Università degli Studi di Padova

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Dottorato in Astronomia

Non-Gaussianity in CMB analysis: bispectrum estimation and foreground subtraction

Supervisore: Dott. Michele Liguori
Co-Supervisore: Prof. Roberto Turolla
Coordinatore: Prof. Giampaolo Piotto

Valutatori: Prof. Patricio Vielva
Prof. Bartjan van Tent

Dottorando: Filippo Oppizzi

Corso di Dottorato in Astronomia

Tesi di Dottorato

Ciclo XXXI

"Even the astronomer cannot prevent himself from seeing the moon larger at its rising than some time afterwards"

- Immanuel Kant, *Critique of Pure Reason.*

Sommario

Il tema centrale di questa tesi è lo sviluppo di metodi statistici e numerici per lo studio di caratteristiche non gaussiane e/o anisotrope in esperimenti mirati alla misura della radiazione cosmica di fondo (CMB, dall'inglese *Cosmic Microwave Background*). Ci concentriamo su due tipi molto diversi di segnali non gaussiani: il primo è la *non Gaussianità primordiale*, che si ipotizza venga generata nell'Universo primordiale durante l'epoca inflazionaria. Lo studio di questo tipo di non-Gaussianità permette di ottenere preziose informazioni cosmologiche. Il secondo è invece la non-Gaussianità generata dalla contaminazione dovuta al *foreground* astrofisico. In questo caso, invece, il nostro obiettivo è utilizzare la non-Gaussianità come tracciante per identificare e rimuovere le componenti spurie non cosmologiche (ovviamente l'emissione di *foreground* contiene informazioni astrofisiche rilevanti, ma il tema di questa tesi verte sulla cosmologia, quindi verrà considerata solo in virtù dell'effetto contaminante in esperimenti che mirano a ricostruire la CMB).

Sforzi considerevoli sono stati spesi finora nel tentativo di misurare piccole deviazioni dalla Gaussianità nelle anisotropie della CMB, che fornirebbero informazioni inestimabili sull'epoca dell'*Inflazione*. La teoria prevede che l'Inflazione produca un campo di fluttuazioni isotropo e quasi Gaussiano. Tuttavia, una grande quantità di modelli prevede anche l'insorgenza di piccole componenti non Gaussiane, le cui caratteristiche dipendono fortemente dal modello inflazionario sottostante. Questa è la ragione principale del grande interesse della comunità cosmologica per la misura della non Gaussianità. Naturalmente, nella ricerca della non-Gaussianità primordiale è necessario ricorrere a statistiche di ordine superiore rispetto allo *spettro di potenza*. Ci si aspetta che la maggior parte del segnale non Gaussiano prodotto durante l'Inflazione si presenti sotto forma di correlazioni a tre punti, che possono essere misurate nello spazio armonico dal *bispettro*. Purtroppo, a causa dell'elevato tempo computazionale richiesto, non è possibile calcolare direttamente il bispettro dai dati. La ricerca di segnali non gaussiani consiste quindi nel misurare la correlazione tra il bispettro dei dati e determinati modelli teorici che riproducono il segnale predetto da specifici modelli inflazionari. Molte teorie inflazionarie producono correlazioni ad alto ordine il cui bispettro presenta un ampiezza dipendente dalla scala. Questo è il motivo per cui una parte significativa di questa tesi

sarà dedicata allo sviluppo di tecniche statistiche per la stima di bispettri con un esplicita dipendenza dalla scala in osservazioni della CMB. I risultati presentati in questa tesi sono ottenuti dalle osservazioni dei satelliti WMAP e *Planck*.

La seconda parte di questo lavoro riguarda invece il problema dell'identificazione delle diverse fonti che contribuiscono alla luminosità del cielo nelle frequenze delle microonde. L'emissione di *foreground* potenzialmente produce grandi deviazioni dalla Gaussianità, che in linea di principio possono essere utilizzate per identificare e rimuovere i componenti spuri dalle mappe del cielo a microonde. Il nostro obiettivo è lo sviluppo di una tecnica di pulizia dai *foreground* basata sull'ipotesi che, se i dati vengono rappresentati nella base appropriata, il segnale delle emissioni di *foreground* appare sparso. La sparsità implica che la maggior parte del segnale sia concentrata in pochi elementi della base, che possono essere usati per ricostruire il componente corrispondente ricorrendo a una tecnica detta *thresholding*. Abbiamo verificato che il frame delle *needlet sferiche* ha le proprietà ideali per separare il segnale coerente del *foreground* dal segnale isotropo e stocastico della CMB. I principali vantaggi della nostra tecnica di *needlet thresholding* sono, in primo luogo, che non richiede di avere osservazioni a diverse frequenze e inoltre che può essere utilizzata in combinazione con altri metodi. Pertanto può essere uno strumento prezioso in esperimenti che osservano il cielo in un limitato intervallo di frequenza come, per esempio, gli attuali esperimenti che mirano a misurare la CMB da terra.

La tesi è strutturata come segue:
Nel primo capitolo viene fornita una breve panoramica dei concetti teorici fondamentali della cosmologia e della statistica.
Nel secondo capitolo sono illustrate le tecniche utilizzate nell'analisi dei dati CMB.
Il terzo capitolo introduce modelli di bispettro dipendenti dalla scala e presenta la misura della non Gaussianità primordiale dai dati WMAP.
Il quarto capitolo espande l'analisi dei modelli di bispettro con dipendenza di scala sui dati *Planck* e presenta diversi miglioramenti nello stimatore.
Nel quinto capitolo viene delineato il problema della separazione dei componenti *foreground* in osservazioni della CMB.
Il sesto capitolo contiene una descrizione del telaio dell'ugello sferico e la base della regressione dell'ago.
Nel settimo capitolo introduciamo una nuova tecnica di rimozione del *foreground* basata sulla sparsità del segnale nello spazio *needlet*.
Infine, l'ottavo capitolo è dedicato alle conclusioni e alle prospettive future.

Abstract

The focus of this work is the development of statistical and numerical methods for the study of non-Gaussian and/or anisotropic features in cosmological surveys of the microwave sky. We focus on two very different types of non-Gaussian (NG) signals. The former is *primordial non-Gaussianity* (PNG), generated in the very Early Universe during the inflationary expansion stage. In this case the aim of our study will be that of exploiting the NG component in order to extract useful cosmological information. The latter is *non-Gaussianity* generated by *astrophysical foreground contamination*. In this case, the goal is instead that of using non-Gaussianity as a tool to help in removing these spurious, non-cosmological components (of course foregrounds themselves contain relevant astrophysical information, but the focus in this thesis is on Cosmology, therefore foregrounds are regarded here only as a contaminant).

Considerable efforts have been put so far in the search for deviations from Gaussianity in the CMB anisotropies, that are expected to provide invaluable information about the Inflationary epoch. Inflation is in fact expected to produce an isotropic and nearly-Gaussian fluctuation field. However, a large amount of models also predicts very small, but highly *model dependent* NG signatures. This is the main reason behind the large interest in primordial NG studies. Of course, the pursuit for primordial non-Gaussianity must rely on beyond power spectrum statistics. It turns out that the most important higher order correlator produced by interactions during Inflation is the three point function, or, more precisely, its Fourier space counterpart, called the bispectrum. To overcome the issue of computing the full bispectrum of the observed field, that would require a prohibitive amount of computational time, the search for PNG features is carried out by fitting theoretically motivated bispectrum templates to the data. Among those, one can find bispectrum templates with a scale-dependent (SD) bispectrum amplitude. Such templates have actually received little attention so far in the literature, especially as long as NG statistical estimation and data analysis are concerned. This is why a significant part of this thesis will be devoted to the development and application of efficient statistical pipelines for CMB scale-dependent bispectra estimation. We present here the results of the estimation of several primordial running bispectra obtained from WMAP 9 year and *Planck* data-set.

The second part of this thesis deals instead, as mentioned iin the beginning, with the *component separation* problem, *i.e.* the identification of the different sources that contributes to the microwave sky brightness. Foreground emission produces several, potentially large, non-Gaussian signatures that can in principle be used to identify and remove the spurious components from the microwave sky maps. Our focus will be on the development of a foreground cleaning technique relying on the hypothesis that, if the data are represented in a proper basis, the foreground signal is *sparse*. Sparseness implies that the majority of the signal is concentrated in few basis elements, that can be used to fit the corresponding component with a *thresholding* algorithm. We verify that the *spherical needlet* frame has the right properties to disentangle the coherent foreground emission from the isotropic stochastic CMB signal. We will make clear in the following how sparseness in needlet space is actually in several ways linked to the coherence, anisotropy and non-Gaussianity of the foreground components.. The main advantages of our needlet thresholding technique are that it does not requires multi-frequency information as well as that it can be used in combination with other methods. Therefore it can represent a valuable tool in experiments with limited frequency coverage, as current ground-based CMB surveys.

The layout of this thesis is structured as follows.

In the first chapter we provide a brief review of the basic concepts about theoretical cosmology and estimation theory.

In the second chapter we give an insight of the methodologies used in CMB data-analysis.

The third chapter introduces scale-dependent bispectrum templates and provides estimates of the PNG running from WMAP 9 year data-set.

The fourth chapter expands the analysis of SD templates to the *Planck* data-set, and it provides various improvement of the SD estimator.

In the fifth chapter we outline the foreground component separation problem in CMB surveys.

The sixth chapter contains a description of the spherical needlet frame and the essential notions about needlet regression.

In the seventh chapter we introduce a new component separation pipeline based on the sparseness of the foreground signal in needlet space.

Finally, the eighth chapter is dedicated to the conclusions and to an overview of future prospects.

Contents

List of Figures

List of Tables

Chapter 1

Theoretical Background

In this chapter we will outline an overview of the basic notions about cosmology and CMB data analysis. We will provide a brief review of the standard cosmological model and its most important extension, the theory of inflation. We will then introduce the essential concepts about the description and the analysis of random fields and some basic principles on statistics and estimation theory.

1.1 The Standard Cosmological Model

The first developments of what is now known as "the standard Cosmological model" trace back to the beginning of the last century, when the recently formulated General theory of Relativity allowed the development of a complete and testable theory of the structure and the evolution of the Universe. The basis of this model are the so called *Cosmological Principle* and the expansion of the Universe. The former states that the Universe is homogeneous and isotropic on large scales, the latter derived form observations. The crucial piece of evidence in favour of Universe expansion comes from the Edwin Hubble's discovery that the galaxies recede from us at a velocity proportional to the distance. The resulting picture is a Universe homogeneous and isotropic, that expands starting from a initial state of high energy density, the so called Big Bang.

In 1929 Edwin Hubble found the relation between the distance of galaxies and redshift z called *Hubble Law*:

$$z = \frac{H_0 d}{c}, \tag{1.1}$$

where H_0 is the so called *Hubble constant*. The redshift is the shift of the photons wavelength to the red end of the spectrum due to the recession of galaxies from us

$$z = \frac{\lambda_{obs} - \lambda_{emit}}{\lambda_{emit}} \simeq \frac{v_{rec}}{c}, \tag{1.2}$$

where the relation between redshift and velocity applies only for low redshift. The Hubble law states so that the galaxies are receding from us with velocity that increase linearly with distance with a slope given by H_0.

General relativity allows us to extrapolate the expansion backwards in time. This yields to the conclusion that the Universe evolved from a state in which energy density diverges to infinite. The term "Big Bang" refers to this initial singularity and to the following hot phase in which the Universe expanded while dramatically cooling down.

It was in this moment that part of the chemical composition of the Universe was established. The atomic nuclei are predicted to form when the temperature fell below $\sim 0.1MeV$ during the so called "Big Bang Nucleosynthesis" (BBN from now). From the condition of nuclear statistical equilibrium it is possible to derive the expected abundances of the nuclei formed. The theoretical predictions depend on the combined density of protons and neutrons, which, in Cosmology, is called "baryon density". Thus BBN allows to measure the primordial baryon density, as well as the relative abundances of several light elements. These predictions are consistent with observational measurements of light elements, obtained for example with the analysis of distant Quasar spectra. This consistency provides one of the main confirmation of the Big Bang models.

But the evidences of an early hot phase are not exhausted with elements abundance. The photons that, in equilibrium with matter, formed this primordial plasma are still observable today as a diffuse background of black-body radiation, the *Cosmic Microwave Background* (from now CMB). The predictions of elements abundance and of CMB represent the great successes that allowed the Hot Big Bang model to establish itself as the Standard Cosmological Model.

It was however soon realised that a simple Universe filled with standard matter could not explain a vast range of observations, like rotational curves of galaxies and gravitational lensing from distant clusters. The observations reveal many inconsistencies between the distribution of luminous matter and the gravitational potential. In order to explain this we need to introduce a new "Dark Matter" component (DM) that doesn't interact directly with matter and radiation but contributes to the gravitational

potential. In other words, DM feels gravitational interactions, but does not interact electromagnetically.

Moreover recent measurements of the expansion rate using the supernovae Ia as standard candles led to the identification of another new component: Dark Energy. This is the dominating component and it is supposed to be the source of the late time acceleration in the expansion rate.

1.1.1 Friedmann-Lemaître-Robertson-Walker Metric

The Einstein equations describe gravitation as a result of the relation between the energy density and the geometry of space-time. The geometric properties are described by the *metric* tensor that defines how coordinates are related to physical distance.

First of all we must emphasise that the Cosmological Principle is valid only if we consider all the Universe at the same instant. The geometrical interpretation then implies that space, but not space-time, must be maximally symmetric. Since a symmetry corresponds to a conservation rule, a 3 dimensional space must be invariant under both translation and rotations to agree with the cosmological principle.

The Friedmann-Lemaître-Robertson-Walker metric (FLRW from now) describes a 4-dimensional space-time admitting an infinity of 3-dimensional subspaces maximally symmetric. Therefore, using the signature $(-,+,+,+)$, the FLRW metric tensor is:

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

Here we use *comoving coordinates*: the metric is free of time-space cross-terms so the space components are proportional to a single function of t. The result is that a physical point keeps the same spatial coordinates at any time. The expansion is quantified by the scale factor $a(t)$. It allows to pass from *comoving distance*, that remains constant in the expansion, to physical distance that grows in time. A maximally symmetric space will be certainly spherically symmetric, so the line element is usually defined in spherical coordinates giving:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right]. \tag{1.4}$$

This leads to the definition of the parameter k. It describes the *curvature* of space-time: the case $k = 0$ corresponds to a flat space with no curvature ; $k = 1$ corresponds to a positive curvature, or to an closed space; $k = -1$ corresponds to a negative curvature and to a open space.

With the metric we can compute the Ricci tensor $R_{\mu\nu}$ and the Ricci scalar R and then write down the Einstein equation:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.5}$$

where $T_{\mu\nu}$ is the *energy momentum tensor*, in our case we consider a perfect fluid:

$$T_{\mu\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix}, \tag{1.6}$$

here ρ is the energy density and P is the pressure for unit of surface of the fluid. The energy density can be expressed in function of the density of each component $\rho = \rho_m + \rho_\Lambda + \rho_\gamma + \rho_\nu$ where the subscript m indicates the matter density, Λ the Dark Energy density, γ the photons density and ν the neutrinos. The matter density term contains both baryon and Dark matter contribution $\rho_m = \rho_b + \rho_{DM}$. From the solution of (1.5) (for further details see appendix A or (Dodelson 2003) for a complete discussion) comes the *Friedmann equations*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{1.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P). \tag{1.8}$$

Equation (1.7) defines how rapidly the scale factor changes, this is called *Hubble rate*:

$$H(t) = \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} \tag{1.9}$$

the most important fact is that it is relate to the energy density, the measurement of the Hubble rate provide precious information about the history of the Universe. The Hubble rate's value at present time is the Hubble constant that appears in equation (1.1).

The composition of the Universe is defined in relation to the critical density ρ_c by the density parameter Ω.

$$\rho_c = \frac{3H^2}{8\pi G}, \tag{1.10}$$

$$\Omega = \rho \frac{8\pi G}{3H^2}. \tag{1.11}$$

Rewriting (1.7) it's trivial to understand the meaning of critical density:

$$\Omega - 1 = \frac{k}{H^2 a^2}. \tag{1.12}$$

ρ_c is then the density for which the Universe is spatially flat. The relation between the curvature and the energy content of the Universe can be summarised as:

$$\begin{array}{lllllll} \text{Open} & \leftrightarrow & k = -1 & \leftrightarrow & \Omega < 1 & \leftrightarrow & \rho < \rho_c \\ \text{Flat} & \leftrightarrow & k = 0 & \leftrightarrow & \Omega = 1 & \leftrightarrow & \rho = \rho_c \\ \text{Close} & \leftrightarrow & k = 1 & \leftrightarrow & \Omega > 1 & \leftrightarrow & \rho > \rho_c \end{array} \tag{1.13}$$

Current experiments suggest a total density very close to ρ_c, that is $\Omega = \rho/\rho_c \simeq 1$.

1.1.2 The discovery of The Cosmic Microwave Background

The detection of the CMB by Penzias and Wilson in 1965 set the milestone for the affirmation of the Big Bang as standard cosmological model. The CMB was predicted for the first time by Gamow in the 40s as the natural consequence of an expanding Universe; whereas it is totally incompatible with a steady state Universe.

Actually the controversy was not immediately decided. The CMB is the relic of the hot ancient Universe, coming from an epoch in which radiation and matter were tightly coupled by the Compton and Coulomb scattering with free electrons, forming the so called *photon-baryon fluid*. Being emitted in a condition of high equilibrium the CMB photons were expected to possess a black-body spectrum. However it was very difficult for the time to avoid atmospheric absorption and measure the spectrum with the required narrow accuracy. Finally in the 90s the COBE spacecraft, with its spectrophotometer FIRAS, measures the most perfect black-body spectrum in nature, with a temperature of $2.753 \pm 0.002K$. Furthermore, measurement of its color temperature shows that it is extraordinarily uniform over the whole sky providing strong evidence for large scale isotropy.

The CMB photons scatter the last time at the epoch called *recombination*. Recombination signs the moment in which neutral hydrogen began to form lowering the number of free electrons. Before recombination due to the high temperature electrons could not bind to the nuclei to form neutral atoms, in this early times the free electrons gave an high contribution to the optical depth via Compton scattering. When the temperature was lowered enough to allow the formation of the first atoms the number of free electrons fell down; without their contribution optical depth fell down too and radiation and matter decoupled. The matter began to cluster, forming the structure we observe today, while the photons were free to travel through space with nearly no interactions, giving rise to a diffuse electromagnetic signal of Cosmological origin, hence the name, that we already been using, of "Cosmic Microwave Background". Looking at the CMB we see the image of a spherical section of this primordial photon-baryon fluid just after the decoupling, this is the so called *Last Scattering Surface*. It is like to see a snapshot of a section of the primordial plasma taken at the recombination. From the statistical mechanics we know that photons decoupling occurred when the temperature of the Universe was $\sim 3000\,K$. The energy density of uniform blackbody radiation is defined by:

$$\rho_\gamma = 2\int \frac{\mathrm{d}^3 p}{(2\pi)^3}\frac{1}{e^{p/T}-1} = \frac{\pi^2}{15}T^4, \tag{1.14}$$

were p is the momentum of the photons and T is the temperature. The evolution of energy can be computed from the conservation law for the energy momentum tensor (1.6):

$$\nabla_\mu T^\mu_{\ \nu} = \partial_\mu T^\mu_{\ \nu} + \Gamma^\mu_{\alpha\mu} T^\alpha_{\ \nu} - \Gamma^\alpha_{\nu\mu} T^\mu_{\ \nu} = 0, \tag{1.15}$$

considering the $\nu = 0$ component, and computing the connection (A.2) for the FLRW metric, we have the conservation law in an expanding universe:

$$\frac{\partial \rho}{\partial t} + H[3\rho + 3P] = 0. \tag{1.16}$$

The radiation pressure is $P_\gamma = \rho_\gamma/3$, so that (1.16) gives $\rho_\gamma \propto a^{-4}$, this implies that the temperature scales as a^{-1}. We can so compute the growth of the scale factor from the recombination to our day. Since the CMB temperature decreased by about a factor ~ 1100, $a(t)$ should be increased equally, consistently with the expansion of the Universe. The redshift of the CMB photons can easily be computed by applying the Wien law, it

is so:

$$z = \frac{\lambda_{max}^{obs} - \lambda_{max}^{*}}{\lambda_{max}^{*}} = T_* \left(\frac{1}{T} - \frac{1}{T_*} \right) \simeq \frac{T_*}{T} \simeq 1100, \tag{1.17}$$

were $*$ indicates the values at recombination.

As we said the CMB is extremely uniform, but at a sensitivity level of about one part in 1000 of the total intensity a dipolar anisotropy can be observed. It is naturally attributed to the kinematic Doppler effects associated with the Earth's motion through a reference frame in which the radiation is at rest. This allows to compute the peculiar motion of our galaxy with respect the cosmic reference frame. The result is $v \simeq 600\, Km/s$ in direction of the constellation of Hydra Centaurus, in the same direction of the other galaxies of our group.

Observations with a sensitivity of 10^{-5} part of the total intensity opens new scenarios in the study of the CMB: since they show the presence very small anisotropies. These anisotropies are the result of gravitational fluctuations on the last scattering surface and their origin cannot be explained by the Standard Cosmological Model. The large scale structure we observe today nearly certainly originated from these little perturbation. The most accredited theory to explain the origin of these anisotropies is Inflation we will describe in the next section.

1.1.3 Inflation

As it often happens in science, the great discovery of the CMB opened a number of new questions about the formation of our Universe. As we saw, the Standard Cosmological Model does not explain the origins if the little temperature anisotropies, but there is an even most evident problem. One of the main issues is the impressive level of uniformity of the CMB sky; this is a wonderful proof of the Cosmological principle but, on the other hand, is a fact rather difficult to justify.

To better understand this issue we have to introduce the concept of *comoving horizon* η. The comoving horizon is the maximum comoving distance traveled by light since the beginning of the Universe. Obviously, for two regions to communicate, they should be at a comoving distance smaller than the comoving horizon. If they aren't, they cannot have exchanged any information.

If recombination happened $\sim 380,000\,\mathrm{yr}$ after the Big Bang, as the model predicts, and the Universe was radiation dominated, the size of comoving horizon at that time subtends a very small angle from our point of view. Furthermore, from the analysis of

CMB anisotropies it is possible to obtain information about the size of causally connected regions at the recombination, resulting in an estimate of $\sim 2°$.

Despite this, the CMB appears to be thermalized at all angular scales, so that physical processes must have acted on all scales in order to allow the Universe to reach an equilibrium state.

If this couldn't happen, how can all the sky appear to have been in thermal equilibrium, when it couldn't even be considered a single thermodynamic system? This apparent impasse is the so called *Horizon Problem*, a possible explanation to this problem comes from *Inflationary theory*. As we will see, this theory can also explain even the origin of the inhomogeneities.

Inflation states that, at very early times, the Universe underwent a period of accelerated expansion. To understand how this can solve the horizon problem, it is more illuminating to reason in terms of *Hubble radius* instead of comoving horizon. Acceleration implies $\mathrm{d}^2a/\mathrm{d}t^2 > 0$, that is:

$$\frac{\mathrm{d}^2 a}{\mathrm{d}t^2} = \frac{\mathrm{d}}{\mathrm{d}t}\left[a\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right] = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{aH}\right)^{-1} > 0. \tag{1.18}$$

The quantity $1/aH$ is the Hubble radius: it is the physical distance a particle can travel in one expansion time. It is related to the comoving horizon by the equation:

$$\eta = \int_0^a \frac{\mathrm{d}a'}{a'}\frac{1}{Ha'}. \tag{1.19}$$

The difference between the two is that regions farther than η never could have communicated in the past (*i.e*, from t=0 up to today) whereas if they are at distance grater than $1/aH$ they aren't connected only at given time. From (1.18) we see that the consequence of an accelerated expansion is the decrease of the Hubble radius, as it's shown in the right panel of figure 1.1. Different region of the CMB, thus, are effectively not connected at the recombination, but only in terms of Hubble radius. The comoving horizon instead grew together with the scale factor, so that it was bigger than the last scattering surface yet at the recombination. If Inflation succeed, zones outside the Hubble radius at recombination can have been causally connected at the early times in which they thermalized, as it can be seen in the left panel in figure 1.1. Quantitatively, to solve the horizon problem, during Inflation the Universe must have expanded over approximately 28 orders of magnitude (Guth 1981).

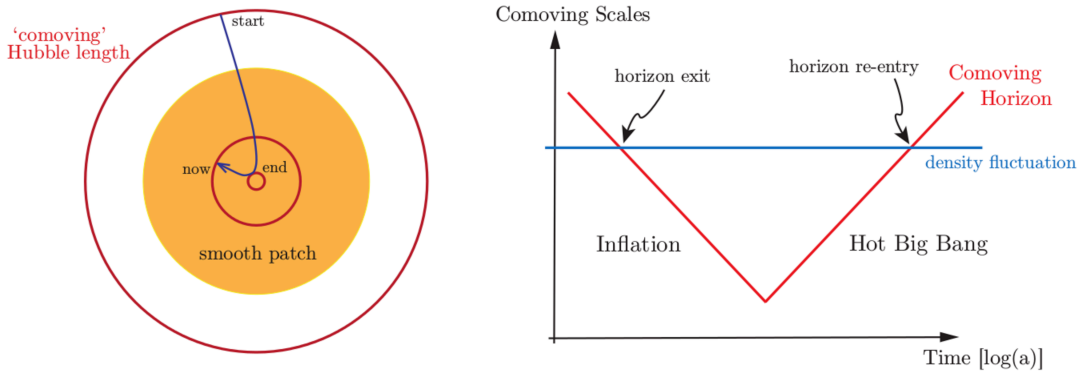

FIGURE 1.1: Evolution of Hubble radius in an inflationary Universe. The left figure shows the Hubble sphere before and after Inflation, and how it is now, still inside the smooth patch. The right panel shows how Inflation solves the horizon problem. All scales that are relevant to cosmological observations today were larger than the Hubble radius until $a \sim 10^{-5}$. However, at sufficiently early times, these scales were smaller than the Hubble radius and therefore causally connected. Image taken from (Baumann 2011).

