approximation the noise, foreground signals have more complex distributions.

In CMB maps we measure a little residual non Gaussian component, which as now we believe is mainly to sources different than the CMB. Foregrounds could be linked to this non Gaussianity that we measure in terms of a non zero bispectrum. Substracting this components from CMB maps would improve our knowledge about the physics governing our Universe. Indeed, we saw that foregrounds are a problem not only in CMB temperature maps but also in the polarization ones. We saw that in particular foregrounds are a big problem when we try to have access to the CMB $B-$mode polarization signal, signal that could shed light on our theories about the primordial Universe. From this, the importance of removing contaminations in the best way based on our assumptions.

Based on the prior knowledge of foregrounds having nearly Gaussian distributions we took a Bayesian approach to estimate the CMB. We said that we measure the bispectrum as extra information to put in our prior. There are infinite distributions that can have the same power spectrum and bispectrum. So how do we choose our nearly Gaussian distribution? We took an Edgeworth expansion, that is a very good approximation for several nearly Gaussian distributions. With an Edgeworth expansion we can do a truncation and use only the two and three point harmonic correlation function to define a unique distribution. With the Bayesian approach we used this as our prior for foregrounds and we found a posterior probability that define completely the CMB signal[1].

We then applied our results to a 1 point-pdf case (toy model), imagining a general signal processing problem, with the example to a Chi-Squared distributed 'foreground'. We saw that using the Chi-Squared skewnees in the Edgeworth expansion led to a better result with respect to the Bayesian ILC. This is also natural: we used extra information, with respect to a completely Gaussian case, and this led to a better result.

1 Furthermore, in the way we did it we could also include higher order correlation functions.

The fact that the 1-pdf case worked doesn't mean that the multivariate one will do the same. Indeed, in the problem will enter cross correlations and we will also have to take in case that if we want to mantain isotropy with a Non Gaussian pdf the a_{lm}s will become dependent. Furethermore, one drawback of any Bayesian approach is that our estimations are biased, because in general our priors are never accurate(and this is more true for foregrounds, the physics of which we still don't know very well).

We also made the assumption that the foreground signals have values that belongs to all $\mathbb{R}$. This is not true

So, we showed that the NGILC methods do better than ILC in the 1 point pdf case when the underlying foreground distribution is nearly Gaussian. The future step will be to apply our results to a general $N-$dimensional case. We have the full posterior pdf. So, the next step is to do simulations and then use real CMB data to, hopefully, improve the existing maps.

$45Users_{o}mardarwish_{D}ropbox_{T}ESI_{M}AGISTRALE_{T}E_{s-LyX2_{2}-v4_{3}pi_{b}iblatex_{b}iber_{B}ibliography}$

BIBLIOGRAPHY

[1] L. Raul Abramo and Thiago S. Pereira. "Testing Gaussianity, Homogeneity, and Isotropy with the Cosmic Microwave Background." In: *Hindawi Publishing Corporation Advances in Astronomy* (2010) (cit. on p. 15).

[2] Daniel Babich. "Optimal Estimation of Non-Gaussianity." In: *https://arxiv.org/abs/astro-ph/0503375* (2005) (cit. on p. 20).

[3] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro. "Large-Scale Structure of the Universe and Cosmological Perturbation Theory." In: *https://arxiv.org/abs/astro-ph/0112551* (2001) (cit. on pp. 9, 15).

[4] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006 (cit. on p. 15).

[5] S. Blinnikov and R. Moessner. "Expansions for nearly Gaussian distributions." In: (1997) (cit. on pp. 15, 17).

[6] Gianluca Calcagni. *Classical and Quantum Cosmology*. Springer, Graduate Texts in Physics, 2017 (cit. on pp. 24, 25).

[7] Anthony Challinor. "CMB anisotropy science: a review." In: (2012) (cit. on p. 38).

[8] Planck Collaboration. "Planck 2015 results. XIII. Cosmological parameters." In: *https://arxiv.org/abs/1502.01589* (2015) (cit. on p. 25).

[9] J. Delabrouille and J.-F. Cardoso. "Diffuse Source Separation in CMB Observations." In: *https://arxiv.org/abs/astro-ph/0702198* (2007) (cit. on pp. 33–36, 40).