Once it's proven that Inflation can provide solutions to these inconsistencies, the following step is to define what type of energy is needed to produce it. The second Friedmann equation (1.8), under the condition (1.18) gives $(\rho + 3P) < 0$, so Inflation requires:

$$P < -\frac{1}{3}\rho. \tag{1.20}$$

The energy density ρ is positive, so the the pressure must be negative. We don't know any type of ordinary matter that has a negative pressure. We have to suppose that what drives Inflation is some kind of field arising at the high energy scale of the primordial Universe. Objects with the required properties can indeed be found in the high energy extension of the Standard Model of Particle Physics.

1.1.3.1 Primordial Potential

At early time the energy density of the Universe reached incredibly high levels, about $10^{15} GeV$ when $\sim 10^{-37}\, s$ passed from Big Bang. This energy is 9 order of magnitude above the limits of modern particle accelerators. Physics in this range of energy is, from an experimental point of view, still unexplored. If Inflation occurred, the more reasonable explanation seems that it was driven by one or more quantum scalar fields, called *inflaton*, arising at these scales.

In order to understand how a scalar field component can produce acceleration in the expansion rate, we have to take a look at the Einstein equations. Let us start by

considering the energy momentum tensor of a homogeneous scalar field $\psi(t)$, for now we neglect the first order perturbation $\delta\psi(x,t)$, so we have

$$T^{\alpha}_{\ \beta} = g^{\alpha\nu}\frac{\partial\psi}{\partial x^{\nu}}\frac{\partial\psi}{\partial x^{\beta}} - g^{\alpha}_{\ \beta}\left[\frac{1}{2}g^{\mu\nu}\frac{\partial\psi}{\partial x^{\mu}}\frac{\partial\psi}{\partial x^{\nu}} + V(\psi)\right], \tag{1.21}$$

were $V(\psi)$ is the potential of the field and for the metric we used the signature $(-,+,+,+)$. Since the field is homogeneous we consider only the time derivative, the time-time component for a perfect isotropic fluid is $T^{0}_{\ 0} = -\rho$:

$$\rho = \frac{1}{2}\left(\frac{\mathrm{d}\psi}{\mathrm{d}t}\right)^{2} + V(\psi), \tag{1.22}$$

and the space-space component, which represent the pressure, gives:

$$P = \frac{1}{2}\left(\frac{\mathrm{d}\psi}{\mathrm{d}t}\right)^{2} - V(\psi). \tag{1.23}$$

The first of these equations shows that the dynamic of a scalar field is totally analogue to that of a single particle moving in a potential. The second tells us that a negative pressure requires more potential energy than kinetic. Little kynetic energy implies that the energy density and the potential remain nearly constant in time. From equation (1.7) it's clear that this condition produces the exponential expansion required to solve the horizon problem.

These assumptions form the basis for the simplest model of Inflation: the so called "single-field slow-roll" scenario (Linde 1982) (Albrecht & Steinhardt 1982). In the slow-roll picture, the scalar field is initially characterized by a very flat potential. This makes $V >> E_k$ in (1.22) and (1.23), thus producing the conditions for acceleration. Inflation ends when the field reaches the end of the plateau, and falls down into its ground state (*i.e* we have a transition from a "false vacuum" to a "true vacuum"), as shown in figure 1.2. At this point the field oscillates around the minimum, and decays into ordinary matter and radiation. Initial energy in the false vacuum state thus get converted into the various components of the cosmological fluid. This phase is called *reheating* and it sets initial conditions in the Standard cosmological Model.

The name *slow-roll* derives from the analogy with the dynamics of a classic point particle rolling down a potential.

The equation governing the evolution of the field in an expanding Universe is:

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}t^2} + 3H\frac{\mathrm{d}\psi}{\mathrm{d}t} + \frac{\mathrm{d}V(\psi)}{\mathrm{d}\psi} = 0. \tag{1.24}$$

During Inflation the Hubble rate remains nearly constant. Meanwhile the scale factor grows nearly exponentially, so its relation with conformal time can be simplified as:

$$\eta = \int_{a_e}^{a} \frac{\mathrm{d}a}{a^2 H} \simeq -\frac{1}{aH}. \tag{1.25}$$

The slow-roll model is usually characterized by the parameters:

$$\epsilon = -\frac{\dot{H}}{aH^2}, \tag{1.26}$$

$$\eta = -\frac{1}{aH\dot{\psi}}\left[3aH\dot{\psi} + a^2\frac{\partial V}{\partial \psi}\right], \tag{1.27}$$

that quantify how slowly the field is rolling. By definition, an Inflationary epoch has $\epsilon < 1$.

So far we have presented the zeroth order picture, this provides a good solution for the horizon problem but it doesn't explain the onset of the fluctuations. It has been shown that a perturbative expansion of the homogeneous scalar field may account for the initial power spectrum of perturbations (Bardeen et al. 1983; Brandenberger et al. 1983; Guth & Pi 1982; Starobinsky 1982). We can express the field as:

$$\psi(\vec{x}, t) = \psi(t) + \delta\psi(x, t), \tag{1.28}$$

where the first term is the homogeneous part and the second represents the perturbation. The zeroth order part is responsible for the accelerated expansion, while the first order part induces perturbations on the metric. The existence of this perturbation component is unavoidable in quantum context, since uncertainty principles imply small quantum fluctuations of the field, giving rise to this $\delta\psi$ term. From General Relativity we know that the density of matter and radiation is coupled to the scalar field perturbation (1.28). Therefore it is the scalar part that eventually provided the initial conditions for density fluctuations after reheating. Given the limited scope of this brief introduction, we omit here the full treatment needed to derive the power spectrum for the perturbative term (see*e.g.* (Dodelson 2003) for a full calculation), and we simply outline the main aspects of the derivation.

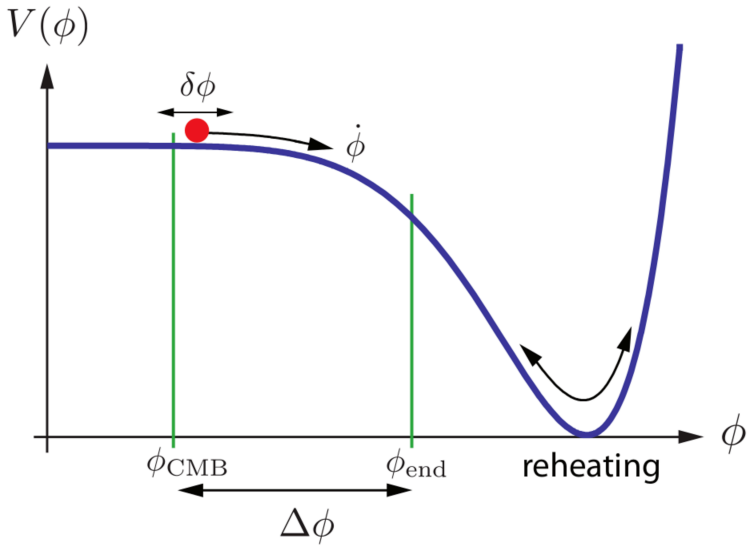

FIGURE 1.2: Inflation potential (the inflaton field is ϕ instead of ψ). ϕ_{end} indicates the point in which the kinetic energy equals the potential and the Inflation ends. CMB fluctuations are created at ϕ_{CMB} about 60 e-folds before the end of Inflation. At reheating, the energy density of the inflaton is converted into radiation. Image taken from (Baumann & Peiris 2009).

The most common way is to derive the equation for $\delta\psi$ in a gauge[1] in which the scalar field is spatially decoupled from the density fluctuations, called *spatially flat slicing*. In this gauge the spatial part of the metric remains unperturbed. So the first order part of the inflaton scalar field can be directly derived from the conservation of the energy momentum tensor, leaving:

$$\ddot{\delta}\psi + 2aH\dot{\delta}\psi + k^2\delta\psi = 0, \tag{1.29}$$

were dots indicate derivatives with respect to conformal time. Given this equation, we have to find a gauge invariant variable[2] related to $\delta\psi$ in the flat slicing gauge and to the metric perturbation Φ in the conformal Newtonian gauge. Again, we will not formally define the Newtonian gauge here. It suffices to say for our purposes that, when perturbing the metric with this gauge choice, the perturbation variable Φ takes a clear Physical meaning: it is the Newtonian potential generated by the small fluctuation in the homogeneous background curvature.

Once we found a definition for the inflaton perturbation, and understand how it can be related to the density perturbation, we need to define their power spectrum, so that

[1] A gauge theory is a type of field theory in which the Lagrangian is invariant under a continuous group of local transformations. The terms *gauge* refers to the coordinate systems connected by these transformations. In General Relativity a gauge is defined by a specific choice ot the functional relation linking coordinates in the unperturbed background to coordinates in the perturbed space.

it can be compared with its observed counterpart. The choice of the power spectrum as observable is strictly related to the quantum nature of the field. If we rearrange (1.29) into the form of a harmonic oscillator and quantize it (for further details about the quantum harmonic oscillator see (Peskin & Schroeder 1995)) it's easy to define the variance of the operator that would represent the field. This is important because, in the simplest picture of the single scalar field, in consequence of their quantum mechanical origins, the fluctuations are Gaussian.

A Gaussian distribution is fully characterized by mean and variance. In our case the mean is 0 by definition (1.28), therefore its variance is sufficient to totally define the field. The variance is related to the Power Spectrum by:

$$\langle|\delta\psi(\vec{x})\delta\psi(\vec{x}')|\rangle = (2\pi)^3 P_\psi(k)\delta^3(\vec{k}-\vec{k}'), \tag{1.30}$$

where $P_\psi(k)$ is the power spectrum. As we were saying above, being linked to the variance, the power spectrum fully defines the perturbations fields. It is thus a quantity of paramount importance in Cosmology.

Omitting a bit of mathematics, we leap to the definition of the spectrum in the spatially flat slicing gauge:

$$P_{\delta\psi} = \langle|\delta\psi|^2\rangle = \frac{H^2}{2k^3}. \tag{1.31}$$

The next step is to find the gauge invariant variable, in the flat slicing gauge it is:

$$\zeta = -\frac{aH}{\dot{\psi}}\delta\psi. \tag{1.32}$$

The power of this variable is thus related with (1.31):

$$P_\zeta = \langle|\zeta|^2\rangle = \left(\frac{aH}{\dot{\psi}}\right)^2 \langle|\delta\psi|^2\rangle = \left(\frac{aH}{\dot{\psi}}\right)^2 P_{\delta\psi}. \tag{1.33}$$

In the Newtonian gauge at the end of Inflation $\zeta = 3\Phi/2$, here Φ represent the gravitational potential fluctuation, finally we can define the power spectrum of the primordial Newtonian potential generated by density perturbations:

$$P_\Phi = \frac{4p_\zeta}{9} = \left.\frac{8\pi G H^2}{9\epsilon k^3}\right|_{aH=k}. \tag{1.34}$$

[2]Gauge invariant are quantities remaining unchanged under gauge transformation. They are useful to connect one gauge to another.

Observing equation (1.34) we see that Inflation produces a so called *scale-invariant* power spectrum, that is $P(k)k^3$ is constant. As we will see in detail in the next chapter, the CMB power spectrum is indeed nearly scale-invariant. A scale-invariant (or scale-free) spectrum is characterized by a constant logarithmic slope, physically this means that the process from which it derives acted in the same way at all scales. The little deviation from scale invariance, detected with a high degree of significance in recent surveys (*Planck* satellite), is also matching a specific prediction of Inflation, once corrections to the power spectrum in a slow-roll parameter expansion are accounted for. CMB observation thus provide an outstanding confirmation of the Inflationary paradigm.

1.1.4 Non-Gaussianity

Until now most predictions of Inflation passed the observational tests. The Universe appears to be remarkably flat, homogeneous and isotropic and the measured power spectrum is very near to the Harrison-Zel'dovich one.

However, the standard single-field slow-roll scenario briefly outlined above is not the only possible Inflationary model we can build. On the contrary, literally hundreds of models have been proposed in the literature, eventually involving different scalar fields potentials and kinetic terms, or possibly including multiple fields as well as different mechanisms for the generation of perturbations. Moreover, some alternatives to Inflation are still compatible with current data. If we want to discriminate between all these possibilities, we need to consider observables beyond the power spectrum of scalar perturbation. A good way to discriminate between different scenarios is the measurement of primordial non-Gaussianity. The standard single-field slow-roll paradigm implies very small deviations from Gaussianity since the field is essentially a quantum harmonic oscillator in its ground state. A certain degree of non-Gaussianity does arise from small non-linear coupling with gravity, but it has been shown to be tiny and undetectable. However a number of more complex models exist predicting some degree of non-Gaussianity in the perturbations.

If a field is not Gaussian then its power spectrum is no longer enough to fully define it, thus the natural step is to look at higher order correlator. The lowest order indicator of Non-Gaussianity is the three-point function ("skewness"), because it is identically 0 for a Gaussian field. Actual calculations in the Inflationary framework (beyond the scope of

this Introduction) show that the three point function is generally and by far the largest correlator produced in non-standard models.

The study of the primordial three-point function in CMB data is thus the focus of this thesis. The most interesting aspect is that different models predict different 3-point functions, so that measurements of this type allow to strongly discriminate between different scenarios, that would otherwise be totally degenerate if looking just at the power spectrum.

We must emphasize that a non-Gaussian signal, even if present, is predicted to be totally subdominant with respect to the Gaussian component of primordial perturbations, and require very accurate, high resolution data-sets to be studied. The recent CMB data provided by the WMAP, and especially *Planck* satellite, are ideal suited to this purpose.

1.2 Perturbations Power Spectrum and Random Fields

In this section we briefly review some key elements of the formalism used to describe cosmological perturbations, that will be widely used in the following. With this aim we introduce the concepts of *random field* and of perturbations *Power Spectrum*.

We have seen that the expansion history of the Universe, as well as is curvature, depends on the evolution of the energy density of different components: matter radiation and cosmological constant. These density fields are homogeneous in the background FLRW model, so its natural to describe them as:

$$\rho(x,t) = \rho(t) + \delta(x,t), \tag{1.35}$$

where $\rho(t)$ represents the background density and $\delta(x,t)$ is the density perturbation field, representing small deviations from homogeneity. We saw how in the standard Cosmological scenario the density of different components presents small initial fluctuations , possibly set during the inflationary epoch, that later on grow via gravitational instability, forming all the structures we see today in the Universe.

Since $\delta(x,t)$ is a random field, we cannot make deterministic predictions about one specific realization. What we can do is to make a statistical description of the Universe; taking a CMB related example, we can predict *e.g* how many hot and cold spot there are *on average*, and the expected fluctuation of this number, but we cannot tell *exactly*

how many of them we are going to see, and where they are going to be located in our observed universe.

In other words, and more precisely, when we say that the perturbations are random fields we mean that for every position x_i at a given time t (wich we will omit to explicitly write from now)the quantity $\delta(x_i)$ is a random variable.

The value of a random variable cannot be predicted but we can associate to the variable the *probability density function* (PDF) $P_x(\hat{\delta})$. It quantifies the probability that the variable $\delta(x)$ takes a certain value $\hat{\delta}$ in a certain point x. Considering N points we define the joint probability $P_{(x_1 x_2 \dots x_N)}(\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_N)$ as the probability that in every point $\{x_1, x_2, \dots, x_N\}$ the random variable takes respectively the values $\{\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_N\}$.

The physical quantities described by random fields, as the Universe density distribution, are so defined by the statistical properties of the field. The statistical distribution of a random variable δ can be described through the *moments* of the PDF, which are the expectation value of the quantity:

$$\mu_k = \int \mathrm{d}\delta(x)\, [\delta(x) - \langle\delta(x)\rangle]^k P_x(\delta(x)). \tag{1.36}$$

Following a most common notation we define the first 4 moments as

$$\mu_1 = \mu = \langle\delta(x)\rangle \qquad \text{Average,} \tag{1.37}$$

$$\mu_2 = \sigma^2 = \langle\delta^2(x)\rangle - \langle\delta(x)\rangle^2 \qquad \text{Variance,} \tag{1.38}$$

$$\mu_3 = \int \mathrm{d}\delta(x)\, [\delta(x) - \langle\delta(x)\rangle]^3 P_x(\delta(x)) \qquad \text{Skewness,} \tag{1.39}$$

$$\mu_4 = \int \mathrm{d}\delta(x)\, [\delta(x) - \langle\delta(x)\rangle]^4 P_x(\delta(x)) \qquad \text{Kurtosis.} \tag{1.40}$$

Furthermore if the random variables are more than one, as it happens when we consider random fields, which are collections of random variables, we can define the *correlation function* $\xi(r)$ and the *covariance*, for example taking the density perturbation in x and in $x + r$:

$$\xi(r) = \langle\delta(x)\delta(x + r)\rangle \tag{1.41}$$

$$\mathrm{cov}(\delta(x), \delta(x + r)) = \langle\delta(x)\delta(x + r)\rangle - \langle\delta(x)\rangle\langle\delta(x + r)\rangle. \tag{1.42}$$

Given N variables we can define the *covariance matrix*: it is a $N \times N$ matrix that quantifies the covariance for every combination of variable.

The most important characteristic of cosmological density perturbations, inherited, as we will see, from quantum fluctuations, is their *Gaussianity*. Gaussian distribution is:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \tag{1.43}$$

from this definition we can note that the Gaussian distribution is completely defined by the first two moments: the average μ and the variance σ^2. The higher ones are null if odd ($\mu_3, \mu_5 \dots$) while the evens ones can be defined as powers of σ.

The energy density of the different cosmological components Universe is thus described, for each components, by a "random Gaussian scalar field". This allows us to totally describe it with its *Power Spectrum*. In fact the power spectrum is, as we will shortly see, the analogous for random fields to what the variance is for random variable. As a Gaussian random variable is entirely described by average and variance, a Gaussian random field will entirely be defined by its power spectrum. To define the power spectrum we start taking the Fourier transform of the field:

$$\delta(x) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, \delta(k) e^{ik\cdot x}, \tag{1.44}$$

the perturbation in real space is a random variable, then, so is its Fourier transform. Furthermore, since the field is stationary and isotropic, the covariance matrix in Fourier space is diagonal, and its elements are defined as:

$$\langle \delta(k)\delta(k')\rangle = (2\pi)^3 P_\delta(k)\delta_D(k-k'), \tag{1.45}$$

where $P_\delta(k)$ is the perturbation Power Spectrum: it is defined as the diagonal elements of the covariance matrix of the perturbation in Fourier Space and, if the perturbation field is Gaussian, the Power Spectrum entirely defines it.

The *Wiener-Khintchine theorem* states that the Power spectrum is the Fourier transform of the correlation function:

$$\xi(r) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, P_\delta(k) e^{ik\cdot r}. \tag{1.46}$$

The Power Spectrum is directly related to variance of the field:

$$\sigma^2 = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} P_\delta(k), \tag{1.47}$$

the variance of the field is so defined by the three dimensional integral on all scales k of the Power Spectrum. Before concluding this short digression on random fields, we would like to stress that, whereas here we have stressed Cosmological perturbation as Gaussian fields, in the following of this thesis we will focus our attention on non-Gaussianity. This apparent contradiction is due to the fact that some degree of non-Gaussianity is allowed by theories of primordial Inflation, but is expected to be always very small. So to study the evolution of the Universe, formation of structure, and cosmological parameters, we can safely work assuming that everything is Gaussian. On the other hand, detecting a tiny primordial non-Gaussian component in itself would be hugely important, as it could shed light on the Physics of the Early Universe.

1.3 Statistical Techniques

The starting point for most problems of parameter inference in statistic is the so called *Likelihood Function* $\mathcal{L}$. Assume we have a theoretical model depending on a set of parameters $\{\lambda_i\}$. We also have a set of measurement $\{x_i\}$, and our data depend on the parameter we want to estimate. We denote the probability of A given B as $P[A|B]$, so that we define the Likelihood function as:

$$\mathcal{L} = P[x|\lambda_i], \tag{1.48}$$

it represents the probability to get a certain data given the parameter. Before carrying out our experiment we have some preliminary degree of knowledge of the parameter we want to measure. This is described in terms of a starting probability distribution $P[\lambda_i]$, wich is called the *prior* [1]. If we have no preliminary information on the parameters we want to measure (as it is often the case), then we can define an uninformative prior, expressing the fact that we do not have any preference for picking some specific value of a parameter over the others. In many cases, it is intuitive to see how a totally uninformative prior is simply a uniform distribution over the allowed range of λ_i, that

[1] In our treatment we are defining probability as a "degree of belief", and treating parameters as random variables. That defines the so called *Bayesian* approach to parameters inference, as opposed to the "frequentist" approach, where parameters are just numbers, probability is defined as the frequency of occurrence of an event over many trials, and the concept of a parameter prior simply cannot be defined

is $P[\lambda_i] = k$, where k is a constant used to normalize $P[\lambda_i]$ to 1 over the domain $[\lambda_{min}, \lambda_{max}]$. Given two random variables A and B the *Bayes theorem* states that:

$$P[A|B] = \frac{P[B|A]P[A]}{P[B]}. \tag{1.49}$$

We can apply this crucial result to our problem, in order to get the probability distribution of our parameter, given the measured Likelihood and the prior:

$$P[\lambda_i|x] = \frac{P[x|\lambda_i]P[\lambda_i]}{P[x]}, \tag{1.50}$$

where $P[\lambda_i]$ is the prior and $P[x]$ is called *evidence* and simply acts as normalization. Since it is independent of the parameters, the evidence can be omitted as long as we are concerned with parameter inference problems. The probability distribution of the parameters given the data, is called *posterior*. A natural estimate of λ_i is the the choice which maximizes the posterior. Note that, if we have an uniform prior from (1.50) we can write:

$$P[x|\lambda_i] = \mathcal{L} \propto P[\lambda_i|x]. \tag{1.51}$$

That means that the choice of the parameters which maximizes the posterior is also the one that maximizes the Likelihood. Our parameter estimator is thus a *Maximum Likelihood* estimator. This estimator plays a crucial role in parameters inference, as can be shown that in many case it is the "Best Unbiased Estimator" (BUE) of a set of parameters given the data. Unbiased means that the average is equal to the true value of the parameter. "Best" means that it gives the most accurate estimate of the parameters, in a sense that we are now going to clarify better in order to define the uncertainty on the parameters, or rather the "error bars" (we stressed that we are still assuming an uniform prior). Our estimate is derived from the peak of the Likelihood, so its natural to assume, as a measurement of the precision of our measurement, the width of the Likelihood itself, determined in its maximum. The "width" is nothing else than the curvature of the Likelihood: if it is strongly peaked,*i.e.* the curvature is large, the errors on the estimates will be smaller. On the contrary, if around the peak the Likelihood is rather flat,*i.e.* the curvature is small, it means that the errors are larger. Performing a Taylor expansion of the logarithm of the Likelihood around the peak values

$\{\hat{\lambda}_i\}$ we can approximate it with a multivariate Gaussian:

$$\ln \mathcal{L}(x, \lambda) = \ln \mathcal{L}(x, \hat{\lambda}) + \frac{1}{2}(\lambda_i - \hat{\lambda}_i)\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i \partial \lambda_j}(\lambda_j - \hat{\lambda}_j), \tag{1.52}$$

from which it derives:

$$\mathcal{L}(x, \lambda) = \mathcal{L}(x, \hat{\lambda}) \exp\left[-\frac{1}{2}(\lambda_i - \hat{\lambda}_i)H_{ij}(\lambda_j - \hat{\lambda}_j)\right]. \tag{1.53}$$

Here $H_{ij} = -\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i \partial \lambda_j}$ is the *Hessian matrix*, which quantifies the curvature of the Likelihood, so it is the key to compute the uncertainties on the parameter. The Hessian matrix quantifies also how much the estimates on the parameters λ_i and λ_j are correlated, if it is diagonal the measurements of the two parameters are uncorrelated. Note that this doesn't means that the parameters itself are correlated, but that so are their estimates: for example the parameters can be completely independent but they can have a similar effect on the data.

If we fix all the parameters except one, or equivalently if we consider the case of a totally independent estimate on a parameter, the error is given by the curvature along the axes representing this parameter (in parameter space), this is called *conditional error* and is defined by:

$$\sigma_{cond,i} = \frac{1}{\sqrt{H_{ii}}}, \tag{1.54}$$

however this error is rarely relevant since is not realistic assumes a totally independent parameter or an infinite precision on the determination of the other. To obtain the marginal error on a parameter, taking into account all the correlation, we have to compute the inverse of the Hessian matrix, so we have:

$$\sigma_i = \sqrt{H_{ii}^{-1}}. \tag{1.55}$$

To estimate the error for a future experiment what is used is the expectation value of the Hessian matrix. This quantity is called *Fisher Matrix*:

$$F_{ij} = \langle H_{ij} \rangle = \langle -\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_i \partial \lambda_j} \rangle. \tag{1.56}$$

There is an important lower limit to error bars derived from the so called *Cramer-Rao inequality*. It states that for any unbiased estimator the minimum uncertainty on the

parameters is given by:

$$\sigma_\lambda = \frac{1}{\sqrt{F_{\lambda\lambda}}}, \tag{1.57}$$

and if other parameters are derived from the same data, we have the *marginal error*:

$$\sigma_\lambda = \sqrt{F^{-1}_{\lambda\lambda}}. \tag{1.58}$$

We are now finally in the condition to quantify our previous statement on the Maximum Likelihood estimator, namely that it is the "best" estimator. By best we mean that the errors bars obtained from a Maximum Likelihood procedure, using the procedure outlined just above, saturate the lower limit provided by the Cramer-Rao inequality. No estimator can thus obtain errors bar smaller than the Maximum Likelihood estimator. Hence its central role in parameter estimation theory. A common case is that in which the observables O^{obs} we measure come from a Gaussian distribution with mean $O(\lambda)$ and variance σ. We assume that our observations are unbiased and the mean of distribution corresponds to the real value. The Likelihood is

$$\mathcal{L} = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \sum_{i=1}^{N} - \frac{\left[O_i^{obs} - O_i(\lambda)\right]^2}{2\sigma_i^2} \equiv e^{\frac{\chi^2}{2}}, \tag{1.59}$$

in our case the observable will be the angular averaged bispectrum $B_{\ell_1\ell_2\ell_3}$ and the summation will be on all possible combination of indices. For a Gaussian Likelihood minimizing the χ^2 is equivalent to maximize the Likelihood. Let us study the curvature of χ^2 about its minimum. We have

$$\chi^2(\lambda) = \chi^2(\hat{\lambda}) + \frac{1}{2} \left.\frac{\partial^2\chi^2}{\partial\lambda^2}\right|_{\lambda=\hat{\lambda}} (\lambda - \hat{\lambda})^2. \tag{1.60}$$

The coefficient of the quadratic term is the curvature of the Likelihood. In this approximation, the Likelihood around the peak is Gaussian, hence we can compute the curvature from formula (1.60), so that we have:

$$\frac{1}{2} \left.\frac{\partial^2\chi^2}{\partial\lambda^2}\right|_{\lambda=\hat{\lambda}} = \mathcal{F} = \sum_i \frac{1}{\sigma_i^2} \left[\left(\frac{\partial O_i(\lambda)}{\partial\lambda} \right)^2 + \left(O_i(\lambda) - O^{obs} \right) \frac{\partial^2 O_i(\lambda)}{\partial\lambda^2} \right], \tag{1.61}$$

the second term in the summation can be ignored since we assume that our observations are unbiased: if this is true the difference will oscillate around zero resulting in cancellations. With this hypothesis, the curvature can be replaced by the Fisher matrix.

Generalizing to many parameters we get:

$$F_{\alpha\beta} = \sum_i \frac{1}{\sigma_i^2} \frac{\partial O_i(\lambda)}{\partial \lambda_\alpha} \frac{\partial O_i(\lambda)}{\partial \lambda_\beta}. \tag{1.62}$$

This is a very powerful results, it states that the Fisher matrix can be computed with no data, so for a given fiducial choice of the parameters it is possible to estimate the expected error bars on a parameter before the experiment is performed.

Chapter 2

CMB Analysis

For decades after its discovery, the only known anisotropy in the Cosmic Microwave Background was the dipole due to the Earth motion (Smoot et al. 1977). Starting from the 1990s, however, technological progress, especially space telescopes, allowed the detection of $\mathcal{O}(10^{-5})$ fluctuations in its temperature field (Smoot et al. 1992). This discovery marked a key moment in the development of the modern cosmological model, lifting a veil on the very early Universe and enabling astrophysicists to constrain a number of cosmological parameters with unprecedented accuracy. The most important feature of these anisotropies is that they are small enough to consider their evolution in linear regime. This entails that the correspondence between their sources and what we observe on the Last Scattering Layer is easy to model and understand analytically.

The main tool used in the study of these fluctuations is the *angular power spectrum* C_ℓ, which is essentially the variance derived from the spherical harmonic expansion of the temperature distribution. The comparison between observed C_ℓ and theoretical ones provides accurate information about the physics of the primordial plasma of baryons and photons.

In the pursuit for non-Gaussianities instead, the two-point function and the power spectrum are no longer sufficient to characterize the Cosmological field of interest. As we saw in section 1.2, the power spectrum completely defines a Gaussian field. If the field is no longer Gaussian, the distribution cannot be described only in terms of the power spectrum, but also the higher correlators are needed. The first of these correlators, *i.e.* the the three-point correlator in harmonic space, the Bispectrum, is the statistic most sensitive to a non-Gaussian signal.

In the first section we will review the most important notions about CMB analysis: the

spherical harmonics representation of a random fields on the sphere and the connection between the angular power spectrum and the properties of the primordial field. In the next section we will then introduce the basic concepts about the measurement of non-Gaussianity of the CMB field and the estimation of the bispectrum.

2.1 Power Spectrum

As we saw in the previous chapter, the CMB presents the most perfect black body spectrum in nature. Furthermore, it is highly isotropic with mean temperature $T = 2.753K$, so that we can describe it in terms of small temperature fluctuations $\Theta(\vec{x}, \hat{n}, \tau) = \Delta T/T$, where $\vec{x}$ is the spatial coordinate, $\hat{n}$ is the direction of the incoming photons and τ is the conformal time, so a local observer today has $\vec{x} = 0$ and $\tau = 0$. It's important to emphasize that, deriving from quantum fluctuations, $\Theta(\vec{x}, \hat{n}, \tau)$ is a random field and it's impossible to predict its magnitude in a deterministic way. For this reason what we observe and study is its statistical distribution over the full sky. CMB fluctuations lie over the surface of a sphere, for which the spherical harmonics are a complete set of orthonormal functions. They can be expressed as a Fourier series writing:

$$\Theta(\vec{x}, \hat{n}, \tau) = \sum_{\ell=0}^{\infty} \sum_{m=\ell}^{\ell} a_{\ell m}(\vec{x}, \tau) Y_\ell^m(\hat{n}). \tag{2.1}$$

We now want to derive a useful relation which links the multipoles $a_{\ell m}$ on the right hand side of (2.1) to the Fourier coefficients of the temperature field on the left hand side. We thus perform a plane-wave expansion for the perturbation at a given point:

$$\Theta(\vec{x}, \hat{n}, \tau) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} e^{i(\vec{x}\cdot\vec{k})} \Theta(\vec{k}, \hat{n}, \tau) \equiv \int \frac{\mathrm{d}^3 k}{(2\pi)^3} e^{i(\vec{x}\cdot\vec{k})} \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell+1) \Theta_\ell(\vec{k}, \tau) P_\ell(\vec{k}\cdot\hat{n}), \tag{2.2}$$

where the functions $P_\ell(\vec{k}\cdot\hat{n})$ are the *Legendre Polynomials* (see appendix B.1.1). From the addition theorem for spherical harmonics we know that:

$$P_\ell(\vec{k}\cdot\hat{n}) = \frac{4\pi}{2\ell+1} \sum_{m=\ell}^{\ell} \overline{Y_\ell^m}(\hat{k}) Y_\ell^m(\hat{n}), \tag{2.3}$$

then from direct substitution we can derive:

$$\Theta(\vec{x},\hat{n},\tau) \equiv 4\pi \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, e^{i(\vec{x}\cdot\vec{k})} \sum_{\ell=0}^{\infty}(-i)^\ell \Theta_\ell(\vec{k},\tau) \sum_{m=\ell}^{\ell} \overline{Y_\ell^m}(\hat{k}) Y_\ell^m(\hat{n}). \tag{2.4}$$

Equating (2.1) and (2.4) we obtain the definition of the coefficients of the CMB temperature multipoles as functions of the Legendre coefficients on the temperature field in Fourier space:

$$a_{\ell m}(\vec{x},\tau) = 4\pi(-i)^\ell \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, e^{i(\vec{x}\cdot\vec{k})} \Theta_\ell(\vec{k},\tau)\overline{Y_\ell^m}(\hat{k}), \tag{2.5}$$

this formula will come handy later on.

Since for $\Theta(\vec{x},\hat{n},\tau)$ it's impossible to make predictions about any particular $a_{\ell m}$, what is studied is the distribution from which they are drawn. The mean value is zero for all $a_{\ell m}$, so if they are Gaussian (and to a good approximation they are), they are fully characterized by their variance. The covariance matrix of these coefficients defines the angular power spectrum:

$$\langle a_{\ell m}\overline{a_{\ell' m'}}\rangle = C_\ell \delta_{\ell\ell'}\delta_{mm'}, \tag{2.6}$$

where the average is performed over a theoretical ensemble of many different realization of Universe. Unfortunately there is only one CMB sky from where we can measure the $a_{\ell m}$, therefore, in order to estimate the CMB multipoles, we have to rely on an "ergodic approximation", assuming that averaging over the azimuthal number m (*i.e.* over different multipole orientation on the sky) is equivalent to averaging on the full ensemble. This is of course justified by rotational invariance, which make the variance C_ℓ independent of m. Since, for a fixed ℓ, we have only $2\ell+1$ elements in our sample, we obtain a fundamental lower limit to the precision with which we can determine the C_ℓ. This is called *cosmic variance* and is defined as:

$$\left(\frac{\Delta C_\ell}{C_\ell}\right)_{cv} = \sqrt{\frac{2}{2\ell+1}}. \tag{2.7}$$

The counterpart of the angular power spectrum in real space is the angular two-point correlation function of the temperature field, they are related by:

$$\langle\Theta(\hat{n}_1)\Theta(\hat{n}_2)\rangle = \frac{1}{4\pi}\sum_{\ell=0}^{\infty}(2\ell+1)C_\ell P_\ell(\hat{n}_1\cdot\hat{n}_2). \tag{2.8}$$

In the C_ℓ lies a huge amount of information about the primordial potential, the physics

of the primordial plasma at recombination and the evolution of the Universe, we have to be able to discern those we are looking for. In order to separate contribution from initial condition (*i.e.* Inflation) and radiative transport effects (*i.e.* micro-physics of the baryon-photon plasma before recombination) we write Fourier counterpart of the fluctuations $\Theta_\ell(\vec{k},\tau)$ in the form:

$$\Theta_\ell(\vec{k},\tau) = \Phi(\vec{k})\Delta_\ell(k,\tau), \tag{2.9}$$

$\Phi(\vec{k})$ is the primordial potential fluctuations that we already encountered in the previous chapter and $\Delta(k,\tau)$ is the solution of a Boltzmann equation derived by keeping into account all gravitational and physical interactions between matter and radiation components in the Universe before recombination. $\Delta(k,\tau)$ is called the *CMB radiation transfer function*.

The Boltzmann equation describes the statistical behavior of a thermodynamic system out of thermodynamic equilibrium. In other words it tells us how the abundance and the energy distribution of every component of the Universe changes in consequence of their interactions. This allows to predict how the perturbations evolve when they cross the horizon and they begin to be affected by micro-physical effects. This evolution is formalized by the transfer function: it quantifies the variations of energy density due to the energy transport and how these affect the CMB temperature field. A full treatment of the Boltzmann equation for CMB anisotropies is beyond the scope of this thesis, and can be found in *e.g.* (Dodelson 2003). The Boltzmann equation is generally solved numerically by public available "Boltzmann code", such as CMBfast and CAMB, yielding the final transfer function and allowing the computation of the CMB power spectrum, as well of the bispectrum, as we will see in the following section.

The power spectrum of initial fluctuations is the two-point correlator of $\Phi(\vec{k})$:

$$\langle\Phi(\vec{k})\Phi(\vec{k'})\rangle = (2\pi)^3 P_\Phi(k)\delta^3(\vec{k}+\vec{k'}). \tag{2.10}$$

To relate it to the angular power spectrum we combine (2.10), (2.5) and (2.6) to get:

$$\langle a_{\ell m}\overline{a_{\ell' m'}}\rangle =$$
$$= (-i)^{\ell+\ell'}(4\pi)^2 \iint \frac{\mathrm{d}^3k}{(2\pi)^3}\frac{\mathrm{d}^3k'}{(2\pi)^3}\langle\Phi(k)\overline{\Phi}(k')\rangle e^{i(\vec{x}\cdot\vec{k}-\vec{x}\cdot\vec{k'})}\Delta_\ell(k,\tau)\Delta_{\ell'}(k',\tau)\overline{Y_\ell^m}(\hat{k})Y_{\ell'}^{m'}(\hat{k}'), \tag{2.11}$$

applying (2.10), for the Dirac delta's proprieties this equation becomes:

$$\langle a_{\ell m}\overline{a_{\ell' m'}}\rangle = (-i)^{\ell+\ell'}(4\pi)^2 \int \frac{\mathrm{d}^3k}{(2\pi)^3} P_\Phi(k)\Delta_\ell(k,\tau)\Delta_{\ell'}(k,\tau)\overline{Y_\ell^m}(\hat{k})Y_{\ell'}^{m'}(\hat{k}). \tag{2.12}$$

The angular part of the integral over the spherical harmonics leaves two δ factor, giving the final formula:

$$\begin{aligned}\langle a_{\ell m}\overline{a_{\ell' m'}}\rangle &= C_\ell\delta_{\ell\ell'}\delta_{mm'} = (-i)^{\ell+\ell'}\frac{2}{\pi}\int \mathrm{d}k\, k^2 P_\Phi(k)\Delta_\ell(k,\tau)\Delta_{\ell'}(k,\tau)\delta_{\ell\ell'}\delta_{mm'},\\ C_\ell &= \frac{2}{\pi}\int \mathrm{d}k\ k^2 P_\Phi(k)\Delta_\ell^2(k,\tau).\end{aligned} \tag{2.13}$$

This relation connects, for a given ℓ, the observable C_ℓ with the theoretical quantity $\Theta_\ell(\vec{k},\tau)$, which depends on the composition of the primordial plasma, and on initial conditions at the end of inflation, providing a fundamental test for the cosmological models.

It's important to notice that at first order the C_ℓ and $P_\Phi(k)$, as well as $a_{\ell m}$ and Φ, are linked by a linear operator, hence if the primordial perturbation field is Gaussian also the CMB temperature field must show the same distribution. This makes the CMB a perfect observable to test primordial non-Gaussianity.

If we were working with galaxies, primordial non-Gaussianity will be concealed by the effects of the highly non-linear evolution undergone by matter distribution in recent eras. In this context the effect of primordial non-Gaussian signal constitutes only a correction on the galaxy bispectrum, so that to measure the primordial signal we have first to distinguish it from non-Gaussianity due to non-linear evolution.

The relations presented in this chapter show that knowing $\Delta_\ell(k,\tau)$ allows to fully determine the C_ℓ, and account for all radiative effects. As we were mentioning earlier, the

calculation of $\Delta_\ell(k,\tau)$ can be done through a full numerical treatment of the Einstein-Boltzmann system.

2.2 The Bispectrum

Measurements of primordial *non-Gaussianity* have been shown to be a very powerful tool to discriminate between different Inflationary models. We know that, in case of a non Gaussian distribution the odd moments are no longer vanishing. In the same way as the variance is related to the Power spectrum and to the two-point correlator, we can relate the third moment, the skewness, to the three-point correlator and to its Fourier transform counterpart, the *bispectrum*. Since it is related to the third moment, also the bispectrum vanishes if the field is Gaussian. Measurements of the bispectrum amplitude thus provide a very sensitive indicator of departures from Gaussianity.

The bispectrum is usually characterized by an amplitude parameter: f_{NL}, where "NL" stands for "non-linear". It is a dimensionless parameter representing the overall amplitude of the curvature bispectrum. Bispectrum measurements are the most promising tool to test small primordial departures from Gaussianity. Other statistics, like the 4-point function, generally are not expected to have the same constraining power (Creminelli et al. 2007). An intuitive argument, which clarifies why this is the case, can be derived by remembering that non-Gaussianity from Inflation is generally very small, and expanding the primordial non-Gaussian field in a Taylor series around its Gaussian component, with f_{NL} providing the amplitude of the three-point function of expanded field. So that we expand the primordial potential in terms of an underlying Gaussian process:

$$\Phi(x) = \Phi_L(x) + f_{NL}(\Phi_L^2(x) - \langle\Phi_L^2(x)\rangle) + g_{NL}\Phi_L^3(x) + \dots , \tag{2.14}$$

here g_{NL} represents the amplitude of the non-Gaussian contribution from the four-point correlator; since the magnitude of the potential fluctuation is $\Phi_L \simeq 10^{-5}$ we see that, unless $g_{NL} >> f_{NL}$, higher order correlator are suppressed by the increasing power of Φ_L. We will see in this chapter that this expansion, called "local", strictly applies only to some Inflationary model, namely, those generating non-Gaussianity of the so called local type. However the general argument holds in nearly all cases.

As for the power spectrum, since it is defined on the surface of a sphere the CMB angular power spectrum is obtained in terms of an harmonic expansion instead of the

classical Fourier transform. Preserving the formalism of the previous section we can now define the angular bispectrum as:

$$B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} = \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle. \tag{2.15}$$

In order to distinguish between geometrical and physical proprieties the bispectrum can be expressed as (Komatsu & Spergel 2001):

$$B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} = \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} b_{\ell_1 \ell_2 \ell_3}, \tag{2.16}$$

here $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ is the *Gaunt integral* and $b_{\ell_1 \ell_2 \ell_3}$ is called the *reduced bispectrum*. The Gaunt integral is defined by:

$$\begin{aligned} \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \equiv & \int \mathrm{d}^2 \hat{n}\ Y^{m_1}_{\ell_1}(\hat{n}_1) Y^{m_2}_{\ell_2}(\hat{n}_2) Y^{m_3}_{\ell_3}(\hat{n}_3) \\ = & \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \end{aligned} \tag{2.17}$$

where the matrix is the Wigner-3j symbol , $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ is a real function which ensured that the bispectrum satisfies the triangle conditions and selections rules: $m_1 + m_2 + m_3 = 0$, $\ell_1 + \ell_2 + \ell_3 = even$ and $|\ell_i - \ell_j| \leq \ell_k \leq \ell_i + \ell_j$; in turn, these conditions guarantee the invariance under translation. For further details about the Wigner symbols proprieties see the appendix B.3.

The fact than we can find a factorization in which the reduced bispectrum is m-independent describes instead rotational invariance (isotropy) of the CMB sky.

As for the power spectrum we can estimate the Bispectrum as an average on all the m values. The observable quantity is thus the angle averaged bispectrum, given by:

$$B_{\ell_1 \ell_2 \ell_3} \equiv \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}. \tag{2.18}$$

Inserting (2.16) in (2.18) and applying the orthogonality relation of the Wigner symbols (B.14), given the definition of $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ (2.17), the angle averaged bispectrum can be expressed as:

$$B_{\ell_1 \ell_2 \ell_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1 \ell_2 \ell_3}. \tag{2.19}$$

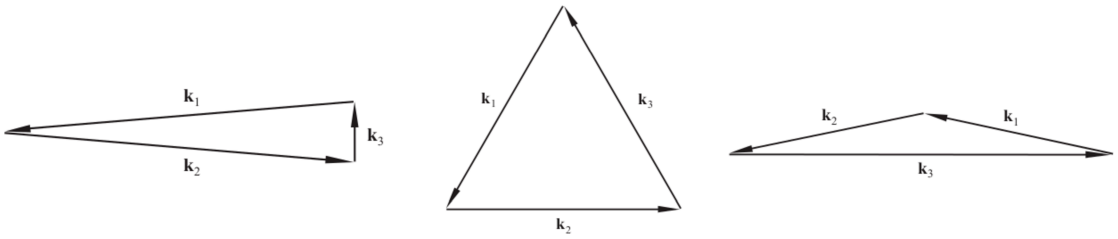

FIGURE 2.1: Triangle types contributing to the bispectrum corresponding to "squeezed" configurations with $k_3 << k_1 = k_2$ (left), equilateral configurations with $k_3 \simeq k_2 \simeq k_1$ (center) and flattened configurations with $k_3 \simeq k_1 + k_2$ (right). Image taken from (Liguori et al. 2010).

2.2.1 Primordial Bispectrum

To obtain information about the primordial potential from the CMB bispectrum we must define in which way the two quantities are linked. Equation (2.19) can be rewritten inserting the definition of angular bispectrum (2.15) and linking $a_{\ell m}$ to Φas in formula (2.5), so that we get:

$$\begin{aligned} B_{\ell_1\ell_2\ell_3} = (4\pi)^3(-i)^{\ell_1+\ell_2+\ell_3} \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times \\ \times \iiint \frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\frac{\mathrm{d}^3k_3}{(2\pi)^3} \overline{Y_{\ell_1}^{m_1}}(\hat{k}_1)\overline{Y_{\ell_2}^{m_2}}(\hat{k}_2)\overline{Y_{\ell_3}^{m_3}}(\hat{k}_3) \times \\ \times \langle \Phi(\vec{k}_1)\Phi(\vec{k}_2)\Phi(\vec{k}_3)\rangle \Delta(k_1,\tau)\Delta(k_2,\tau)\Delta(k_3,\tau), \end{aligned} \tag{2.20}$$

where $\Delta(k,\tau)$ is the transfer function and $\langle \Phi(\vec{k}_1)\Phi(\vec{k}_2)\Phi(\vec{k}_3)\rangle$ is the three-point function of the primordial potential $\Phi(\vec{k})$ that can be defined as (Liguori et al. 2010):

$$\langle \Phi(\vec{k}_1)\Phi(\vec{k}_2)\Phi(\vec{k}_3)\rangle = (2\pi)^2\delta_D^3(k_1+k_2+k_3)B_\Phi(k_1,k_2,k_3). \tag{2.21}$$

$B_\Phi(k_1,k_2,k_3)$ is the primordial bispectrum, as it is function of the primordial potential only. Every model of Inflation predict a different potential $\Phi(\vec{k})$ from which it derives a specific prediction for the primordial bispectrum.

The primordial bispectrum is usually described by the *amplitude parameter* f_{NL} and the *shape function* (Liguori et al. 2010):

$$B(k_1,k_2,k_3) = f_{NL}F(k_1,k_2,k_3) \tag{2.22}$$

here $F(k_1,k_2,k_3)$ is the shape function that defines the dependence of the bispectrum

on particular triangular configurations, the so called *shapes*. The most important bispectrum shapes, encompassing a large number of Inflationary scenarios, are showed in figure 2.1. The reason for which studies of the primordial bispectrum focus on the determination of a single amplitude parameter is that the amplitudes of the individual triangle configurations are way too small to be detected. Cosmic variance is too large to allow the measurement of a single configuration. The signal-to-noise ratio of a single configuration is:

$$\frac{S}{N} \simeq \frac{B_{\ell_1\ell_2\ell_3}}{\sigma} \simeq \frac{B_{\ell_1\ell_2\ell_3}}{\sqrt{C_{\ell_1}C_{\ell_2}C_{\ell_3}}} \simeq \sqrt{C_\ell} \simeq 10^{-5}, \tag{2.23}$$

here $\sigma \propto C_\ell C_\ell C_\ell$ is the variance of the bispectrum.Compressing all the bispectrum information into a single, integrated, amplitude parameter is thus the only way to hope to achieve a sufficiently high signal-to-noise ratio for a potentially significant detection. Do not forget, in fact, that the simplest models of Inflation, characterized by slow-roll a single scalar field, predict $f_{NL} \sim 10^{-2}$, while even models allowing in principle "large" non-Gaussianity make predictions[1] for $f_{NL} \sim 1$ to 10. As one can deduce from formula (2.14), a scale-invariant bispectrum will generally be linked to $P_\Phi(k)$ as

$$f_{NL} \sim \frac{B_\Phi(k,k,k)}{P^2_\Phi(k)}. \tag{2.24}$$

In addition to the f_{NL}, theories predict what kind of triangular configurations provide the greatest contribution to the overall signal (i.e. they say on which triangles the bispectrum in peaked), this is encoded in the shape function $F(k_1, k_2, k_3)$. Various Inflationary Models can be classified according to the configurations they mainly generate in the bispectrum. The most common and studied shapes are the *local shape*, the *equilateral shape* and the *orthogonal shape* (Fergusson et al. 2012; Chen 2010).

Local shape This shape covers a wide range of models, including standard single-field slow-roll (Falk et al. 1993; Gangui et al. 1994; Gangui & Martin 2000; Wang & Kamionkowski 2000; Komatsu & Spergel 2001; Mangilli & Verde 2009). However f_{NL} in standard single-field is predicted to be tiny, and undetectable ($f_{NL} \sim 10^{-2}$). Sizable local non-Gaussianities can be produced in Inflationary model with multiple interacting fields. The deduction of large local non-Gaussianities would thus rule out the simplest slow-roll single-field Inflationary scenarios.

[1] note that, for this order of magnitude estimate of the strength of the signal, we are defining f_{NL} using expansion (2.14)

The local bispectrum is peaked on "squeezed" configuration with $k_1 \simeq k_2 \gg k_3$. This shape derives from the expansion (2.14) we used earlier to show the suppression of higher order correlator. It's called "local" because non-Gaussianities are typically produced on large super-horizon scales (represented by the smallest k). Since different regions on super-horizon scales cannot interact, non-Gaussianity will not correlate different point in real space. This is evident from (2.14), where the non-Gaussian part of the potential is a *local* functional of the Gaussian part, *i.e.* $\Phi_{NL}(x) = f_{NL}(\Phi_L^2(x) - \langle\Phi_L^2(x)\rangle)$ depends only on the value of $\Phi_L(x)$ in the *same* point. The form of the local shape function is:

$$F(k_1, k_2, k_3) = 2A^2 \left[\frac{1}{k_1^{4-n_s} k_2^{4-n_s}} + \frac{1}{k_1^{4-n_s} k_3^{4-n_s}} + \frac{1}{k_2^{4-n_s} k_3^{4-n_s}} \right], \tag{2.25}$$

here A is the normalization of the Power Spectrum and n_s is the spectral index $P_\Phi(k) = A/k^{n_s-4}$.