[10] Clive Dickinson. "CMB foregrounds - A brief overview." In: *https://arxiv.org/pdf/1606.03606.pdf* (2016) (cit. on pp. 35, 36).

[11] Scott Dodelson. *Modern Cosmology*. Academic Press, 2003 (cit. on pp. 24–26).

[12] Wayne Hu. "Wandering in the Background: A Cosmic Microwave Background Explorer." In: (1995) (cit. on p. 21).

[13] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. Wiley, 2001 (cit. on p. 39).

[14] Kiyotomo Ichiki. "CMB foreground: A concise review." In: *Progress of Theoretical and Experimental Physics, Volume 2014, Issue 6, 1 June 2014, 06B109, https://doi.org/10.1093/ptep/ptu065* () (cit. on p. 40).

[15] Robert D. Klauber. "Student Friendly Guide to the Cosmic Microwave Background." In: *http://www.quantumfieldtheory.info/CMB.pdf* (2015) (cit. on p. 11).

[16] Eiichiro Komatsu. "The Pursuit of Non-Gaussian Fluctuations in the Cosmic Microwave Background." In: *http://wwwmpa.mpa-garching.mpg.de/ komatsu/phdthesis.html* (2001) (cit. on pp. 11, 12).

[17] Eiichiro Komatsu. "Non-Gaussianity." In: *http://wwwmpa.mpa-garching.mpg.de/ komatsu/cmb/lecture$_N$G$_i$ucaa$_2$011.pdf* (2011) (cit. on p. 18).

[18] Fabien Lacasa. "Non-Gaussianity and extragalactic foregrounds to the Cosmic Microwave Background." In: *https://arxiv.org/abs/1406.0441* (2014) (cit. on p. 6).

[19] Manzoor A. Malik. "Cosmology with the Cosmic Microwave Background." In: *http://article.sapub.org/pdf/10.5923.j.astronomy.20130202.01.pdf* (2013) (cit. on p. 30).

[20] Domenico Marinucci. "Testing for Non-Gaussianity on Cosmic Microwave Background Radiation: A Review Testing for Non-Gaussianity on Cosmic Microwave Background Radiation: A Review." In: *Statist. Sci. Volume 19, Number 2 (2004), 294-307. Statist. Sci. Volume 19, Number 2 (2004), 294-307.* (2004) (cit. on pp. 13, 22, 23).

[21] Yasin Memari. "Polarization of the Cosmic Microwave Background Radiation." In: *http://www.roe.ac.uk/ifa/postgrad/pedagogy/2007$_m$emari.pdf* (2007) (cit. on pp. 27, 28).

[22] Guido Walter Pettinari. *The Intrinsic Bispectrum of the Cosmic Microwave Background*. Springer Theses, 2016 (cit. on pp. 3, 7, 9).

[23] Cristiano Porciani. "Cosmolgical Perturbations." In: *bo.astro.it/ school/school09/Present* () (cit. on p. 8).

[24] Dodelson Scott and G.F. Smoot. "COSMIC MICROWAVE BACKGROUND." In: *http://pdg.lbl.gov/2011/reviews/rpp2011-rev-cosmic-microwave-background.pdf* (2011) (cit. on p. 23).

[25] Flavien Vansyngel, Benjamin D. Wandelt, Jean-François Cardoso, and Karim Benabed. "Semi-blind Bayesian inference of CMB map and power spectrum." In: *https://arxiv.org/abs/1409.0858* (2014) (cit. on pp. 43, 50, 69).

[26] Licia Verde. "Statistical methods in cosmology." In: *https://arxiv.org/abs/0911.3105* (2009) (cit. on pp. 3, 10).

[27] Licia Verde, Raul Jimenez, Luis Alvarez-Gaume, Alan F. Heavens, and Sabino Matarrese. "Multi-variate joint PDF for non-Gaussianities: exact formulation and generic approximations." In: *JCAP* (2013) (cit. on p. 19).

[28] Matías Vidal Navarro. *Diffuse Radio Foregrounds: All-Sky Polarisation, and Anomalous Microwave Emission*. Springer Theses, 2016 (cit. on pp. 23, 25, 27, 28).