Equilateral shape For the equilateral bispectrum, the shape function peaks at the equilateral triangle limit $k_1 = k_2 = k_3$ and vanishes as $\sim k_3/k_1$, it has the form (Creminelli et al. 2006):

$$\begin{aligned} F(k_1, k_2, k_3) =& 6A^2 \Bigg[- \frac{1}{k_1^{4-n_s} k_2^{4-n_s}} - \frac{1}{k_1^{4-n_s} k_3^{4-n_s}} - \frac{1}{k_2^{4-n_s} k_3^{4-n_s}} - \frac{2}{(k_1 k_2 k_3)^{2(4-n_s)/3}} + \\ &+ \left(\frac{1}{k_1^{(4-n_s)/3} k_2^{2(4-n_s)/3} k_3^{(4-n_s)/3}} + 5\,\text{perm.} \right) \Bigg]. \end{aligned} \tag{2.26}$$

This kind of bispectrum is generated mainly from single-field Inflationary Models with non-standard kinetic terms in the Inflaton Lagrangian. In single field large wavelenght modes are frozen from the moment they leave the horizon, therefore they can not have large interactions with short wavelength modes. On the other hand modes that are still within the horizon average out any contribution to non-Gaussianities with their oscillations. For this reason the signal in squeezed triangles is now suppressed. Therefore the greatest contribution comes from the modes that have similar wavelengths and exit the horizon at about the same time sourcing an equilateral bispectrum signal.

Orthogonal shape This shape is peaked both on equilateral configurations and flat-triangle configurations (with $k_1 + k_2 \simeq k_3$). The Physical arguments that bring to this

shape are the same of the equilateral one. It is derived from the combination of similar-equilateral shapes (Senatore et al. 2010). It can be approximated by the template:

$$F(k_1,k_2,k_3) = 6A^2\Bigg[-\frac{3}{k_1^{4-n_s}k_2^{4-n_s}}-\frac{3}{k_1^{4-n_s}k_3^{4-n_s}}-\frac{3}{k_2^{4-n_s}k_3^{4-n_s}}-\frac{8}{(k_1k_2k_3)^{2(4-n_s)/3}}+ \\ +\left(\frac{3}{k_1^{(4-n_s)/3}k_2^{2(4-n_s)/3}k_3^{(4-n_s)/3}}+5\,\text{perm.}\right)\Bigg]. \tag{2.27}$$

2.2.2 Factorizability

Measurements of the CMB angular bispectrum are a very powerful tool for the detection of primordial non-Gaussianity. However they also present a huge computational challenge. The bispectrum $B_{\ell_1\ell_2\ell_3}^{m_1m_2m_3}$ has six degrees of freedom. Thus its brute-force calculation would scale like $\mathcal{O}(\ell_{max}^6)$, that becomes $\mathcal{O}(\ell_{max}^5)$ once triangle and parity conditions constraints are kept into account. However this is still huge for a typical modern CMB experiment, with $\ell_{max}\sim 10^3$ or larger. It's now clear that for the analysis of a survey like *Planck*, with $\ell_{max}\sim 2000$, a more efficient method is needed.

The solution is to write the bispectrum in *factorizable* form, allowing to compute three independent integral, one in every k_i, instead of a triple one. As we are going to see, in this way the problem cost reduces to $\mathcal{O}(\ell_{max}^3)$.

A factorizable, or separable form, is defined by an ansatz of the kind:

$$B_\Phi(k_1,k_2,k_3) = f_{NL}F(k_1,k_2,k_3) = f_{NL}\sum_{pqr}c_{pqr}F_p(k_1)F_q(k_2)F_r(k_3) + \text{perm.} \tag{2.28}$$

To obtain the general form for a separable bispectrum, we shall insert (2.28) in equation (2.21), then expand the Dirac delta in plane wave:

$$\delta(\vec{k}_1+\vec{k}_2+\vec{k}_3) = \int\frac{\mathrm{d}^3x}{(2\pi)^3}\; exp\;\imath\big[(\vec{k}_1+\vec{k}_2+\vec{k}_3)\cdot\vec{x}\big], \tag{2.29}$$

and then apply the spherical wave expansion of vector plane wave (for further details see appendix B.2), so we obtain:

$$\begin{aligned}\delta(\vec{k}_1+\vec{k}_2+\vec{k}_3) = (4\pi)^3 \int \frac{\mathrm{d}^3x}{(2\pi)^3} \Bigg\{ \Big[\sum_{\ell_1=0}^{\infty} \imath^{\ell_1} j_{\ell_1}(k_1 r) \sum_{m_1=-\ell_1}^{\ell_1} \overline{Y}_{\ell_1}^{m_1}(\hat{n}_1) Y_{\ell_1}^{m_1}(\hat{k}_1) \Big] \times \\ \times \Big[\sum_{\ell_2=0}^{\infty} \imath^{\ell_2} j_{\ell_2}(k_2 r) \sum_{m_2=-\ell_2}^{\ell_2} \overline{Y}_{\ell_2}^{m_2}(\hat{n}_2) Y_{\ell_2}^{m_2}(\hat{k}_2) \Big] \times \\ \times \Big[\sum_{\ell_3=0}^{\infty} \imath^{\ell_3} j_{\ell_3}(k_3 r) \sum_{m_3=-\ell_3}^{\ell_3} \overline{Y}_{\ell_3}^{m_3}(\hat{n}_3) Y_{\ell_3}^{m_3}(\hat{k}_3) \Big]. \end{aligned} \tag{2.30}$$

Here the coordinate r in the Bessel functions argument represents the look-back conformal time $r = c\tau$. Now returning to the definition (2.20) for the angular bispectrum, we can substitute $\langle \Phi(\vec{k}_1)\Phi(\vec{k}_2)\Phi(\vec{k}_3)\rangle$ with the expanded δ_D and the separable function (2.28), at last we can eliminate the summations recalling the normalization propriety of the spherical harmonics obtaining:

$$\begin{aligned} B_{\ell_1\ell_2\ell_3}^{m_1m_2m_3} &= \left(\frac{2}{\pi}\right)^3 \int \mathrm{d}^3x \iiint \mathrm{d}k_1\mathrm{d}k_2\mathrm{d}k_3\, k_1^2k_2^2k_3^2 \sum_{pqr} \Big[c_{pqr} F_p^{(1)}(k_1) F_q^{(2)}(k_2) F_r^{(3)}(k_3) \\ &\quad + \mathit{perm.} \Big] \Delta_{\ell_1}(k_1,\tau)\Delta_{\ell_2}(k_2,\tau)\Delta_{\ell_3}(k_3,\tau) j_{\ell_1}(k_1r) j_{\ell_2}(k_2r) j_{\ell_3}(k_3r) \times \\ &\quad \times \overline{Y_{\ell_1}^{m_1}}(\hat{n_1}) \overline{Y_{\ell_2}^{m_2}}(\hat{n_2}) \overline{Y_{\ell_3}^{m_3}}(\hat{n_3}) \\ &= \mathcal{G}_{\ell_1\ell_2\ell_3}^{m_1m_2m_3} \times \left(\frac{2}{\pi}\right)^3 \sum_{pqr} \Bigg[\int r^2\mathrm{d}r \int k_1^2\mathrm{d}k_1\, F_p^{(1)}(k_1)\Delta_{\ell_1}(k_1,\tau) j_{\ell_1}(k_1r) \times \\ &\quad \times \int k_2^2\mathrm{d}k_2\, F_q^{(2)}(k_2)\Delta_{\ell_2}(k_2,\tau) j_{\ell_2}(k_2r) \times \\ &\quad \times \int k_3^2\mathrm{d}k_3\, F_r^{(3)}(k_3)\Delta_{\ell_3}(k_3,\tau) j_{\ell_3}(k_3r) + \mathit{perm.} \Bigg]. \end{aligned} \tag{2.31}$$

This is the general factorized form for the angular bispectrum, $F_{p,q,r}$ are now generic functions, their form and their number depend on the model for the primordial potential.

It is now clear why factorizing the shape allows a strong reduction in computational cost. The triple integral $\mathrm{d}k_1\mathrm{d}k_2\mathrm{d}k_3$, over all configurations, has been written as the product of three *one dimensional* integral over $\mathrm{d}k_1$, $\mathrm{d}k_2$, $\mathrm{d}k_3$, separately. This operation is performed in $\sim \ell_{max}$ operations instead of the original $\sim \ell_{max}^3$ operations. With $\ell_{max} \sim 10^3$ this is a huge gain, making the problem numerically tractable.

2.3 KSW estimator

The KSW estimator exploits separability over the wavenumbers k of the bispectrum shape, to achieve high numerical efficiency (Komatsu et al. 2005). Starting from the assumption that the bispectrum configurations follow a Gaussian distribution, maximizing the likelihood is equivalent to minimizing the following χ^2:

$$\chi^2 = \sum_{\ell_1 \leq \ell_2 \leq \ell_3} \frac{\left(B^{obs}_{\ell_1\ell_2\ell_3} - f_{\rm NL} B^{th}_{\ell_1\ell_2\ell_3}(f_{\rm NL}=1)\right)^2}{\sigma^2} \tag{2.32}$$

where $B^{th}_{\ell_1\ell_2\ell_3}$ and $B^{obs}_{\ell_1\ell_2\ell_3}$ are respectively the theoretical and observed angle averaged CMB bispectra, and σ^2 is the variance of the bispectrum, defined as:

$$\sigma^2_{\ell_1\ell_2\ell_3} = \Delta_{\ell_1\ell_2\ell_3} C_{\ell_1} C_{\ell_2} C_{\ell_3} \quad \begin{cases} \Delta_{\ell_1\ell_2\ell_3} = 1 & \ell_1 \neq \ell_2 \neq \ell_3 \\ \Delta_{\ell_1\ell_2\ell_3} = 2 & \ell_i = \ell_j \neq \ell_k \\ \Delta_{\ell_1\ell_2\ell_3} = 6 & \ell_1 = \ell_2 = \ell_3. \end{cases} \tag{2.33}$$

The estimator is derived by inserting the explicit expressions for bispectra, and differentiating with respect $f_{\rm NL}$. Using the notation introduced in (Komatsu & Spergel 2001), the angle averaged CMB bispectrum is expressed as:

$$B_{\ell_1\ell_2\ell_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1\ell_2\ell_3}. \tag{2.34}$$

where the matrix is the Wigner-3j symbol, encoding the geometric conditions while $b_{\ell_1\ell_2\ell_3}$ is the so called "reduced bispectrum", accounting for the shape dependence. We saw that a general separable bispectrum template is written as as a linear combination of products of three one-dimensional functions:

$$B(k_1,k_2,k_3) = f_{\rm NL} \sum_{pqr} c_{pqr} F_p(k_1) F_q(k_2) F_r(k_3) + perm. \tag{2.35}$$

Projecting on the sphere, we obtain the expression for the reduced bispectrum (for the sake of clarity, but without losing in generality, here we assume a shape involving a single set of permutations):

$$b_{\ell_1\ell_2\ell_3} = f_{\rm NL} \int {\rm d}r\, r^2 \left[\, X_{\ell_1}(r) Y_{\ell_2}(r) Z_{\ell_3}(r) + perms. \right]. \tag{2.36}$$

For the local shape we have:

$$b^{loc}_{\ell_1\ell_2\ell_3} = 2f_{\rm NL}\int_0^\infty r^2{\rm d}r\ \left[\alpha_{\ell_1}(r)\beta_{\ell_2}(r)\beta_{\ell_3}(r) + \alpha_{\ell_2}(r)\beta_{\ell_3}(r)\beta_{\ell_1}(r) + \alpha_{\ell_3}(r)\beta_{\ell_1}(r)\beta_{\ell_2}(r)\right], \tag{2.37}$$

where:

$$\alpha_\ell(r) \equiv \frac{2}{\pi}\int_0^\infty k^2{\rm d}k\,\Delta_\ell(k,\tau)j_\ell(kr), \tag{2.38}$$

$$\beta_\ell(r) \equiv \frac{2}{\pi}\int_0^\infty k^2{\rm d}k\,P_\Phi(k)\Delta_\ell(k,\tau)j_\ell(kr). \tag{2.39}$$

Here, $P_\Phi(k)$ is the primordial fluctuations power spectrum and $\Delta_\ell(k,\tau)$ is the radiation transfer function. The field Φ is linked to the curvature perturbations ζ by $\Phi = (3/5)\,\zeta$. The reduced bispectrum for the equilateral shape is instead:

$$b^{equil}_{\ell_1\ell_2\ell_3} = 6f_{\rm NL}\int_0^\infty r^2{\rm d}r\ \left[-2\delta_{\ell_1}\delta_{\ell_2}\delta_{\ell_3} - (\alpha_{\ell_1}\beta_{\ell_2}\beta_{\ell_3} + 2perms.) + (\beta_{\ell_1}\gamma_{\ell_2}\delta_{\ell_3} + 5perms.)\right], \tag{2.40}$$

where, for simplicity of notation, we have avoided to write explicitly the r-dependence of quantities in the integral above. The new functions are defined as:

$$\gamma_\ell(r) \equiv \frac{2}{\pi}\int_0^\infty k^2{\rm d}k\,P_\Phi(k)^{\frac{1}{3}}\Delta_\ell(k,\tau)j_\ell(kr), \tag{2.41}$$

$$\delta_\ell(r) \equiv \frac{2}{\pi}\int_0^\infty k^2{\rm d}k\,P_\Phi(k)^{\frac{2}{3}}\Delta_\ell(k,\tau)j_\ell(kr). \tag{2.42}$$

The orthogonal bispectrum is:

$$b^{equil}_{\ell_1\ell_2\ell_3} = 6f_{\rm NL}\int_0^\infty r^2{\rm d}r\ \left[-8\delta_{\ell_1}\delta_{\ell_2}\delta_{\ell_3} - 3\,(\alpha_{\ell_1}\beta_{\ell_2}\beta_{\ell_3} + 2perms.) + 3\,(\beta_{\ell_1}\gamma_{\ell_2}\delta_{\ell_3} + 5perms.)\right]. \tag{2.43}$$

Under these assumptions, the general form of the "cubic" estimator is:

$$\mathcal{E}^{cubic} = \frac{1}{\mathcal{N}}\sum_{\ell_1\ell_2\ell_3}\frac{B^{th}_{\ell_1\ell_2\ell_3}(f_{\rm NL}=1)B^{obs}_{\ell_1\ell_2\ell_3}}{C_{\ell_1}C_{\ell_2}C_{\ell_3}} = \frac{1}{\mathcal{N}}\sum_{\ell_1\ell_2\ell_3}^{m_1m_2m_3}\frac{\mathcal{G}^{m_1m_2m_3}_{\ell_1\ell_2\ell_3}b^{f_{\rm NL}=1}_{\ell_1\ell_2\ell_3}}{C_{\ell_1}C_{\ell_2}C_{\ell_3}}a_{\ell_1m_1}a_{\ell_2m_2}a_{\ell_3m_3} \tag{2.44}$$

where $\mathcal{N}$ is a normalization factor and $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ is the Gaunt integral, encoding the geometric properties and defined as:

$$\mathcal{N} = \sum_{\ell_1 \ell_2 \ell_3} \frac{\left(B^{th}_{\ell_1 \ell_2 \ell_3}(f_{\rm NL}=1)\right)^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \tag{2.45}$$

$$\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \equiv \int \mathrm{d}^2\hat{n}\; Y^{m_1}_{\ell_1}(\hat{n}) Y^{m_2}_{\ell_2}(\hat{n}) Y^{m_3}_{\ell_3}(\hat{n}). \tag{2.46}$$

Furthermore, in case of partial sky coverage, the rotational invariance is broken. This introduces a spurious NG signal. As a consequence the estimator becomes sub-optimal. To correct for this effect, an additional "linear" term must be added (Creminelli et al. 2007):

$$\mathcal{E}^{lin} = -\frac{3}{f^{sky}\mathcal{N}} \sum_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} \frac{\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} b^{f_{\rm NL}=1}_{\ell_1 \ell_2 \ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle a_{\ell_3 m_3}. \tag{2.47}$$

where f^{sky} is the fraction of the sky covered by the survey, and the brackets represent ensemble average. The best fit value of $f_{\rm NL}$ is therefore:

$$\hat{f}_{\rm NL} = \frac{\mathcal{E}^{cubic}}{f^{sky}} + \mathcal{E}^{lin}. \tag{2.48}$$

To obtain a numerical efficient expression, is useful to define the filtered maps:

$$\begin{aligned} \mathrm{M}_X(r,\hat{n}) &= \sum_{\ell,m} \frac{a_{\ell m} Y^m_\ell(\hat{n})}{C_\ell} X_\ell(r), \\ \mathrm{M}_Y(r,\hat{n}) &= \sum_{\ell,m} \frac{a_{\ell m} Y^m_\ell(\hat{n})}{C_\ell} Y_\ell(r), \\ \mathrm{M}_Z(r,\hat{n}) &= \sum_{\ell,m} \frac{a_{\ell m} Y^m_\ell(\hat{n})}{C_\ell} Z_\ell(r). \end{aligned} \tag{2.49}$$

These maps can be computed resorting to fast harmonic transform. Therefore, we can write the estimator as:

$$\mathcal{E}^{cubic} = \frac{1}{\mathcal{N}} \int_0^\infty \mathrm{d}r\, r^2 \int \mathrm{d}^2\hat{n}\, \mathrm{M}_X(r,\hat{n}) \mathrm{M}_Y(r,\hat{n}) \mathrm{M}_Z(r,\hat{n}) + perms. \tag{2.50}$$

$$\mathcal{E}^{lin} = -\frac{3}{f^{sky}\mathcal{N}} \int_0^\infty \mathrm{d}r\, r^2 \int \mathrm{d}^2\hat{n}\, \mathrm{M}_X(r,\hat{n}) \langle \mathrm{M}_Y(r,\hat{n}) \mathrm{M}_Z(r,\hat{n}) \rangle + perms. \tag{2.51}$$

This expression is immediately implementable and numerically efficient. This estimator has been widely used in the literature to extract the amplitude of the NG signal from CMB maps.

Chapter 3

CMB constraints on running non-Gaussianity

The content of this chapter has been published in (Oppizzi et al. 2018a). Primordial cosmological non-Gaussianity (NG) is nowadays constrained very tightly by Cosmic Microwave Background (CMB) data, with the most stringent bounds coming from measurements of the angular bispectrum and trispectrum of *Planck* temperature and polarization maps (Planck Collaboration et al. 2014b, 2016b). Besides the most common "local", "equilateral" and "orthogonal" NG shapes, arising from a large variety of either single (equilateral, orthogonal) or multi-field (local) inflationary scenarios, many additional bispectra have been tested in the *Planck* analysis. A non-exhaustive list includes anisotropic, flattened and parity-odd shapes, as well as a large family of strongly scale-dependent, oscillatory models. However, to date, no analysis of *Planck* data includes scale-dependent models with a mild $f_{\rm NL}$ running, described by a non-Gaussian spectral index.

The main goal of this paper is to build a set of bispectrum estimators for a complete set of theoretically motivated running NG models. The only existing experimental constraint on running NG, to the best of our knowledge, was obtained by the authors of (Becker & Huterer 2012). They considered a local-type running shape that is expected to appear in curvaton and modulated reheating scenarios. In this work we will extend the analysis in various directions. First of all, we will include additional shapes, peaking both in the local and equilateral limit and generated in different multi-field inflationary models and in Dirac-Born-Infeld (DBI) scenarios. For the latter, we will implement the full, non-separable, shape reported in theoretical derivations and carefully compare it to

the separable phenomenological parametrization that was employed in previous forecast analyses (Sefusatti et al. 2009). Finally, before obtaining our experimental constraints, we will pay particular attention to the validation of our pipeline, via generation of NG CMB maps, including all the running bispectra under examination.

The paper is organized as follows: in section 3.1, we will provide a short overview of the models included in our analyses, focusing on their primordial bispectrum predictions; in section 3.2, we will describe in detail our data analysis pipeline, including bispectrum estimation and generation of NG maps; in section 3.3 we will show the outcome of validation tests, final WMAP experimental bounds on all running shapes and forecasts for future CMB surveys; we will finally summarize our main results in section 3.4.

3.1 Scale-dependent models

In this work we will consider inflationary models that produce a running of the NG parameter $f_{\rm NL}$, parametrized via a NG spectral index $n_{\rm NG}$. Running primordial NG can be sourced by a wide range of different physical processes, such as non-linear evolution of perturbations, interactions in multi-field inflation, variation of the sound speed in single-field inflationary models, peculiar properties of the background metric. Note that scale-dependence can arise also in very simple models and can therefore be considered as a fairly general prediction of Inflation.

It is actually not possible to encompass such a variety of scenarios using just a single bispectrum shape, with a specific ansatz for $f_{\rm NL}(k)$. The aim of this work is to build a quite general class of CMB bispectrum estimators, which can account for most of the theoretically motivated, scale-dependent NG parametrizations proposed so far in the literature. We start in this section by briefly reviewing them. Scale-dependent (SD) NG in the context of slow-roll inflation was studied for example in (Byrnes et al. 2010; Shandera et al. 2011; Tzavara & van Tent 2013), where the authors propose explicit expressions for the primordial three point function, in the cases of one or two fields contributing to the curvature perturbations. The SD local generalization, when only one field contributes to primordial perturbations reduces to:

$$B(k_1,k_2,k_3) \propto f_{\rm NL}\left[(k_1k_2)^{n_\zeta-4}k_3^{n_{\rm NG}} + 2 \text{ perms.}\right], \tag{3.1}$$

where n_ζ denotes the usual spectral index of curvature perturbations ζ. This shape

was constrained using WMAP data in (Becker & Huterer 2012). It describes multi-field models (e.g curvaton or modulated reheating) in which the Inflaton contribution to perturbations is subdominant. A large scale-dependence arises here as a consequence of strong self interactions of the field (Byrnes et al. 2010) . In the following, we will use the superscript "*1f*" to address this model.

If two fields contribute to the curvature perturbation, the dependence on k follows a different parametrization. This kind of template arises, for example, from the mixed inflaton-curvaton theory (assuming the curvaton field has a quadratic potential), and also in general two-field models when the test field has a quadratic potential (Byrnes et al. 2010). The resulting shape is:

$$B(k_1, k_2, k_3) \propto f_{\rm NL} \left[(k_1 k_2)^{n_\zeta + (n_{\rm NG}/2) - 4} + 2 \text{ perms.} \right]. \tag{3.2}$$

We will refer to this model with the superscript *2f*.

Both these shapes are separable over wavenumbers, therefore it will be fairly straightforward to implement them in a generalized version of the classic KSW bispectrum estimator (Komatsu et al. 2005), as we will discuss in the next section. In the theoretical derivation of the previous two shapes, it was assumed that the fields are slow-rolling and that $|n_{\rm NG} \ln (k_{max}/k_{min})| << 1$, where k_{max} and k_{min} are the largest and smallest wavenumbers included in the analysis. For WMAP, $\ln (k_{max}/k_{min}) \lesssim 7$ so $n_{\rm NG}$ can be at most of order of 0.1. However, in this work, we wish to argue that from a purely phenomenological point of view, the previous two bispectrum shapes are interesting templates to constrain with data even when $n_{\rm NG}$ is larger than 0.1. The value of the running, for this class of models, is proportional to higher order derivatives of the inflationary potential. Since the power spectrum is insensitive (to lowest order) to these quantities, measuring $n_{\rm NG}$ can provide additional information about primordial perturbations. If we move to single-field scenarios with a non-canonical kinetic term, a mild running of the NG can be also produced. A typical such example is DBI-inflation (Chen 2005) (see (Bartolo et al. 2010) for a generalization within effective field theory of Inflation). In this context, the NG amplitude $f_{\rm NL}$ is promoted to a function of the triangular wavenumbers configurations. A first study of the testability of these models is presented in (Sefusatti et al. 2009). In that work the authors proposed a parametrization

assuming a dependence on the geometric mean of the three wavenumbers

$$B(k_1, k_2, k_3) \propto f_{\rm NL} \left(\frac{k_1 k_2 k_3}{\mathbf{k}_{piv}^3} \right)^{n_{\rm NG}/3} F(k_1, k_2, k_3)\,, \tag{3.3}$$

where $\mathbf{k}_{piv}$ is a pivot scale and F is the shape function. Being the scale independent part of the bispectrum, F depends only on the ratios of the three wavenumbers. In the theoretical literature (see e.g. (Chen et al. 2007)) a different parametrization is however generally found, in terms of the arithmetic, rather than the geometric mean. Namely:

$$B(k_1, k_2, k_3) \propto f_{\rm NL} \left(\frac{k_1 + k_2 + k_3}{3\mathbf{k}_{piv}} \right)^{n_{\rm NG}} F(k_1, k_2, k_3). \tag{3.4}$$

The geometric mean parametrization can be seen as an approximation of the theoretical shape, which is expected to work well for equilateral shapes (typical case for DBI, and in more general single-field models), where the significant contribution comes from configurations with $k_1 \sim k_2 \sim k_3$ (in that case of course, the two parametrizations coincide). Its practical advantage lies in its explicit separability. On the other hand, for an accurate measurement, it is important at least to compare the two parametrizations, and explicitly verify their level of correlation. The technical problem with the arithmetic mean is that it is not explicitly separable, that is, the factor $(k_1 + k_2 + k_3)^{n_{\rm NG}}$ is not trivially factorizable. There are many well-known ways in the literature to circumvent this problem, based for example on the modal (Fergusson et al. 2010, 2012) or binned (Bucher et al. 2016; Bucher et al. 2010) decomposition of the shape. In this work, we take a different approach. We stick with a KSW-type estimator, and factorize the shape by resorting to the so-called Schwinger parametrization (Smith & Zaldarriaga 2011)

$$B(k_1, k_2, k_3) \propto f_{\rm NL} \frac{F(k_1, k_2, k_3)}{\Gamma(1 - n_{\rm NG})} \frac{1}{\mathbf{k}_{piv}^{n_{\rm NG}}} \int_0^\infty {\rm d}t\; t^{-n_{\rm NG}} \left[k_1 e^{-t(k_1+k_2+k_3)} + perm \right], \tag{3.5}$$

where Γ is the Gamma function. This form is valid for $n_{\rm NG} < 1$, which is not a very limiting assumption, since all the models predict a running $n_{\rm NG} \lesssim 10^{-1}$. We will refer to this shape with the superscript *"am"*. For an overview of the explicit form of the resulting CMB templates, see Appendix 3.A. The advantage of this approach, in this specific case, is that we will not have to re-expand the shape every time we change $n_{\rm NG}$, while exploring the parameter space.

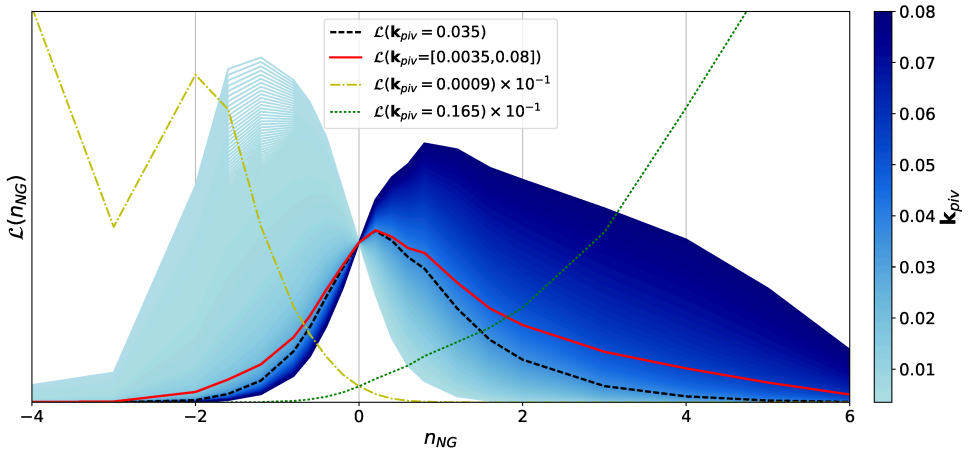

FIGURE 3.1: From light blue to dark blue: likelihood functions for increasing values of the pivot scale, in the range $[0.0035, 0.08]\,\mathrm{Mpc}^{-1}$. Dashed black line: likelihood for the best fit $\mathbf{k}_{piv}$ (see the text). Continuous red line: likelihood averaged over $\mathbf{k}_{piv}$ in the same interval. Dash-dotted yellow line: likelihood for $\mathbf{k}_{piv} = 9 \times 10^{-4}\,\mathrm{Mpc}^{-1}$ rescaled by a factor 10^{-1}. Dotted green line: likelihood for $\mathbf{k}_{piv} = 0.165\,\mathrm{Mpc}^{-1}$ rescaled by a factor 10^{-1}.

3.2 Methodology

We aim to extend the KSW NG estimation technique (Komatsu et al. 2005) (for a comprehensive review about CMB NG estimation see e.g. (Liguori et al. 2010), and references therein) to include all types of bispectra described in the previous section. We follow the indirect approach developed in (Becker & Huterer 2012): we first apply an estimator of the NG amplitude $f_{\rm NL}$, for different, fixed values of the running $n_{\rm NG}$. With this set of measurements we reconstruct the full likelihood function $\mathcal{L}(\hat{f}_{\rm NL}, n_{\rm NG})$, and finally obtain a joint fit of both parameters. A brief description of the standard KSW $f_{\rm NL}$ estimator is reported in Appendix 2.3. We review here its extension to include the running parameter $n_{\rm NG}$. We will follow the approach firstly introduced in (Becker & Huterer 2012), introducing some modifications, that will be discussed in the following. Under the assumption of normally distributed bispectrum configurations, the likelihood can be written as:

$$\mathcal{L}(f_{\rm NL}, n_{\rm NG}) \propto \exp\left[-\frac{1}{2}\sum_{\ell_1\ell_2\ell_3}\frac{\left(B^{obs}_{\ell_1\ell_2\ell_3} - f_{\rm NL}B^{th}_{\ell_1\ell_2\ell_3}(f_{\rm NL}=1, n_{\rm NG})\right)^2}{C_{\ell_1}C_{\ell_2}C_{\ell_3}}\right]. \tag{3.6}$$

Dropping constant terms, this equation can be rearranged in the form:

$$\mathcal{L}(f_{\rm NL}, n_{\rm NG}) \propto \exp\left[-\frac{f_{\rm NL}^2}{2}\sum_{\ell_1\ell_2\ell_3}\frac{\left(B^{th}_{\ell_1\ell_2\ell_3}\right)^2}{C_{\ell_1}C_{\ell_2}C_{\ell_3}} + f_{\rm NL}\sum_{\ell_1\ell_2\ell_3}\frac{B^{obs}_{\ell_1\ell_2\ell_3}B^{th}_{\ell_1\ell_2\ell_3}}{C_{\ell_1}C_{\ell_2}C_{\ell_3}}\right]. \tag{3.7}$$

It is easy to recognize, in the first term, the KSW normalization factor $\mathcal{N}$ (see formula (2.45)) and, in the second term, the unnormalized estimator (see formula (2.44)). We can therefore re-write the expression above as:

$$\mathcal{L}(f_{\rm NL}, n_{\rm NG}) \propto \exp\left[\mathcal{N}\left(-\frac{f_{\rm NL}^2}{2} + f_{\rm NL}\hat{f}_{\rm NL}\right)\right], \tag{3.8}$$

where $\hat{f}_{\rm NL}$ is the value of the NG amplitude recovered from the KSW estimator. Assuming a constant prior on $f_{\rm NL}$, we can integrate to find the marginalized likelihood of $n_{\rm NG}$:

$$\mathcal{L}(n_{\rm NG}) \propto \frac{\mathbf{k}_{piv}^{n_{\rm NG}}}{\sqrt{N}}\exp\left(\frac{\hat{f}_{NL}^2\mathcal{N}}{2}\right), \tag{3.9}$$

here we have explicitly highlighted the dependence of the likelihood on the pivot scale $\mathbf{k}_{piv}$, and N denotes the normalization without the pivot factor. Due to the limitations in resolution, the experimental sensitivity is not constant for different choices of different $\mathbf{k}_{piv}$. As a consequence, the correlation between the two parameters depends on the pivot scale, and this reflects on the shape of the marginalized likelihood. An example of the dependence of $\mathcal{L}(n_{\rm NG})$ on $\mathbf{k}_{piv}$ is shown in figure 3.1, where we consider the one-field model likelihood obtained from a Gaussian simulation (as extreme examples, we show also cases in which we set the pivot outside the accessible scales, resulting in a likelihood that diverges at the edges). The standard approach, (see e.g. (Sefusatti et al. 2009)(Becker & Huterer 2012)), which we also follow here, is to start with an arbitrary value of $\mathbf{k}_{piv}$, compute the likelihood and finally rescale $\mathbf{k}_{piv}$ in order to minimize the correlation between the parameters at the peak of the likelihood. As mentioned above, the estimated amplitude $\hat{f}_{\rm NL}(n_{\rm NG})$ is computed via KSW bispectrum estimation, for fixed $n_{\rm NG}$. The likelihood is then profiled by iterating this operation over a sufficiently wide $n_{\rm NG}$ interval. Finally, with the full profiled likelihood in hand, we can extract the best fit value of $n_{\rm NG}$. In case of partial sky coverage and non-stationary noise, it is well known that a linear term must be added to the KSW cubic statistic, to restore optimality. This is generally computed as a Monte Carlo average over Gaussian realizations of the masked CMB sky, including realistic instrumental noise properties. When producing

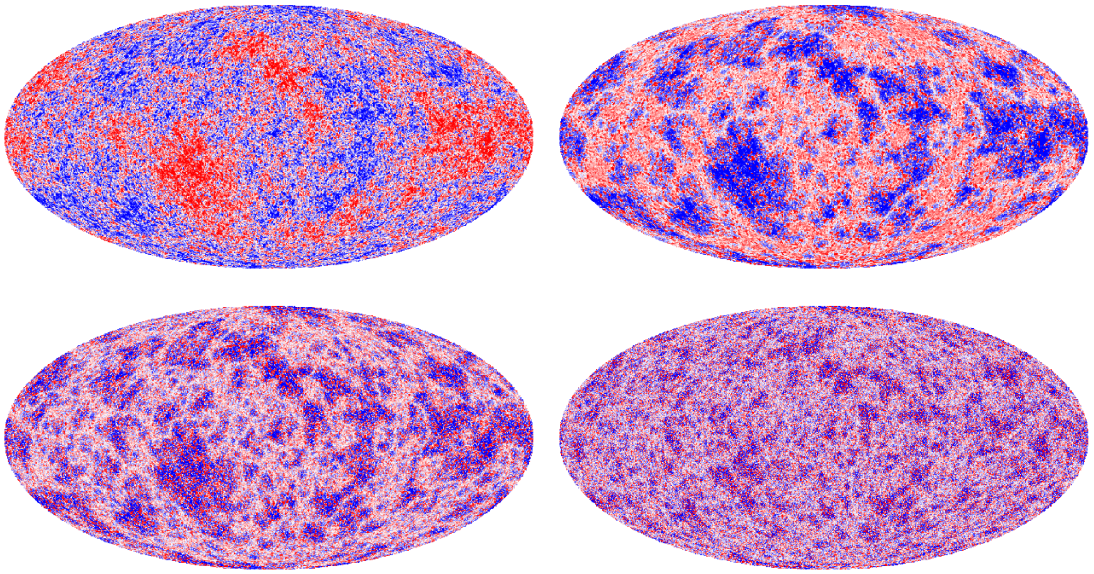

FIGURE 3.2: Random realizations of CMB sky with a SD bispectrum from model (3.1) for different values of the parameters. Upper left panel: Gaussian field ($f_{\rm NL} = 0$). Upper right, lower left and lower right panels: non-Gaussian residuals (i.e. residuals after the subtraction of the Gaussian component of the field) for simulations with, respectively $n_{\rm NG} = -0.2$, $n_{\rm NG} = 0.2$, $n_{\rm NG} = 0.8$.

these realizations, for the WMAP analysis that follows, we adopt the procedure described in (Komatsu et al. 2009), adding to it an inpainting procedure of masked regions, as done in *Planck* data analysis (Planck Collaboration et al. 2014b). For a complete discussion on inpainting and for studies of its efficiency in the context of CMB bispectrum estimation see (Gruetjen et al. 2017; Bucher et al. 2016).

3.2.1 Simulations of non-Gaussian maps

In order to test our estimators, we produce CMB simulations with a scale-dependent $f_{\rm NL}$. We follow the method described in (Smith & Zaldarriaga 2011), which is valid for any separable primordial shape. We can therefore apply it in all the cases under study, since we are always able to find separable representations for our models, eventually via Schwinger parametrization in the technically most complex case. We can express the CMB multipoles as a sum of a Gaussian and a non-Gaussian part, so that:

$$a_{\ell m} = a^{G}_{\ell m} + f_{\rm NL} a^{NG}_{\ell m}, \tag{3.10}$$

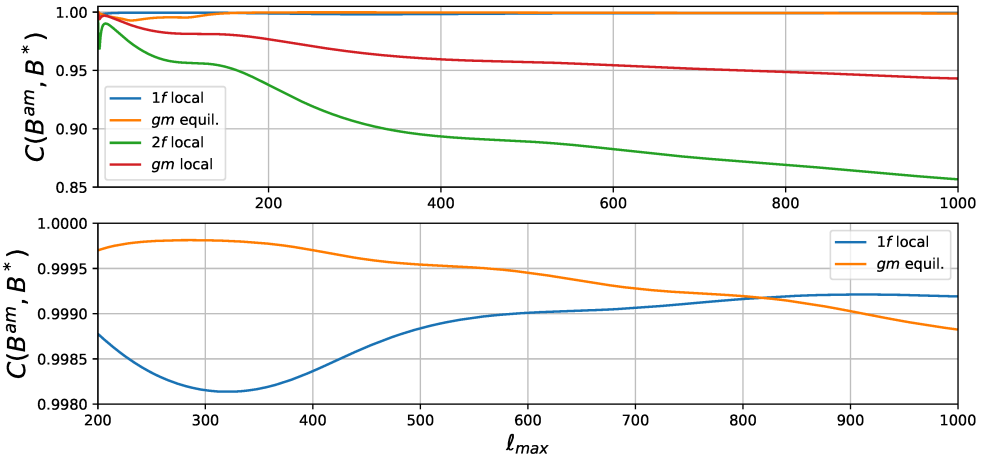

FIGURE 3.3: Correlation between shapes as a function of the maximum multipole number, computed with respect to the arithmetic mean parametrization of the respective template (local for one-field, two-fields and the geometric mean local, equilateral for geometric mean equilateral). All correlations are computed for $n_{\rm NG} = 0.8$. The lower panel shows a zoom on the templates with higher correlation.

where the non-Gaussian component is expressed as:

$$a_{\ell m}^{NG} = \frac{1}{6} \sum_{\ell_1 \ell_2}^{m_1 m_2} B_{\ell \ell_1 \ell_2}^{th} \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ m & m_1 & m_2 \end{pmatrix} \frac{a_{\ell_1 m_1}^G}{C_{\ell_1}} \frac{a_{\ell_2 m_2}^G}{C_{\ell_2}}. \tag{3.11}$$

We compute the NG multipoles inserting in this equation the theoretical separable bispectrum template under study (see appendix 3.A for the expressions). Once computed the $a_{\ell m}^{NG}$, it is straightforward to obtain the final map from Eq. (3.10) for any value of $f_{\rm NL}$. This methods, holding in the limit of small NG, has a drawback in that it can fail in controlling terms $\mathcal{O}(f_{\rm NL}^2)$ in connected N-point functions (Hanson et al. 2009). This can bring a significant, spurious contribution to the power spectrum from the non-Gaussian component. A technique to control and remove this spurious signal has been developed and is described in depth in (Fergusson et al. 2010), to which we refer for further details about NG simulations.

In Figure 3.2 we show the contributions on the CMB sky of a SD bispectrum from model (3.1) for different values of the running. All maps are computed starting from the same Gaussian multipoles. To highlight the differences, we plot the Gaussian and the non-Gaussian components separately. It is evident from these maps how the non-Gaussian modulation peaks on smaller scales for higher values of the running.

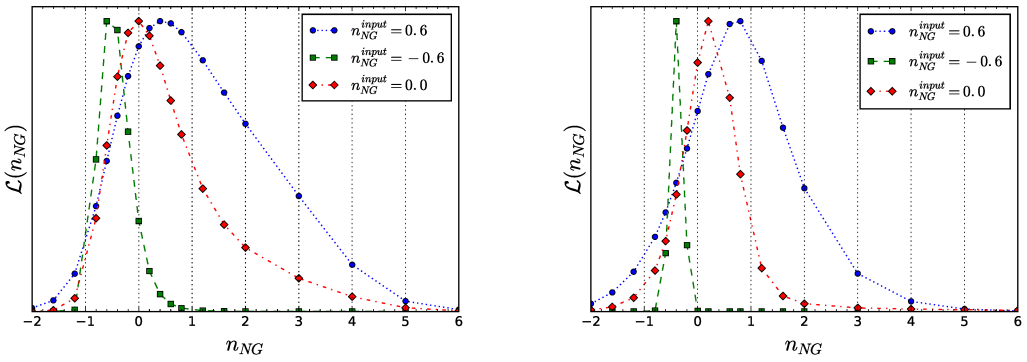

FIGURE 3.4: Likelihood from the one-field (left) and two-field (right) models, for different input value of $n_{\rm NG}$.

3.2.2 Correlation between shapes

We compute the correlation between the shapes described before. The correlation is obviously higher for low values of $n_{\rm NG}$, since all templates reduce to the classical scale invariant case in the limit of zero running. We define the correlation between shapes as the portion of the total NG amplitude recovered fitting a template B' if the true bispectrum is B. The expression of the correlator is:

$$C(B,B') = \frac{1}{N(B,B')} \sum_{\ell_1\ell_2\ell_3} \frac{B_{\ell_1\ell_2\ell_3} B'_{\ell_1\ell_2\ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}, \tag{3.12}$$

where the normalization $N(B,B')$ is:

$$N = \sqrt{\sum_{\ell_1\ell_2\ell_3} \frac{(B_{\ell_1\ell_2\ell_3})^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}} \sqrt{\sum_{\ell_1\ell_2\ell_3} \frac{(B'_{\ell_1\ell_2\ell_3})^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}}. \tag{3.13}$$

This analysis clearly shows that some of the templates under study are highly correlated. As expected, in the equilateral case, the arithmetic mean and the geometric mean parametrizations are almost equivalent, having $C(B^{am}, B^{gm}) \sim 0.999$. In the local case, an high level of correlation is instead displayed between the arithmetic mean parametrization and the one-field model (3.1), where $C > 99.9\%$ for all the multipoles of interest. This behavior is highlighted in figure 3.3, in which we show correlations between the templates, as a function of the maximum multipole considered. For the one-field, two-fields and geometric mean local parametrization the correlations are computed with respect the arithmetic mean parametrization of the local shape. For the geometric mean equilateral shape we show the correlation between the arithmetic mean

equilateral parametrization. To better highlight differences between templates, we consider an extreme value for the running, $n_{\rm NG} = 0.8$. It is also interesting to note how the two-field model (3.2) is highly uncorrelated with the others, showing the importance of separately fitting the different scale-dependent shapes predicted in the literature.

3.3 Results

3.3.1 Test on simulations

To test our estimator, we run it on different sets of NG maps, produced with the method outlined in the previous section. We produce nine different sets, with different spectral index, for three models: the one-field and two-field local models, and the equilateral "geometric mean" model. We choose a value $f_{\rm NL} = 50$ for the local templates and $f_{\rm NL} = 100$ for the equilateral one, at a pivot scale $\mathbf{k}_{piv} = 0.02$ Mpc^{-1}. For each model we consider three different values of the running: $n_{\rm NG} = 0$, $n_{\rm NG} = -0.6$ and $n_{\rm NG} = 0.6$. We compute the Gaussian component assuming the best-fit *Planck* power spectrum. The angular resolution in these test maps is $\ell_{max} = 500$. We test our method both in the case of full sky-coverage and in a more realistic case with 30% of the sky masked.

We find that, in all cases, our estimators recover correctly the initial value of the parameters, within error bars. At the same time, the uncertainties derived from the likelihood are consistent with Fisher matrix predictions. We find that 100 Gaussian simulations are sufficient in the linear term evaluation, to correct for the partial sky coverage effects. As an example, in figure 3.4 we show the likelihoods obtained for the one and two-field local models from simulated NG maps with different value of the spectral index.

3.3.2 Experimental bounds

We apply our technique to provide bounds on the running of non-Gaussianity from the WMAP9 temperature maps, considering the various running models discussed in the previous sections. Our data-set consists on the combination of the V and W WMAP bands, coadded and weighted following the prescription in (Komatsu et al. 2009). To exclude foreground and point-source contaminated pixels we use the KQ75y9 mask, covering 31.2% of the sky. The maps used in this analysis, as well as instrumental specifications, beams and noise per pixel are extensively described in (Bennett et al.

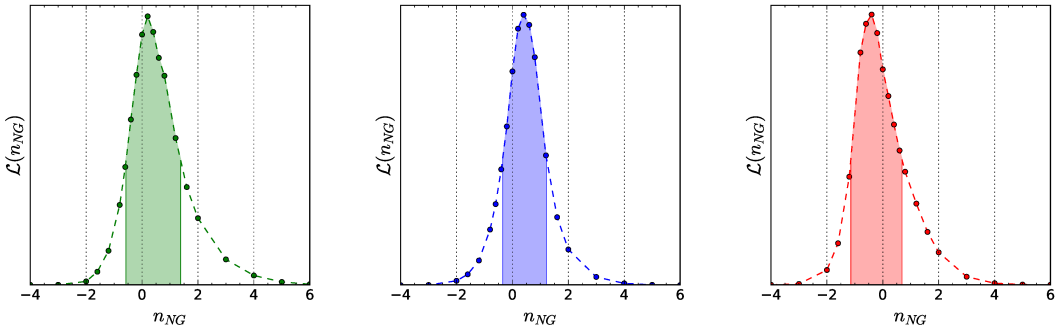

FIGURE 3.5: Marginalized likelihood for the one-field (left, green), two-field (center, blue) models and the geometric mean parametrization for the equilateral model (right, red). The shaded regions represent the 1σ intervals and were found by slicing at constant marginalized likelihood.

model	$n_{\rm NG}$	$\mathbf{k}_{piv}$
one-field (local)	$0.2^{+1.2}_{-0.8}$	$0.035\,{\rm Mpc}^{-1}$
two-field (local)	$0.4^{+0.8}_{-0.7}$	$0.01\,{\rm Mpc}^{-1}$
geometric mean (equil.)	$-0.4^{+1.1}_{-0.7}$	$0.01\,{\rm Mpc}^{-1}$

TABLE 3.1: Experimental constraints for the different models considered in this work. First column: model. Second column: central values and error bars (68% C.L.). Third column: Pivot scale (see section 3.2 for details on the choice of $\mathbf{k}_{piv}$).

2013) and fully available at https://lambda.gsfc.nasa.gov. We obtain our estimates using multipoles up to $\ell_{max} = 800$. To compute the linear correction term (2.47) we use 300 Gaussian realizations obtained with the procedure described in section 3.2, assuming WMAP9 cosmology and instrumental specifications. To further improve the numerical efficiency of our code, we implement the numerical optimization algorithm presented in (Smith & Zaldarriaga 2011). This technique considerably reduces the number of terms needed in the position-space computations of the bispectra. For most models, we start from a conservative number of quadrature points $\mathcal{O}(10^3)$ for the computation of the r integral in (2.44), and reduce it to $\mathcal{O}(10)$ after the optimization. For the arithmetic mean parametrization, which contains an additional integral over the Schwinger parameter t, we reduce the number of terms from $\mathcal{O}(10^4)$ to $\mathcal{O}(10^2)$. The resulting likelihoods for all models are provided in figure 3.5 and the corresponding bounds are summarized in table 3.1. The one-field model (figure 3.5, left) is the only one that was constrained before, using CMB WMAP 7-year data (Becker & Huterer 2012). In that work, the authors found $n_{\rm NG} = 0.30^{+0.78}_{-0.61}$. This is well consistent with our findings, which however lead to a somewhat larger confidence interval. The main source of difference is the shift in $f_{\rm NL}$

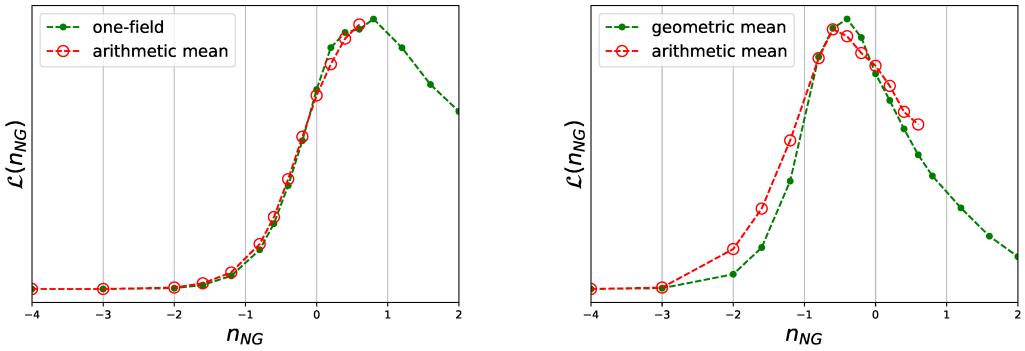

FIGURE 3.6: Comparison between the results from the arithmetic mean parametrization and the corresponding model for local (left) and equilateral (right). Left: green line: one-field local model, red line: arithmetic mean local model. Right: green line: geometric mean equilateral model red line: arithmetic mean equilateral model.

central value, which becomes a bit lower, going from WMAP 7-year to WMAP 9-year data and including inpainting at higher resolution. This lowers the sensitivity to $n_{\rm NG}$ (see also the next section for further discussion on this issue).

In order to obtain the full likelihood, we have in principle to span over the full $n_{\rm NG}$ parameter space, going outside the allowed domain of application of the Schwinger parametrization. Therefore, the arithmetic mean model (3.4) cannot be implemented everywhere. We considered modifications to extend the regime of validity of this parametrization, e.g. using inverse Laplace transform instead of Schwinger parametrization or, more trivially, expanding the non-separable term via multinomial expansion. However these techniques turn out to be too numerical inefficient and unstable for our purposes. However, this is not a big problem. We have in fact explicitly checked (see figure 3.3) that the arithmetic mean template becomes heavily correlated to other models under study: the one-field shape in the local case, and the geometric mean parametrization for the equilateral (note that a reasonably high level of correlation was somewhat assumed, based on general arguments, but not explicitly evaluated and checked in previous works, such as (Sefusatti et al. 2009)). Given current experimental sensitivity and such levels of correlation, fitting the one-field and geometric mean separable templates is therefore equivalent (i.e. leads to the same bounds) to explicitly estimating the arithmetic mean parameterization. This would no longer be the case for very high precision measurements, namely for sensitivities of the order of $\Delta n_{\rm NG} \sim 0.1$. We do not expect such sensitivities to be however achievable, neither with *Planck*, nor with future proposed surveys, due to the very small values of $f_{\rm NL}$ currently constrained by data. As a further

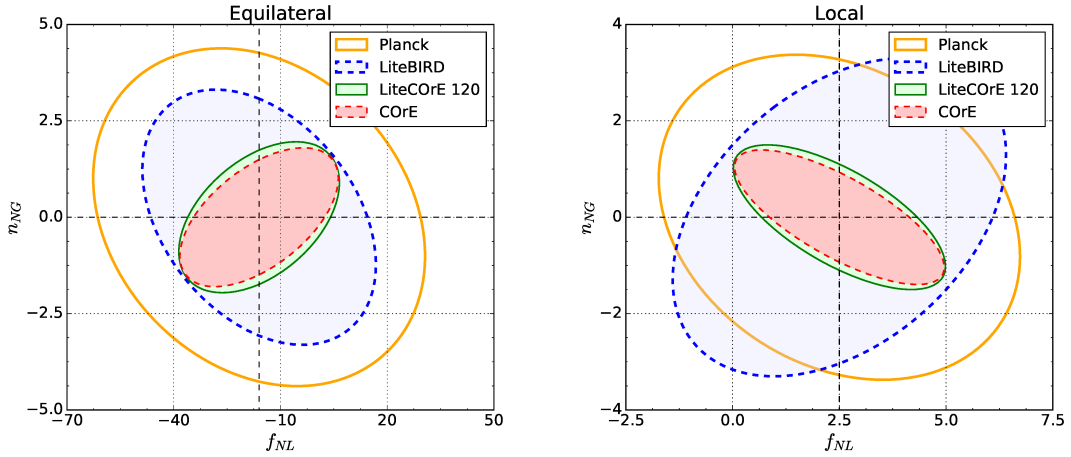

FIGURE 3.7: 1σ error ellipses in the $f_{\rm NL} - n_{\rm NG}$ plane for equilateral (left) and local (right) models arithmetic mean parametrization. We consider a joint temperature and polarization E modes analysis. We assume as central value $f_{\rm NL}^{loc} = 2.5$ and $f_{\rm NL}^{eq} = -16$, $n_{\rm NG} = 0$, $\mathbf{k}_{piv} = 0.055\,{\rm Mpc}^{-1}$. Note that the orientation of the ellipses depends on the sign of $f_{\rm NL}$. We find here positive correlation, for equilateral shapes, between $f_{\rm NL}$ and $n_{\rm NG}$ in higher resolution experiments, because we chose a negative fiducial value for $f_{\rm NL}$.

Experiment	(ℓ_{max})	**local**	**equilateral**
Planck	(2400)	3.4	4.4
LiteBIRD	(1350)	3.3	3.4
LiteCOrE 120	(3000)	1.5	1.9
COrE	(3000)	1.4	1.8

TABLE 3.2: Forecasts for the marginalized 1σ $n_{\rm NG}$ error bars for the arithmetic mean parametrization, assuming joint temperature and polarization analysis. We take as central values $f_{\rm NL}^{loc} = 2.5$ and $f_{\rm NL}^{eq} = -16$, $n_{\rm NG} = 0$, $\mathbf{k}_{piv} = 0.055\,{\rm Mpc}^{-1}$.

verification of this, we show in figure 3.6 the likelihood points from the arithmetic mean model, compared to the two correlated shapes in the equilateral and local cases, over the $n_{\rm NG} < 1$ range, accessible to all templates.

3.3.3 Forecasts

As a final application, we forecast the combined sensitivity to the NG amplitude $f_{\rm NL}$ and the spectral index $n_{\rm NG}$ of the Planck satellite and other proposed CMB projects: LiteBIRD and COrE, in various configurations. For the actual specification used for the various experiments, we refer to (CORE Collaboration et al. 2016) (some $n_{\rm NG}$ forecasts, obtained with the pipeline presented here, where actually already included in (CORE Collaboration et al. 2016)). Figure 3.7 represents the 1σ ellipses in the $f_{\rm NL} - n_{\rm NG}$ plane, for the local and equilateral cases assuming the arithmetic mean parametrization, the corresponding marginalized uncertainties on the running parameter are listed in table

3.2. In this computation we have assumed as fiducial values $n_{\rm NG} = 0$, $f_{\rm NL}^{loc} = 2.5$, $f_{\rm NL}^{eq} = -16$ at a pivot scale of $\mathbf{k}_{piv} = 0.055\,{\rm Mpc}^{-1}$. For all experiments we have considered a sky coverage of 70%. The maximum multipole order accessible is: $\ell_{max} = 2400$ for *Planck*, $\ell_{max} = 1350$ for LiteBIRD, and $\ell_{max} = 3000$ for the other experiments. While these surveys are clearly signal-dominated to much higher ℓ than WMAP, it is not obvious that they will provide significant better constraints on $n_{\rm NG}$. That is because the uncertainty on $n_{\rm NG}$ is inversely proportional not only to $\ell_{\rm max}$ but also to the NG amplitude parameter, $f_{\rm NL}$. The final forecast is therefore crucially dependent on the fiducial value chosen for $f_{\rm NL}$. If we consider the current local scale-independent $f_{\rm NL}$ central value from *Planck* analysis, we know that it has moved much closer to 0, with respect to WMAP, making forecasted *Planck* constraints actually weaker than what we obtained here. A similar reasoning applies also for the equilateral shape, although the shift in measured central values from WMAP to *Planck* is less dramatic there. On the other hand, it is clear that changing the central value of e.g. local $f_{\rm NL}$, in a scale-independent analysis, from approximately 30 at $\ell_{\rm max} = 500$, to $f_{\rm NL} = 2.5$ at $\ell_{\rm max} = 2000$, does display, a posteriori, some degree of running. Allowing for a further running parameter can therefore lead to a shifting of the overall $f_{\rm NL}$ amplitude to larger values. Moreover, for values of $f_{\rm NL}$ which do not correspond to the current best-fit value but are well within the current scale-independent 95% C.L. intervals, significant $n_{\rm NG}$ improvements are expected with future surveys. This is evident from the results shown in table 3.3 and figure 3.8. In summary, while there is a possibility that the constraints obtained here will not be significantly improved with *Planck* or future CMB data, several plausible scenarios do allow for significant tightening of current error bars, up to factors of 2-3. This makes further studies, using more sensitive, higher resolution than WMAP data-sets, clearly worth pursuing.

3.4 Conclusions

Constraining the running of the primordial NG parameter $f_{\rm NL}$ can provide valuable extra-information on the Physics of Inflation, allowing for better discrimination between different scenarios.

In this paper, we presented new constraints on the running of the primordial three-point function, obtained from WMAP 9-year CMB temperature data. We provided bounds on the NG running parameter $n_{\rm NG}$ for different, theory-motivated, bispectrum

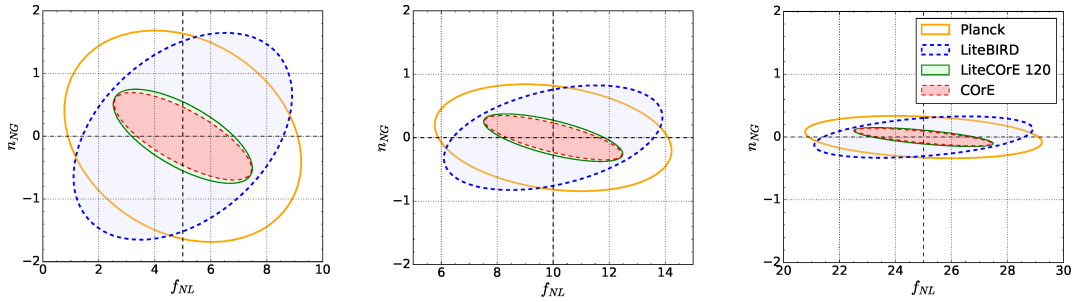

FIGURE 3.8: 1σ error ellipses for the local model arithmetic mean parametrization for different experiments, assuming joint temperature and polarization analysis and different central values of $f_{\rm NL}$, from left to right: $f_{\rm NL} = 5, 10, 25$.

Experiment	(ℓ_{max})	$f_{\rm NL}^{loc} = 5$	$f_{\rm NL}^{loc} = 10$	$f_{\rm NL}^{loc} = 25$
Planck	(2400)	1.7	0.8	0.3
LiteBIRD	(1350)	1.6	0.8	0.3
LiteCOrE 120	(3000)	0.7	0.4	0.1
COrE	(3000)	0.7	0.3	0.1

TABLE 3.3: Forecasts for the marginalized 1σ $n_{\rm NG}$ error bars for the arithmetic mean parametrization of the local case considering different central values for $f_{\rm NL}$. Joint temperature and polarization analysis is assumed.

templates. Of all the models we consider (which summarize all the main running NG scenarios discussed so far in the literature), only the one-field model of equation (3.1) was previously constrained by data (Becker & Huterer 2012). Equilateral shape running scenarios, in particular, were not considered before in data analysis. Moreover, in previous forecasts, the model was approximated using the "geometric mean shape" of equation (3.3), rather than adopting the "arithmetic mean shape", reported in the theoretical literature on the subject (Chen et al. 2007)(Taruya et al. 2008). We consider here the full arithmetic mean model, implementing separability via a Schwinger expansion, explicitly testing its correlation with the geometric mean ansatz and fitting it to data.

Our results were obtained by developing an estimator which extends the standard KSW bispectrum estimator, typically used for studies of the main, scale-independent, local, equilateral and orthogonal shapes. The robustness of our pipeline was successfully tested, using a large set of scale-dependent NG simulations of the CMB sky, generated as part of this work. In our validation, we generated mock data-sets including all the running bispectra under study and we considered various fiducial values of the bispectrum size and running parameters.

Our final constraints (68% C.L.) are $-0.6 < n_{\rm NG} < 1.4$ for the single-field curvaton scenario, $-0.3 < n_{\rm NG} < 1.2$ for two-field curvaton case and $-1.1 < n_{\rm NG} < 0.7$ for DBI.

At this level of sensitivity, information coming from $n_{\rm NG}$ is not particularly useful to set new meaningful bounds on inflationary models. However, the main point of this work was to develop and test a full set of simulation and estimation tools that can be readily applied to future more sensitive data-sets.

In this respect, a natural follow-up of this study, which is actually ongoing work within the *Planck* collaboration, is the application of our pipeline to *Planck* data. Interestingly, we know that local $f_{\rm NL}$, measured with *Planck* as a function of scale, changes from a best-fit value $f_{\rm NL} \sim 40 \pm 20$, at $\ell_{\rm max} = 500$ (fully consistent with WMAP) to $f_{\rm NL} \sim 2 \pm 5$ at $\ell_{\rm max} = 2000$. Fisher matrix forecasts, obtained both for *Planck* and other experimental setups, also show that the error bars obtained in this work could shrink up to a factor ~ 3 with future surveys, depending on the recovered central value of $f_{\rm NL}$. Note also, see figures 3.5, that the actual $n_{\rm NG}$ likelihood can deviate significantly from Gaussianity. This makes the predicted improvements via Fisher analysis an underestimate, for certain intervals of $f_{\rm NL}$ fiducial values, which are consistent with current *Planck* bounds. CMB constraints, obtained via direct bispectrum estimation as presented here, could also in the future be combined with those coming from different probes, such as LSS or, in a more futuristic scenario, CMB spectral distortions and 21 cm anisotropies (Emami et al. 2015; Cooray 2005; Biagetti et al. 2013; Ravenni et al. 2017; Khatri & Sunyaev 2015; Raccanelli et al. 2015), allowing for further, significant improvements.

3.A Scale-dependent templates

The reduced bispectrum for the model (3.1) is a generalization of the local shape (2.37). It has the same expression and the scale-dependence is encoded in the function $\alpha_\ell(r)$, that becomes:

$$\alpha_\ell^{1f}(r) \equiv \frac{2}{\pi}\frac{1}{\mathbf{k}_{piv}^{n_{\rm NG}}}\int_0^\infty k^{2+n_{\rm NG}}\mathrm{d}k\,\Delta_\ell(k,\tau)j_\ell(kr). \tag{3.14}$$

Similarly, the two-field model (3.2) derives from the local template (2.37), but in this case it is the $\beta_\ell(r)$ function that is promoted to a function of $n_{\rm NG}$:

$$\beta_\ell^{2f}(r) \equiv \frac{2}{\pi}\frac{1}{\mathbf{k}_{piv}^{n_{\rm NG}/2}}\int_0^\infty k^{2+n_{\rm NG}/2}\mathrm{d}k\,P_\Phi(k)\Delta_\ell(k,\tau)j_\ell(kr). \tag{3.15}$$

The model (3.3) instead, is valid for the equilateral template (2.40), the new functions being:

$$\alpha_\ell^{gm}(r) \equiv \frac{2}{\pi}\frac{1}{\mathbf{k}_{piv}^{n_{\rm NG}/3}}\int_0^\infty k^{2+n_{\rm NG}/3}\mathrm{d}k\,\Delta_\ell(k,\tau)j_\ell(kr), \tag{3.16}$$

$$\beta_\ell^{gm}(r) \equiv \frac{2}{\pi}\frac{1}{\mathbf{k}_{piv}^{n_{\rm NG}/3}}\int_0^\infty k^{2+n_{\rm NG}/3}\mathrm{d}k\,P_\Phi(k)\Delta_\ell(k,\tau)j_\ell(kr), \tag{3.17}$$

$$\gamma_\ell^{gm}(r) \equiv \frac{2}{\pi}\frac{1}{\mathbf{k}_{piv}^{n_{\rm NG}/3}}\int_0^\infty k^{2+n_{\rm NG}/3}\mathrm{d}k\,P_\Phi(k)^{\frac{1}{3}}\Delta_\ell(k,\tau)j_\ell(kr), \tag{3.18}$$

$$\delta_\ell^{gm}(r) \equiv \frac{2}{\pi}\frac{1}{\mathbf{k}_{piv}^{n_{\rm NG}/3}}\int_0^\infty k^{2+n_{\rm NG}/3}\mathrm{d}k\,P_\Phi(k)^{\frac{2}{3}}\Delta_\ell(k,\tau)j_\ell(kr). \tag{3.19}$$

Obtaining a reduced bispectrum expression for the last model (3.4) is slightly more complex. We have to replace $f_{\rm NL}$ with the new definition (3.5), and put it in separable form, defining new coefficients. For the local shape we compute:

$$\alpha'_\ell(r,t) = \frac{2}{\pi}\int \mathrm{d}k\,k^2\Delta_\ell(k)j_\ell(kr)e^{-tk}, \tag{3.20}$$

$$\beta'_\ell(r,t) = \frac{2}{\pi}\int \mathrm{d}k\,k^2 P_\Phi(k)\Delta_\ell(k)j_\ell(kr)e^{-tk}, \tag{3.21}$$

$$\zeta_\ell(r,t) = \frac{2}{\pi}\int \mathrm{d}k\,k^3\Delta_\ell(k)j_\ell(kr)e^{-tk}, \tag{3.22}$$

$$\xi_\ell(r,t) = \frac{2}{\pi}\int \mathrm{d}k\,k^3 P_\Phi(k)\Delta_\ell(k)j_\ell(kr)e^{-tk}. \tag{3.23}$$

Starting from equation (2.37), with these new coefficients, we can write the reduced scale-dependent local bispectrum as:

$$b^{loc}_{\ell_1\ell_2\ell_3} = \frac{f_{\rm NL}}{\mathbf{k}_{piv}^{n_{\rm NG}}}\frac{1}{\Gamma(1-n_{\rm NG})}\int_0^\infty \mathrm{d}t\, t^{-n_{\rm NG}}\int_0^\infty \mathrm{d}r\, r^2\Big[\Big(\alpha'_{\ell_1}(r,t)\beta'_{\ell_2}(r,t)\xi_{\ell_3}(r,t) + 5\,\text{perm.}\Big) + \\ + \Big(\beta'_{\ell_1}(r,t)\beta'_{\ell_2}(r,t)\zeta_{\ell_3}(r,t) + 2\,\text{perm.}\Big)\Big]. \tag{3.24}$$

In the equilateral case, we need to define other four additional coefficients:

$$\gamma'_\ell(r,t) \equiv \frac{2}{\pi}\int_0^\infty k^2\mathrm{d}k\, P_\Phi(k)^{\frac{1}{3}}\Delta_\ell(k,\tau)j_\ell(kr)e^{-tk}, \tag{3.25}$$

$$\delta'_\ell(r,t) \equiv \frac{2}{\pi}\int_0^\infty k^2\mathrm{d}k\, P_\Phi(k)^{\frac{2}{3}}\Delta_\ell(k,\tau)j_\ell(kr)e^{-tk}, \tag{3.26}$$

$$\epsilon_\ell(r,t) \equiv \frac{2}{\pi}\int_0^\infty k^3\mathrm{d}k\, P_\Phi(k)^{\frac{1}{3}}\Delta_\ell(k,\tau)j_\ell(kr)e^{-tk}, \tag{3.27}$$

$$\eta_\ell(r,t) \equiv \frac{2}{\pi}\int_0^\infty k^3\mathrm{d}k\, P_\Phi(k)^{\frac{2}{3}}\Delta_\ell(k,\tau)j_\ell(kr)e^{-tk}. \tag{3.28}$$

Inserting (3.5) in (2.40) and using all these new coefficients, we obtain the reduced equilateral bispectrum:

$$b^{equil}_{\ell_1\ell_2\ell_3} = \frac{f_{\rm NL}}{\mathbf{k}_{piv}^{n_{\rm NG}}}\frac{1}{\Gamma(1-n_{\rm NG})}\int_0^\infty \mathrm{d}t\, t^{-n_{\rm NG}}\int_0^\infty \mathrm{d}r\, r^2 6\Big[-2\big(\eta_{\ell_{\ell_1}}\delta'_{\ell_2}\delta'_{\ell_3} + 2\,\text{perm.}\big) + \\ - \big(\zeta_{\ell_1}\beta'_{\ell_2}\beta'_{\ell_3} + 2\,\text{perm.}\big) - \big(\alpha'_{\ell_1}\xi_{\ell_2}\beta'_{\ell_3} + 5\,\text{perm.}\big) + \big(\xi_{\ell_1}\gamma'_{\ell_2}\delta'_{\ell_3} + 5\,\text{perm.}\big) + \\ + \big(\beta'_{\ell_1}\epsilon_{\ell_2}\delta'_{\ell_3} + 5\,\text{perm.}\big) + \big(\beta'_{\ell_1}\gamma'_{\ell_2}\eta_{\ell_3} + 5\,\text{perm.}\big)\Big], \tag{3.29}$$

where we adopted a compact notation, removing the explicit dependence on (r,t).

Chapter 4

Analysis of *Planck* data

The ESA *Planck* mission was the third-generation space mission dedicated to measurements of CMB anisotropies. It came after COBE (Smoot et al. 1992) and WMAP (Bennett et al. 2003), and provided large improvements both in the angular resolution, lowered to 5 arcmin, corresponding to $\ell_{max} \sim 3000$, as well as in sensitivity and frequency coverage (the latter being an essential experimental feature for accurate subtraction of Astrophysical foregrounds). Its instrumentation works in in 9 different bands (30, 44, 70, 100, 143, 217, 353, 545, 857 GHz), allowing for an unprecedented estimation of the sky brightness in these frequencies. *Planck* also uses two different types of detector. Radiometers at low frequencies (30-70 GHz) form the so called *Low Frequency Instrument* (LFI). The *High Frequency Instrument* (HFI) is instead composed of bolometers, covering the range $100 - 857$ GHz. This combination of different detector technologies allows for a tight control of instrumental systematics.

Planck results represent the state-of-the-art of precision cosmology. They are the best CMB measurements to date, covering the full range of relevant scales for primary CMB anisotropies (of course also secondary effects, such as Sunyaev Zel'dovich and CMB lensing, are detected and mapped with high accuracy), from the Hubble radius to the diffusion damping scale. More specifically, the *Planck* data-set succeeds in constraining well 18 peaks in the temperature and polarization angular power spectra, providing a strong confirmation and the tightest available constraints on the standard cosmological model, the so called ΛCDM, and its extensions, such as Inflation. *Planck* simultaneously measures five of the six parameters of ΛCDM with percent precision, providing unprecedented precision and overwhelming amount of information. *Planck* data provide the strongest evidence on the existence of non-baryonic dark matter and establish the

flatness of the universe at the 5×10^{-3} level. They also provide the tightest constraints on neutrino masses and allows for constraints on Dark Energy models found and measurements of the cosmological constant. These incredible results are only the tip of the iceberg of the *Planck* products, arrived now at the third and last release as *Planck* Collaboration. For a compelling review of the legacy of *Planck* we refer the reader to (Planck Collaboration et al. 2018d).

The analysis shown in this chapter will be part of the *Planck* legacy release as a contribution to the paper ""IX.Constraints on primordial non-Gaussianity"". At the moment of writing, the analysis is actually still ongoing, so that the results should be considered preliminary.

4.1 Scale-dependent bispectrum estimation from *Planck* data

In this chapter we present the results of the application of the method described in the previous section to the *Planck* data-set. *Planck* constraints on Primordial non-Gaussianity in previous releases, (Planck Collaboration et al. 2016b, 2014b), improved by more than one order of magnitude any previous measurements of the primordial local, equilateral and orthogonal bispectrum amplitudes. Moreover, a wide range of additional templates was tested, including: oscillatory feature models, resonance models, direction dependent NG, tests for deviations from the Bunch-Davies vacuum and parity-odd bispectra. The local, equilateral and orthogonal amplitudes have been measured independently with four different bispectrum estimation techniques, KSW (Komatsu et al. 2005; Munshi & Heavens 2010), binned (Bucher et al. 2010), and two Modal pipelines (Fergusson et al. 2010, 2012), obtaining fully consistent values.

The forthcoming release will update these measurements, including for the first time the full polarization analysis. In the previous release, E-parity polarization analysis was restricted to multipoles $\ell \geq 40$ since the characterization of systematics at larger scales was still unsatisfactory. As a consequence, the previous results, including polarization data, were considered preliminary.

Of particular relevance for the work of this thesis is that the analysis of scale-dependent templates, presented here, is performed for the first time on *Planck* data. To properly exploit the high level of sensitivity of this data-set, we had to improve our technique, considering corrections for spurious NG signal from secondary anisotropies.

Furthermore, we refined our inference procedure by implementing a new, more conservative prior hypothesis and a different approach. In the next sections we discuss these extensions, while we refer to the previous chapter for the description of the basic methodology.

4.2 Bias from the Integrated Sachs-Wolfe-lensing bispectrum

Secondary anisotropies, generated when the evolution of the gravitational potential deviates from the linear regime, are a possible source of non-Gaussianity. At first order, the radiation transfer functions that link the primordial potential to the CMB field are linear. Under this hypothesis, the Gaussianity of the primordial field is preserved in the convolution and any deviation from it detected in CMB fluctuations, would be a sign of PNG. On the other hand, non-linear contributions give rise to NG in the CMB, independently of the statistical properties of the primordial field.

Non primordial NGs are then a possible source of systematic error in our analysis if they generates a detectable bispectrum signal, with a shape which correlates with the primordial templates under exam. Possible sources of secondary NG are: weak-lensing, Integrated Sachs Wolfe and its 2nd order contributions known as Rees-Sciama effect and the Sunyaev-Zel'dovich effect. At *Planck* resolution, the largest effect comes from the ISW-lensing bispectrum (Serra & Cooray 2008; Mangilli & Verde 2009; Mangilli et al. 2013).

These two phenomena are strongly correlated, since both are sourced by the evolution of the gravitational potential field of matter. The ISW effect arises from large scale fluctuations in the gravitational potential that also source the lensing signal on smaller scales. The net result is a coupling between large and small scales in the CMB temperature bispectrum, *i.e.*, a bispectrum signal peaking on squeezed triangles. In polarization, instead, the E-mode anisotropy sourced by the ISW temperature quadrupole is very low, but there is a strong correlation between the large-scale reionization bump due to scattering and the lensing potential (Cooray & Melchiorri 2006; Lewis et al. 2011). Since lensing and ISW are correlated, this results in an induced E-parity polarization bispectrum signal. The angular bispectrum generated by this effect can be written as:

$$B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \equiv \langle a_{\ell_1 m_1} a_{\ell_1 m_1} a_{\ell_3 m_3} \rangle = \langle a^{\rm P}_{\ell_1 m_1} a^{\rm L}_{\ell_2 m_2} a^{\rm ISW}_{\ell_3 m_3} \rangle + 5 \text{ perm}, \tag{4.1}$$

where the apexes P, L, ISW indicate the contribution that sources the fluctuations, respectively primordial, lensing and ISW. The reduced bispectrum template is (Hu 2000; Lewis et al. 2011):

$$b^{X_1X_2X_3}_{\ell_1\ell_2\ell_3} = C^{X_1\phi}_{\ell_1}\tilde{C}^{X_2X_3}_{\ell_2}g^{X_3}_{\ell_3\ell_2\ell_1} + 5 \text{ perm}, \tag{4.2}$$

where X_1, X_2, X_3 represent the CMB field considered, T or E, $C^{X\phi}_{\ell}$ is the cross power spectrum between the temperature or the E-parity polarization and the lensing potential and the tilde on $\tilde{C}^{X_iX_j}_{\ell}$ indicates that the spectrum is lensed. The factor $g^{X}_{\ell_3\ell_2\ell_1}$ depends on the field considered and is:

$$g^{T}_{\ell_3\ell_2\ell_1} = \frac{1}{2}\left[\ell_1\left(\ell_1+1\right)+\ell_2(\ell_2+1)-\ell_3(\ell_3+1)\right] \tag{4.3}$$

$$g^{E}_{\ell_3\ell_2\ell_1} = \frac{1}{2}\left[\ell_1\left(\ell_1+1\right)+\ell_2(\ell_2+1)-\ell_3(\ell_3+1)\right] \times \begin{pmatrix} \ell_3 & \ell_2 & \ell_1 \\ 2 & 0 & -2 \end{pmatrix}\begin{pmatrix} \ell_3 & \ell_2 & \ell_1 \\ 0 & 0 & 0 \end{pmatrix}^{-1}, \tag{4.4}$$

where the matrices represent the Wigner 3-j symbols, note that the order of the subscripts on the left hand side affects the order of the terms on the right. Moreover, like the primordial shapes, also the ISW-lensing bispectrum obeys the the triangle conditions and selections rules: $m_1+m_2+m_3=0$, $\ell_1+\ell_2+\ell_3 = even$ and $|\ell_i-\ell_j| \leq \ell_k \leq \ell_i+\ell_j$ (since those are imposed just by translation and rotation invariance).

The ISW-lensing template has an oscillatory behavior and is peaked in the squeezed configurations, as pointed out earlier. For this reason, it can produce a significant bias especially in the measurement of local NG, even if the differences in the oscillatory pattern between the two shapes put a limit on this effect. Other shapes can also be affected to a smaller but not negligible extent, if they display a significant enough correlation with the local template (e.g., orthogonal shape) The bias induced in the estimation of primordial $f_{\rm NL}$ is:

$$\Delta^{\rm ISW-L}(f^{\rm P}_{\rm NL}) = \frac{1}{\mathcal{N}}\sum_{\ell_1\ell_2\ell_3}\frac{B^{\rm P}_{\ell_1\ell_2\ell_3}(f_{\rm NL}=1)B^{\rm ISW-L}_{\ell_1\ell_2\ell_3}}{\sigma^2_{\ell_1\ell_2\ell_3}} \tag{4.5}$$

where $B^{\rm ISW-L}$ and $B^{\rm P}$ are respectively the ISW-lensing and the primordial angle averaged bispectrum shapes, σ^2 is the bispectrum variance and $\mathcal{N}$ is the normalization factor. This contribution can be modeled and estimated from Gaussian CMB simulations, provided that all significant effects have been included in the computation of the transfer functions.

4.3 Priors

The issue of the selection of prior distribution dates back to the foundation of Bayesian inference. This choice is especially relevant in case of low S/N, since the likelihood becomes less informative. The analysis of the *Planck* data-set, where the measured $f_{\rm NL}$ central values turn out to have lower statistical significance than in the WMAP analysis, therefore requires a more careful treatment. As an example, for the local shape the significance of the detection decreases with the maximum multipole number considered from 2σ at WMAP resolution ($\ell_{max} \simeq 500$) to $<< 1\sigma$ at *Planck* resolution.

In absence of any a-priori information or assumption, a proper prior must express complete ignorance about the expected value of the parameters. A common solution in these cases, is to resort to uniform distributions. This custom follows Laplace's "principle of insufficient reason", for which, without additional information, equal probabilities shall be assigned to each point of parameter space.

In most cases uniform distributions are indeed an appropriate option, but they present the major drawback to not necessarily be invariant under reparametrization. This implies that a prior that is flat, and so apparently uninformative, under a certain parametrization, can be non flat, and so informative, under another.

The first to provide a satisfactory solution to the problem of determining a parametrization-independent uninformative prior was H. Jeffreys (Jeffreys 1946). Jeffreys' solution was to enforce the invariance under reparametrization with the following general rule for the selection of uninformative priors:

$$\pi_J(\theta) \propto \sqrt{\det(\mathcal{I}(\theta))} \tag{4.6}$$

where $\mathcal{I}(\theta)$ is the Fisher Information Matrix. This approach is not the only way to address the issue: different procedures actually exist and can be applied as appropriate (Jaynes 1968). A detailed presentation of this topic is beyond the the purpose of this work; we address the interested reader to the compelling review in (Kass & Wasserman 1996).

In our estimator, the major complication arises from the choice of the arbitrary pivotal scale k_{piv} . As noted already in (Becker & Huterer 2012), a flat prior on $f_{\rm NL}$, defined at a certain scale, corresponds to a non-flat prior for another scale. In this work the authors checked for alternatives, considering also a uniform prior in $log(f_{\rm NL})$. In conclusion they

decide to quote the results for a flat prior in $f_{\rm NL}$, selecting the pivot scale that minimize the correlation between the parameters. We use the same approach in the analysis of WMAP9 data-set presented in the previous chapter.

This is in general a good choice, and would work properly in case of a significant detection of a bispectrum signal. In absence of a clear detection, however, it presents some caveats. Since the range of scales available is obviously finite, a fit performed at a certain pivot scale will tend to favour particular values of $n_{\rm NG}$. Given that, there is not a perfectly "fair" scale at where to perform the fit. This effect is negligible when the data-set is actually informative about the models tested (*i.e.* when the likelihood is sufficiently peaked), but it can become dominant otherwise. When the goodness-of-fit is somewhat constant throughout the parameter space (*i.e.* the likelihood is flattened), the shape of the posterior can be more affected by the pivot choice than by the data.

Since this is a parametrization effect, the implementation of a Jeffreys prior, that is invariant by construction, seems the natural solution. In the next section we will show how to implement an invariant prior in our estimator therefore making the results pivot independent.

4.3.1 New posterior and likelihood profiling

The implementation of the Jeffreys prior implies a modification in the analytical form of the profiled PDF described in the previous chapter. The likelihood is:

$$\mathcal{L}(n_{\rm NG}, f_{\rm NL}) \propto \exp\left[-\frac{\mathcal{N}(f_{\rm NL}-\hat{f}_{\rm NL})^2}{2}\right] \exp\left(\frac{\hat{f}_{\rm NL}^2\mathcal{N}}{2}\right), \tag{4.7}$$

where $\hat{f}_{\rm NL}$ is the value of the NG amplitude recovered from the KSW estimator and $\mathcal{N}$ is the normalization factor defined in formula (2.45) (see section 3.2 for a detailed description).

The Jeffreys prior is defined as the square root of the determinant of the Fisher information matrix $\mathcal{I}(f_{\rm NL}, n_{\rm NG})$. In the case of separable scale-dependent bispectra, the Fisher matrix is:

$$\mathcal{I}_{\alpha,\beta} \equiv \sum_{\ell_1\leq\ell_2\leq\ell_3} \frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \frac{1}{\sigma^2_{\ell_1\ell_2\ell_3}} \frac{\partial b_{\ell_1\ell_2\ell_3}}{\partial\theta_\alpha} \frac{\partial b_{\ell_1\ell_2\ell_3}}{\partial\theta_\beta}. \tag{4.8}$$

Template	Field	SMICA	SEVEM	Commander	NILC
Local	T	-1.6 ± 5.5	-2.2 ± 5.6	-2.4 ± 5.6	-0.03 ± 5.6
	E	47 ± 28	38 ± 29	31 ± 29	21 ± 28
	T+E	-0.83 ± 5.1	-1.7 ± 5.1	-2.1 ± 5.1	-2.7 ± 5.1
Equilateral	T	14 ± 66	17 ± 66	16 ± 66	5.2 ± 69
	E	170 ± 160	180 ± 170	170 ± 170	59 ± 150
	T+E	-18 ± 47	-9.1 ± 47	-10 ± 47	-32 ± 47
Orthogonal	T	-15 ± 36	24 ± 37	25 ± 37	3.9 ± 39
	E	-210 ± 86	-180 ± 88	-180 ± 88	-120 ± 85
	T+E	-37 ± 23	-15 ± 23	-13 ± 23	-21 ± 23

TABLE 4.1: Results for the NG amplitude parameter $f_{\rm NL}$ of the local,equilateral and orthogonal template obtained with the KSW estimator on the data-set from the SMICA, SEVEM, Commander and NILC component separation pipelines. The results are corrected for the ISW-lensing bias, and the error bars represent the 1σ intervals.

where θ_α and θ_β correspond to $f_{\rm NL}$ or $n_{\rm NG}$, depending on the value of the index, and the matrix is a Wigner-3j symbol.

The determinant of the Fisher matrix, after some maths, can be expressed as:

$$\det\mathcal{I}(f_{\rm NL}, n_{\rm NG}) = f_{\rm NL}^2 \mathcal{I}(f_{\rm NL} = 1, n_{\rm NG}), \tag{4.9}$$

so that the bispectrum Jeffreys prior is:

$$\pi_J(f_{\rm NL}, n_{\rm NG}) \propto \sqrt{f_{\rm NL}^2 \det(\mathcal{I}(f_{\rm NL} = 1, n_{\rm NG}))}. \tag{4.10}$$

We search an expression for the Posterior distribution marginalized over $f_{\rm NL}$. In the previous analysis, we computed the marginalized likelihood assuming a constant prior for $f_{\rm NL}$, obtaining:

$$\mathcal{L}(n_{\rm NG}) \propto \frac{1}{\sqrt{\mathcal{N}}} \exp\left(\frac{\hat{f}_{NL}^2 \mathcal{N}}{2}\right), \tag{4.11}$$

this correspond to the posterior when the prior on $n_{\rm NG}$ is uniform. Assuming instead, the Jeffreys prior for both parameters and integrating over $f_{\rm NL}$, we obtain the marginalized posterior:

$$\mathcal{P}(n_{\rm NG}) \propto \left[\hat{f}_{\rm NL}\sqrt{\frac{2\pi}{\mathcal{N}}} \exp\left(\frac{\hat{f}_{\rm NL}^2 \mathcal{N}}{2}\right) \mathrm{erf}\left(\hat{f}_{\rm NL}\sqrt{\frac{\mathcal{N}}{2}}\right) + \frac{2}{\mathcal{N}}\right] \sqrt{\det(\mathcal{I}(f_{\rm NL} = 1, n_{\rm NG}))}. \tag{4.12}$$

The implementation of this expression in the estimator is straightforward. Compared to the analysis presented before, we only need to numerically compute the Fisher matrix determinant for each value of $n_{\rm NG}$ considered. As we wanted, this expression is

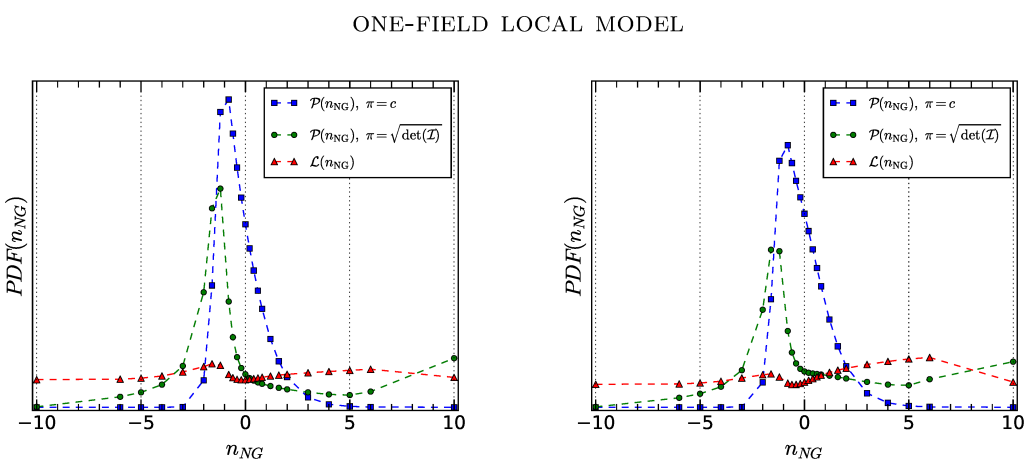

FIGURE 4.1: PDF of the running parameter n_{NG} for the one-field local model. Left panel: SMICA map. Right panel: Commander map. Blue squares: Marginalized posterior assuming constant prior. Green circles: posterior assuming a Jeffreys prior. Red triangles: profiled Likelihood.

independent on the pivot scale used in the computation of the different terms.

A different viable (*i.e.* pivot independent) approach, also starting from (4.7) and taking frequentist approach instead, is likelihood profiling. Instead of marginalizing over $f_{\rm NL}$, the likelihood is sampled along its maximum for every $n_{\rm NG}$ value. For fixed $n_{\rm NG}$, the maximum likelihood $f_{\rm NL}$ is given exactly by the KSW estimator $\hat{f}_{\rm NL}$. Looking at Eq. (4.7), we see that for this condition the first exponential is set to 1 (since $f_{\rm NL} = \hat{f}_{\rm NL}$ in the maximum), and the profiled likelihood reduces to:

$$\mathcal{L}(n_{\rm NG}) \propto \exp\left(\frac{\hat{f}_{\rm NL}^2 \mathcal{N}}{2}\right). \tag{4.13}$$

Notice that also this expression is pivot independent. We will use this expression also to perform a likelihood ratio test between our scale-dependent models and the standard local and equilateral shapes.

In the following, we will show the results from the three different approaches, i.e. from eq (4.11), eq (4.12) and (4.13).

4.4 Results

In this section we will present the results of the analysis obtained with the KSW estimator for the local, equilateral and orthogonal shapes, as well as the first measurements of SD templates performed on *Planck* data-set. In our pipeline we make extensive use of

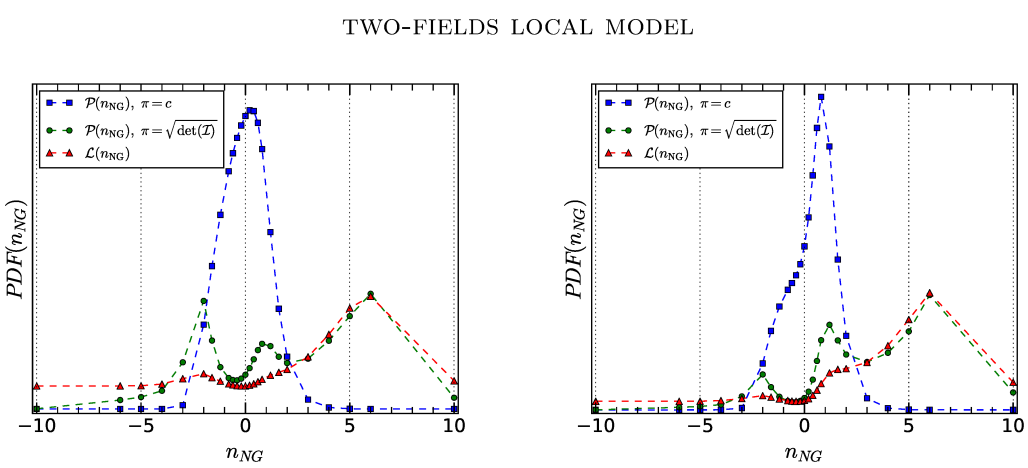

FIGURE 4.2: PDF of the running parameter n_{NG} for the two-fields local model. Left panel: SMICA map. Right panel: Commander map. Blue squares: Marginalized posterior assuming constant prior. Green circles: posterior assuming a Jeffreys prior. Red triangles: profiled Likelihood.

simulations to perform a variety of consistency test, to search for systematics, to correct for partial sky coverage as well as for error bar evaluation. For a detailed description of the simulations and of the data reduction pipeline we refer to (Planck Collaboration et al. 2018c,b). Both the CMB maps and the simulations underwent the SMICA, SEVEM, Commander and NILC component separation pipeline before our analysis (Planck Collaboration et al. 2018e). The comparison of the results obtained from the four foreground cleaning techniques represent an important consistency check. At the moment, the SD analysis has been performed only on the SMICA and Commander temperature maps due to the high computational cost of the estimation pipeline. Since the polarization E-parity maps are not expected to significantly improve the constraints on $n_{\rm NG}$, we decide to include these maps into the analysis only in case of detection from the temperature analysis. We include in the analysis the multipole range from 2 to 2000 and a sky fraction corresponding to the 78% of the sky, masking the residual pixels to reduce the contamination from galactic and extra-galactic sources. The linear term of the KSW estimator, required to correct for the partial sky coverage, is computed from 300 simulations, we tested this is more than sufficient to ensure convergence within the desired accuracy. At this level of resolution and sensitivity, systematic errors induced by ISW-lensing bispectrum become significant. This bias is estimated as the average $f_{\rm NL}$ obtained from the same Gaussian simulations used in the computation of the linear term, and is then subtracted from the final results.

In table 4.1 we present the results of the analysis obtained with the KSW estimator on

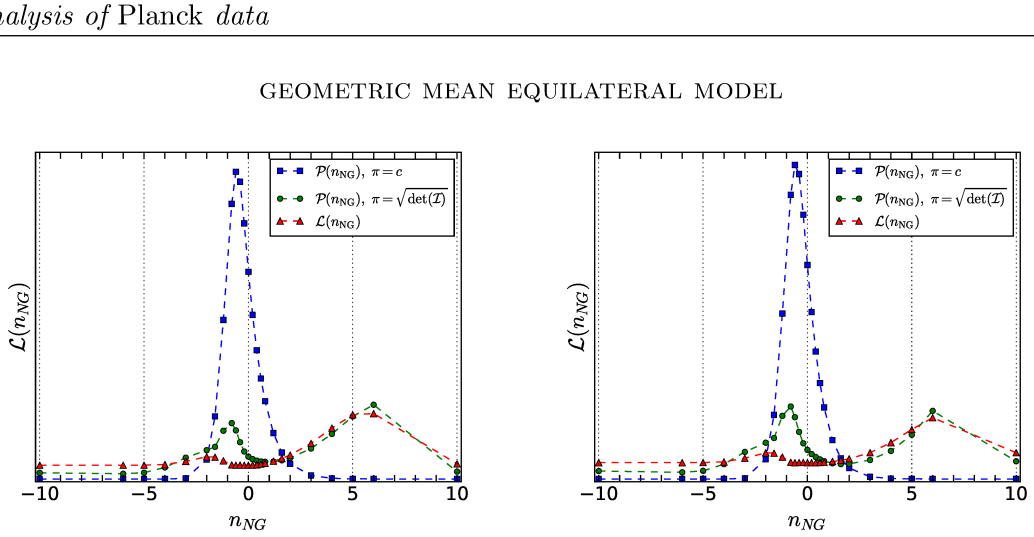

FIGURE 4.3: PDF of the running parameter n_{NG} for the geometric mean equilateral parametrization. Left panel: SMICA map. Right panel: Commander map. Blue squares: Marginalized posterior assuming constant priors. Green circles: posterior assuming Jeffreys priors. Red triangles: profiled likelihood.

the maps from the four foreground cleaning pipelines[1]. These results show in general a good agreement between the different cleaning techniques. Since at the time of writing, final validation tests are still being carried on, we refer to the final release for further comments, including also the comparison with the results from the other estimators.

In figure 4.1, 4.2 and 4.3 we show the results respectively for the one-field local model, the two-fields local model and the geometric mean equilateral model. We show the PDF inferred from all the three methodologies described in the previous section. All curves are normalized to integrate to one. We consider possible values of the running in the interval $n_{\rm NG} = [-10, 10]$. This interval is two order of magnitude wider than the theoretical expectation of the models, that are valid in the regime of mild scale-dependence, *i.e.* $n_{\rm NG} \sim 0.1$. The effects of the prior choice are glaring: while assuming a constant prior (blue squares) we can always identify a peak in the distribution and define proper constraints, this is not the case for the other methods. Implementing an uninformative prior (green circles), the shape of the distribution becomes complex, showing multiple peaks or even diverging on the boundaries, making it impossible to define constraints. A similar behavior appears in the profiled likelihood (red triangles). We used the likelihood also to perform a likelihood ratio test between its maximum value and its value in $n_{\rm NG} = 0$. Notice that, in case of zero running, these models reduce to

[1]The results from the KSW estimator for the local, equilateral and orthogonal shapes has been obtained in collaboration with Alessandro Renzi.

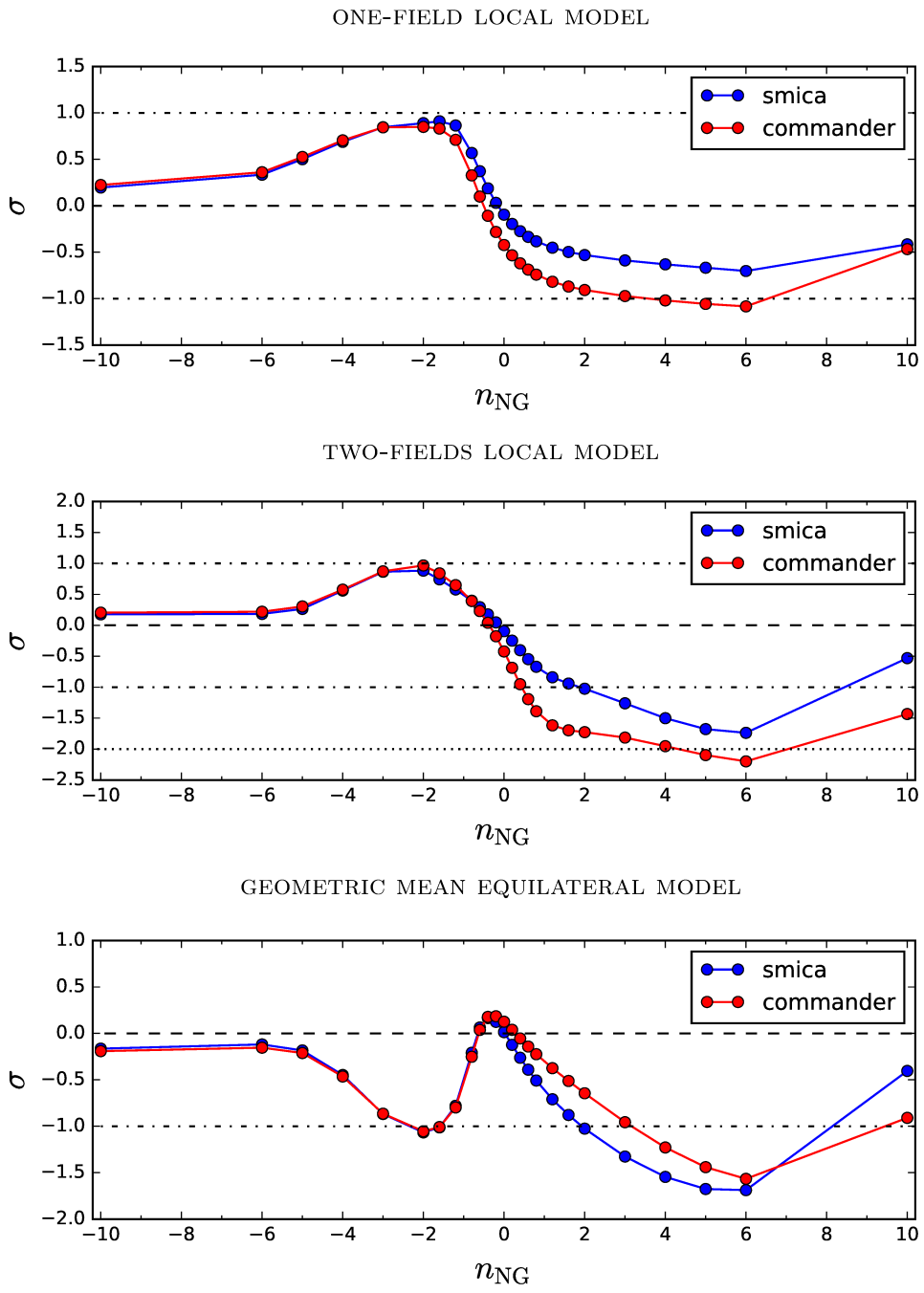

FIGURE 4.4: $f_{\rm NL}$ value, in unit of standard deviation, estimated for different value of the running parameter $n_{\rm NG}$ for the one-field model (upper panel), two-fields model (central panel) and the geometric mean equilateral model (lower panel). Blue line: SMICA, red line: Commander.

the usual local or equilateral shapes. From this test, we do not find evidence in favor of scale-dependent models, assuming an acceptance threshold of $\alpha = 0.01$.

In figure 4.4 we show the amplitude recovered point by point, in units of standard deviation. Also from this analysis, we do not find any evidence of a bispectrum signal for any value of the running. Note that the similar behavior displayed by the two local

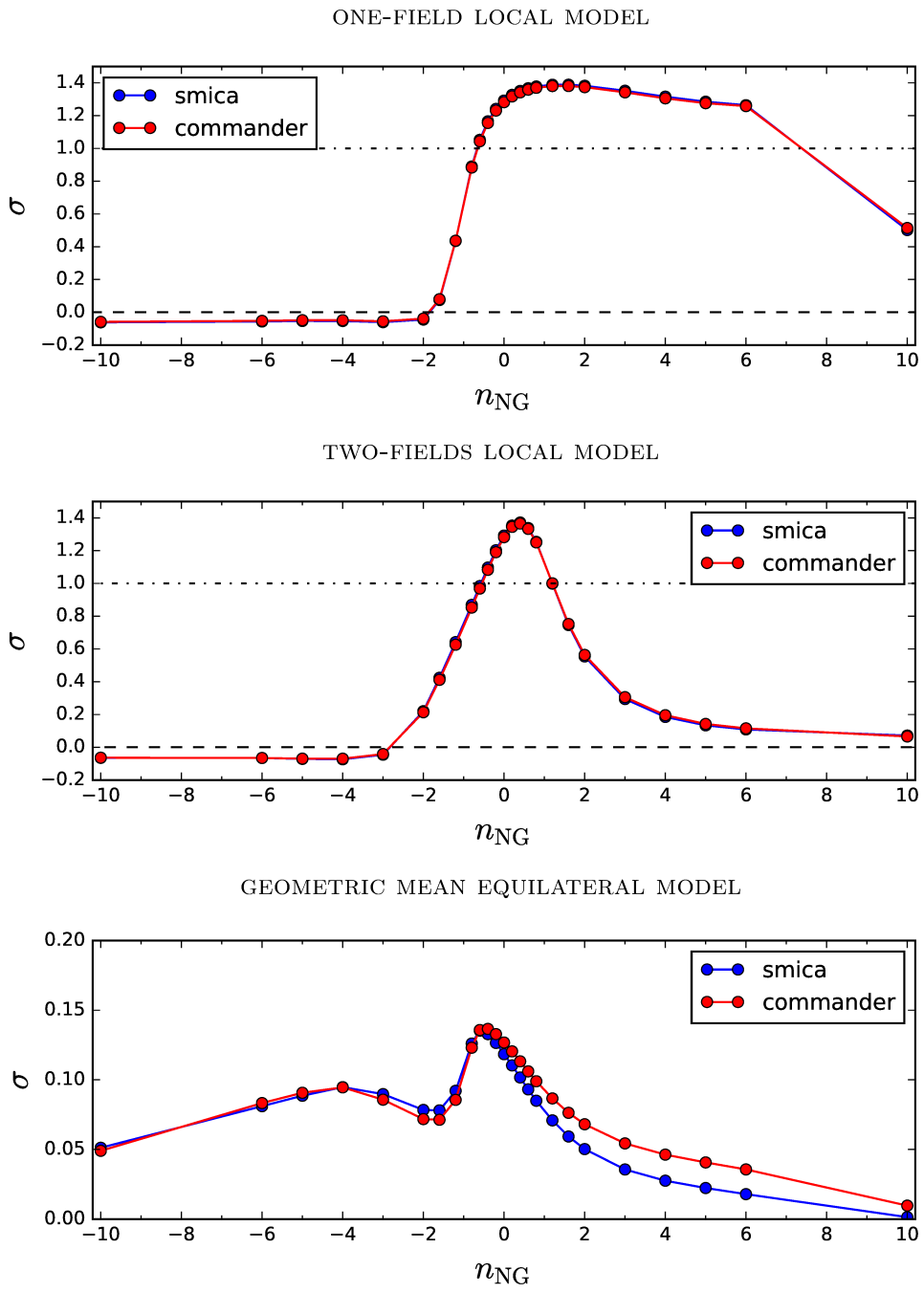

FIGURE 4.5: $f_{\rm NL}$ ISW-lensing bias, in unit of standard deviation, estimated for different value of the running parameter $n_{\rm NG}$ for the one-field model (upper panel), two-fields model (central panel) and the geometric mean equilateral model (lower panel). Blue line: SMICA, red line: Commander.

shapes actually reflects the correlation between the two models.

From this analysis we conclude that scale-dependent bispectrum models cannot be constrained at this level of sensitivity. This is consistent with the reasonable expectation that it is not possible to tightly constrain the scale-dependence of a model without a significant detection of the amplitude. In all tests, we find good agreement between the

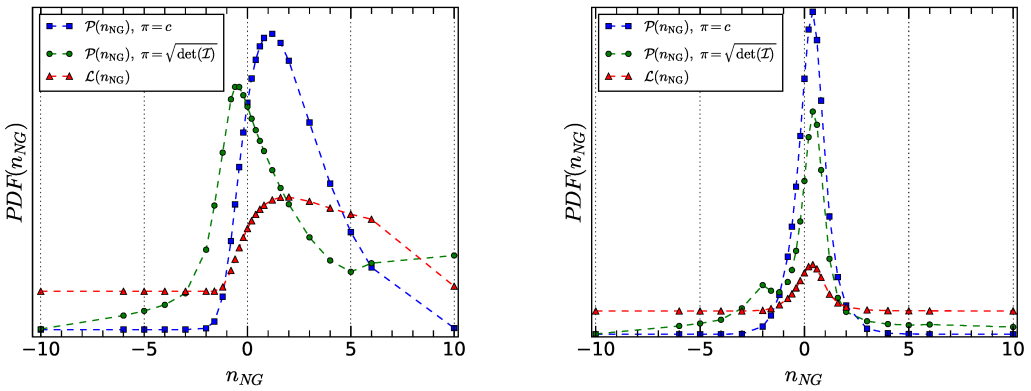

FIGURE 4.6: PDF of the running parameter n_{NG} for the ISW-lensing bias. Left: one-field local model. Right: two-fields local model. Blue squares: Marginalized posterior assuming constant prior. Green circles: posterior assuming a Jeffreys prior. Red triangles: profiled Likelihood.

SMICA and Commander maps. In figure 4.5 we show the results of the same analysis performed on the estimated ISW-lensing NG contamination, again in units of standard deviation. We find a remarkable agreement between the component separation methods. As expected, the equilateral scale-dependent generalization is not affected by this effect. Figure 4.6 shows the distribution obtained from the ISW-lensing signal provided by the SMICA pipeline. We do not shows the Commander counterpart here because they are actually indistinguishable. Also the bias does not show any evidence of scale-dependence.

4.5 Conclusions

We expand the analysis of scale-dependent bispectrum templates presented in the previous chapter. This is the first time that these models have been tested on *Planck* data-set. The adaptation of our estimator to this new data-set required some improvement with respect to the analysis of the lower resolution WMAP data-set. Given the high level of sensitivity of *Planck*, we have now to take into account possible biases, induced by secondary NG, in particular by the ISW-lensing correlation. Moreover we greatly improve our inference, including additional tests resorting to different prior choice as well as different approaches, as frequentist likelihood profiling and model selection via likelihood ratio test.

We notice that, due to the decrease in statistical significance of the estimated primordial NG parameters, the prior choice could now be source of artifacts on the constraints obtained in case of very low significance of the detection of the amplitude $f_{\rm NL}$. This was not an issue in WMAP analysis, since the scale invariant local and equilateral templates, that are nested within our SD models, provided a sufficient level of detection to be unaffected by the prior. To circumvent this risk, we modify our estimator with the implementation of an uninformative Jeffreys prior. We also resort to an alternative frequentist approach in which the likelihood is not marginalized with respect $f_{\rm NL}$ but it is instead profiled along its maximum. Furthermore, we perform an hypothesis test looking at the likelihood ratio between the standard scale invariant shapes and the scale-dependent models.

We run our SD estimation pipeline on the *Planck* temperature maps and simulations, while we also test the standard local, equilateral and orthogonal templates with the KSW estimator (that is a byproduct of our pipeline) on both temperature and E polarization maps. The results for the scale invariant templates show a good agreement between maps provided by the different component separation pipelines, and are in line with the results from the previous release. Since the analysis is still ongoing and the comparison between the different optimal estimators included in the *Planck* NG estimation pipeline is fundamental for the interpretation of the measurements, we refer to the final release for further comments.

From the likelihood ratio test, we do not find evidences in favour of scale-dependent models for any of the templates tested, assuming an acceptance threshold of $\alpha = 0.01$. Coherently, the results from our estimators, both from the frequentist one based on the likelihood profiling and from the Bayesian one based on the likelihood marginalization with a Jeffreys prior, state that at the current level of detection it is not meaningful to put constraints on the additional scale-dependence parameter n_{NG}. The standard technique resorting to a flat prior instead provides very weak constraints that, however, seems to be an artifact induced by the choice of the pivot scales where the fit is performed.

From this analysis we conclude that, since *Planck* constraining power on direct bispectrum estimation almost saturates the theoretical limit for CMB observation, future investigation of scale-dependent bispectrum templates shall rely on different data-set, as LSS, CMB spectral distortion and radio surveys (Emami et al. 2015; Cooray 2005; Biagetti et al. 2013; Ravenni et al. 2017; Khatri & Sunyaev 2015; Raccanelli et al. 2015).

Chapter 5

Component separation in CMB surveys

The Cosmic Microwave Background has proven to be the best benchmark for precision measurements in Cosmology. The simplicity of the underlying physics allows testing fundamental theories and measuring cosmological parameters with unique precision. Given the exquisite sensitivity of current and future surveys, systematic effects are becoming more and more the dominant source of error in CMB data, and need to be understood and modeled with unprecedented accuracy.

One of the major source of systematic contamination in the data, besides instrumental effects, is the astrophysical foreground (Delabrouille & Cardoso 2009). Instrumental effects can be carefully modelled to the necessary accuracy, *e.g.* resorting to laboratory measurements. On the other hand, the required sensitivity can be several order of magnitude lower than the unavoidable astrophysical signals. Foregrounds are nowadays the principal concern in the analysis of current data-sets and in the development of future surveys. In this chapter we briefly describe the astrophysical phenomena contributing to the total sky brightness and the techniques developed to disentagle these emission from the cosmological signal.

Foregrounds as a whole are very complicated to threat. The *Planck* team reported that up to 20 physical components are required to proper modeling the microwave sky (Planck Collaboration et al. 2016a). Moreover, these processes exhibit an highly complex behavior, their emission laws may vary across the sky as well as between different scales and frequencies. The major contribution comes from diffuse galactic emission, but also the solar system and extra-galactic sources play an important role.

All these emissions appear mixed together in the total microwave brightness of the sky. The problem to identify and isolate a particular signal is known as component separation. With the advent of large CMB surveys, component separation becomes a very active field of study in observational cosmology. A large number of methodologies and algorithms have been developed in the attempt to improve our understanding and to consolidate our measurements. Given the heterogeneity of the sources, the importance of component separation transcends cosmology, as it is fundamental to exploit the huge potential held in the sky maps from modern surveys to provide astrophysical information (Planck Collaboration et al. 2016c,d; Ade et al. 2014a,b; Planck Collaboration et al. 2018a; Génova-Santos et al. 2017, 2015; Hilton et al. 2013; Hasselfield et al. 2013; Bierman et al. 2011). Thanks to the careful classification of the various emissions, microwave experiments such as *Planck* and WMAP have provided valuable information on the Galaxy and the Solar System in addition to cosmology and CMB physics.

We review the main emission mechanisms in section 5.1. Some of the most used component separation techniques are described in section 5.2. In section 5.3 we outline the component separation pipeline used in the analysis of *Planck* data.

5.1 Sky components

Diffuse galactic emission dominates the sky brightness in the microwave part of the spectrum. It originates from the interstellar medium (ISM), the diffuse matter filling the voids between stars. The ISM is composed of atomic and molecular gas in a partially ionized state and a small amount of dust. Energetic radiation from young stars stimulates the emission by heating and exciting the ISM, while Supernovae explosions enrich it of high energy particles and cosmic rays. It is concentrated on the galactic plane, where it clusters in vast clouds and nebulae, and its density decrease with the galactic latitude. As a rule of thumb, the integrated emission of the diffuse matter decreases as the galactic latitude cosecant.

The main contribution to diffuse emission in the frequency range between $1-1000$ GHz derives from interstellar dust and free electrons interacting with electric and magnetic fields. The proper modeling of the microwave sky requires detailed knowledge of the physics behind diffuse galactic emission. These information are as precious as the knowledge of the CMB itself for the development of separation technique. The most

distinctive features of each component separation technique often lie in how much of this prior knowledge is exploited and in the way it is included in different approaches.

5.1.1 Dust

Galactic dust is a fraction of the interstellar medium consisting of small particles of various materials, it shows a very heterogeneous chemistry. The principal components are sub-micrometer size silicate and carbonaceous grains, with different shapes, size and physical state, as well as large molecules and organic compound as the polycyclic aromatic hydrocarbons.

Dust represent $\sim 1\%$ of the interstellar medium in our Galaxy, but it play a leading role in the interaction of this diffuse medium with starlight, absorbing high-frequency radiation and re-emitting in the infrared as a grey body. This process is so efficient that it accounts for some 20% of the bolometric luminosity of the Milky Way (Whittet 1992).

It is especially concentrated in the galactic disk, forming structures on a wide range of scales ranging from 10^{-4} pc to 10^{3} pc.

Dust is a longstanding issue in astronomical observations. Given its properties, it places important limitations on observations at any wavelength. In visible light its main effect is the extinction of starlight, so that dusty structures appears as dark patches with no stars. Some of these "holes" in the Galaxy can be noticed with unaided eye, as in the case of the "dark constellations" identified by the Inca and the Aboriginal civilizations.

The identification and the correction for dust induced biases characterized the development of modern astronomy. Evidences of the effects of dust extinction on the determination of stars distances has been found as early as the nineteenth century (Struve 1847). It is after the investigation of Trumpler, during the third decade of the last century (Trumpler 1930a,c,b), that corrections for absorption and reddening [1] have been routinely implemented in astronomical data reduction.

In millimeter and infrared astronomy, on the contrary, the contaminant effects come from the thermal emission and the recently explained rotational emission, once known as Anomalous Microwave Emission (AME) (Planck Collaboration et al. 2014a). Thermal dust emission is particularly relevant for CMB surveys above 100 Ghz, where it is the dominant radiation mechanism (Planck Collaboration et al. 2016a).

[1] The term reddening refers to the distortion of stars spectra due to the selective extinction from dust. Since the absorption efficiency is higher at shorter wavelengths, stars appears redder proportionally to the amount of dust along the line of sight.

Analysis of *Planck* data-set shows that a modified Black Body provides a good description for the dust spectra below 857 GHz, while beyond this limits the underlying physics become too complex. Thermal dust emission is so parametrized by its temperature and emissivity index, defined pixel by pixel to account for variations in dust population and environment.

Moreover, aspherical grains align with galactic magnetic fields. As a consequence, statistical asymmetry arises in their orientation, resulting in a significant degree of linear polarization, proportionally to the strength of magnetic fields. Dust induced E-mode and, especially, B-mode signals represent one the greatest issues for CMB polarization measurements. Spinning dust can emit dipole radiation if grains have a non-zero electric dipole. This results in a continuous emission in the range from ~ 1 Ghz to ~ 100 Ghz with a peak around $20 - 40$ Ghz. It has been first referred as AME because the physical nature was still unclear, due to the significant correlation with synchrotron and bremsstrahlung. These degeneracies are still a limitation for a proper modelling of spinning dust radiation. Forthcoming low frequency surveys are expected to provide valuable information to overcome this issue (Génova-Santos et al. 2015).

Finally, the redshifted emission of dust in distant galaxies is responsible for the Cosmic Infrared Background (CIB) radiation. Since the extra-galactic and the galactic component share similar emission laws, they are usually tracked together. Thermal dust templates often contains both contributions that are subsequently separated.

5.1.2 Emission from charged particles

Charged particles, especially free electrons, represent an important source of spurious signal. Their effects are detectable both on galactic and cosmological scales. In our Galaxy, the contribution of free electrons to the total microwave brightness comes from the synchrotron and the bremsstrahlung (free-free) diffuse radiation. At cosmological scales the interaction of CMB photons with free electrons in hot intra-cluster medium via inverse Compton scattering account for the Sunyaev–Zel'dovich effect.

Bremsstrahlung radiation arises in an ionized medium as a consequence of the acceleration of free electrons in the Coulomb field of ions. It is the dominant emission mechanism in the HII regions, where the hydrogen is partially ionized by the energetic photons from young and hot stars. Moreover, due to the interaction with cosmic rays, a small fraction ($\sim 5\%$) of the hydrogen in the ISM is permanently in ionized state and it

contributes to the diffuse bremsstrahlung. The emission law of thermal bremsstrahlung, *i.e.* assuming equilibrium between electrons and ions, can be modeled from the velocity distribution of electrons, as a function of the temperature and density. The resulting spectrum presents two regimes: a power law with spectral index ~ 2 at frequencies above $\sim$ 1Ghz, where the medium is optically thin, while for lower frequency the ISM is optically thick and the brightness temperature becomes constant and equal to the electron temperature.

Supernovae explosion enrich the ISM with relativistic electrons and cosmic rays. These high energetic electrons are forced in helical trajectories by the Galactic magnetic field and being accelerated they radiate. For relativistic particles this effect is known as synchrotron emission[2]. In the non-relativistic case (cyclotron), particles spiraling into a magnetic field B radiate at a frequency equal to the gyration frequency of the field ω_B. This frequency depend only on the intensity of B and not on the particle energy. Relativistic particles instead emit in a cone of angular amplitude $\propto \gamma^{-1}$, where γ is the Lorentz factor, perpendicular to the acceleration, this is called beaming effect. An observer will thus receive radiation only when this narrow beam points toward him, detecting a series of pulses of duration $\propto (\omega_B \gamma^2)^{-1}$. As a consequence the spectrum of synchrotron extend from the gyration frequency ω_B to $\sim \omega_B \gamma^2$, following the energy distribution of electrons. For example, assuming a power law distribution $n_e E \propto E^n$, the spectrum becomes $I(\omega) \propto \omega^{-\frac{n-1}{2}}$.

Synchrotron emission is dominant in the sky maps at frequency below 100GHz. It can have an high polarization fraction, with a maximum theoretical value $\sim 70\%$. Realistic values are however much lower, $\sim 10\%$ for the diffuse emission and $\sim 30-50\%$ in particular structures (Ade et al. 2016). Furthermore, a number of extra-galactic synchrotron emitters appear as point sources and should be masked.

Finally, the most relevant extra-galactic contribution comes from the SZ effect. It derives from the interaction on CMB photons with the hot ionized gas in galaxy clusters and dark matter halos. The internal motion of this gas provides a distortion in the CMB spectrum as there is a transfer of energy between the hot gas and the low energy CMB photon. In addition there is also a little kinetic contribution from the peculiar velocities of gas clouds. The Sunyev-Zel'dovich effect dominates the power spectrum of CMB secondary anisotropies on angular scale $\sim 1'$.

[2]Synchrotron has been discovered in 1946 in the synchrotron of the General Electric Company as an unexpected blue emission.

5.2 Component Separation Methods

Having described the foreground physics, we now move on to introduce some of the component separation methods used in modern CMB experiments. The basic assumption of the vast majority of component separation techniques is the so called linear mixture model. It describes the overall observed sky brightness at different frequencies in terms of a linear mixture of several diffuse emissions. Assuming to have observations in $N_{\rm chan}$ different frequency bands, with $N_{\rm pix}$ data points for each channel, of a mixture of $N_{\rm comp}$ components, in matrix notation it states:

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}, \tag{5.1}$$

where $\mathbf{y}$ is an $N_{\rm chan} \times N_{\rm pix}$ matrix representing the observed data, $\mathbf{s}$ is an $N_{\rm comp} \times N_{\rm pix}$ matrix containing the template of emission of the components and $\mathbf{A}$ is the $N_{\rm chan} \times N_{\rm comp}$ mixing matrix that weights the contributions of each component at each frequency, *i.e.* it defines the emission laws.

The aim of component separation can be to isolate the contribution of a particular emission of interest, *e.g.* the CMB, as well as to build the full sky model assigning a spatial template and an emission spectrum to each source. Starting from equation (5.1), different techniques can be developed according to the a priori hypotheses, to the parametrization choices and to the bases used to represent the data. In the context of this thesis, we are particularly interested in methods making minimal assumptions on the characteristics of the components, the so called "Blind methods". These techniques try to estimate the mixing matrix by exploiting statistical properties such as either the independence of the components, like for the ICA method described in the following, or other general assumptions. Blind separation has the obvious advantage of working even in presence of unknown contamination, provided the initial ansatz gives a reasonably fair description of the real scenario under study. Clearly, additional information is needed to classify the different sources. Reconstructing all components in the data requires essentially the inversion of equation (5.1). This, in turn, clearly requires as many observation channels as the total number of components. This is the motivation behind multi-channels CMB surveys. A general rules states that we need at least the same number of channels than the components we aim to identify. A description of the main component separation methods used in CMB analysis is contained in the following

sections.

5.2.1 Internal Linear Combination

Internal Linear Combination, ILC, is a blind method providing estimates of a single component. The basic assumption is that the observations are calibrated with respect to the component of interest, that is therefore constant among the different channels. This is equivalent to assuming perfect knowledge of the emission law of the target component. Given that the CMB is the best example of black body spectrum in nature, this method is very well suited for CMB estimation.

ILC maps are weighted averages of the different channels. Assuming model (5.1) and denoting the target component with the vector $\mathbf{s}_0$ of length $N_{\rm pix}$, the general ILC solution is in the form:

$$\hat{\mathbf{s}}_0 = \boldsymbol{\omega}^T \mathbf{y}, \tag{5.2}$$

where $\boldsymbol{\omega}$ is the weights vector of length $N_{\rm chan}$ subject to the constraint $||\boldsymbol{\omega}||_1 = 1$. The weights are defined minimizing the variance of the output map, that is:

$$\sigma^2_{ILC} = \boldsymbol{\omega}^T \langle \mathbf{y}\mathbf{y}^{\rm T} \rangle \boldsymbol{\omega} = \boldsymbol{\omega}^T \mathbf{C} \boldsymbol{\omega}, \tag{5.3}$$

where $\mathbf{C}$ is the estimated covariance between the channels. The solution to the minimization of (5.3) is obtained resorting to the method of Lagrange multiplier, under the constraint provided by the weights normalization. Therefore the general ILC solution states:

$$\boldsymbol{\omega} = \frac{\mathbf{a}^{\rm T}\mathbf{C}^{-1}}{\mathbf{a}^{\rm T}\mathbf{C}^{-1}\mathbf{a}}, \tag{5.4}$$

here $\mathbf{a}$ is the vector of calibration coefficient equalizing the response of each channel to the CMB.

ILC is a very versatile method and can be further refined compared to the simple solution given in (5.4). The weights can be allowed to vary between different regions of the sky or different angular scales, when working in harmonic space, or both if the signal is represented in a more complex basis such as wavelets. This expedient improves the results since it allows to take trace of the great variability of foreground emission along the sky. Once the data are represented an appropriate basis, this implementation is straightforward, it consists in performing ILC independently on different sub samples,

assuming you divide the sky in uncorrelated regions, in pixel space, or you expand the signal in uncorrelated modes, in other representation domains.

If the target component is uncorrelated to the other emissions, it is proven that ILC is equivalent to the inversion of the mixing matrix. On the contrary, in presence of significant correlations the ILC solution is biased. So that, this method is unfitted to recover galactic components, that are all strongly correlated to each other. Particular care is required also in handling small data sets, due to the possible emergence of empirical correlation.

5.2.2 Independent Components Analysis

Independent Component Analysis (ICA) is a class of component separation techniques exploiting statistical independence to identify the components. Since no assumptions are made about the physical properties of the various emissions, ICA is classified among blind methods. In practice, the aim of ICA is the inversion of equation (5.1), *i.e.* to find $\mathbf{W} = \mathbf{A}^{-1}$, under the constraints that the different sources of emissions are mutually independent. ICA is a widely used technique and it is proven to work very well, but there is a caveat to take into account when targets are physically significant sources: its results are defined up to re-scaling and permutations of components. The reason is simply that the correlation between the components is invariant under these operations. The only way to overcome these limitations is to resort to additional information, being however aware that this operation will compromise the "blindness" of the method.

The core of any ICA method is the criterion used to define independence. The obvious choice will be to enforce the decorrelation of the components templates. Denoting as $\mathbf{x} = \mathbf{W}\mathbf{y}$ the ICA estimate of $\mathbf{s}$ in equation (5.1), this condition is equivalent to setting to zero the diagonal elements of:

$$\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle = \mathbf{W} \langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \rangle \mathbf{W}^{\mathrm{T}} = \mathbf{W}\mathbf{C}\mathbf{W}^{\mathrm{T}}, \tag{5.5}$$

where, again , $\mathbf{C}$ represents the estimated covariance between the channels. Unfortunately, due to symmetries, this condition alone cannot provide the correct number of constraints to define $\mathbf{W}$ (*i.e.* N^2_{comp}). Modern ICA techniques solve this problem resorting to more complex criteria, like non-linear correlator and localized correlations.

5.2.3 Template Fitting

Template fitting provides estimates of the amplitude of a component from the fit of a known template to the data of interest. With templates $\mathbf{T}$ of the expected foregrounds sources in hand, the contribution on a given channel can be obtained via linear regression. The results are then subtracted from data to remove the spurious signal.

In the linear mixture model (5.1), the distribution of the components over the data is stored in the matrix $\mathbf{s}$. The templates used should then reproduce the elements of $\mathbf{s}$ other than the CMB, or a linear combination thereof. Assuming to have the exact templates, the linear fit would provide the entries of $\mathbf{A}$ corresponding to the given source and channel.

The result for a collection of known templates $\mathbf{T}$ fitted to a map $\mathbf{y}$ is the standard linear regression solution. Calling $\mathbf{c}$ the vectors of estimated amplitudes we have:

$$\mathbf{c} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}\right)^{-1}\left(\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}\right), \tag{5.6}$$

where C is the $N_{\mathrm{pix}} \times N_{\mathrm{pix}}$ covariance matrix of the map, that depend on the noise and on the CMB. The estimation and the inversion of $\mathbf{C}$ is a major limitations in template fitting since, given the high number of data points collected by modern surveys, it is very large and so computationally expensive.

The choice of templates is, obviously, the a crucial part of any template fitting technique. A possibility is that of resorting on external templates from previous experiment or from theoretical speculations. This approach requires a lot of a-priori information on the emissions of interest, which can be unavailable or unreliable. Relying on external data-sets also runs into the issue of having to deal with additional systematic effects, cross-calibration problems and so on. For this reason, most of the modern CMB surveys have been designed with a number of channels at foregrounds dominated frequencies, allowing us to track spurious contaminant components without having to rely on external information. These templates, obtained directly from the data, are called internal templates. The most straightforward approach would be just to use these foreground dominated maps as template, for example a high frequency channel as dust template and a low frequency as synchrotron. However, these maps contain also the CMB contributions, which would be removed together with the contaminants, so that a correction factor must be introduced in formula (5.6). Another widely used solution is thus to

build linear combinations of these maps, so that the constant component (the CMB, providing the observation are calibrated to black body) vanish from the fitted template.

5.3 The *Planck* Pipeline

The large number of method developed in this field reflects the high complexity of the component separation problem. The difficulty of modeling foreground emission in future, higher sensitivity, CMB surveys, aimed at detecting the faint primordial B-mode signal, encourages this heterogeneous approach. Alternative methods might in fact perform better or worse in correspondence to a different physical and statistical behaviour of the polarized astrophysical emission, and the latter is hard to predict at this stage. Each method exploits different characteristics to identify contaminants and each provides different results. This make difficult comparing them and almost impossible identifying a single technique superior to the others for all scientific targets. Also looking at their performance on data, the problem of uniquely identify the "correct" result still remain.

In light of these points, the component separation effort by the *Planck* collaboration was indeed characterized by the use and cross-validation of a variety of methods, which we briefly review here. Well before the first official data release, the *Planck* Collaboration organized a so called "component separation challenge" among a large number of different algorithms representing the state of the art for each different methodology. These techniques were blindly tested on a series of ad-hoc simulations of *Planck* observations. The results of the challenge are summarized in (Leach et al. 2008): as expected, no method outperformed the others. The official pipeline developed during these preliminary test and fixed after the first release consists of a suite of four different algorithm, involving different techniques and aiming at different targets.

NILC, (Needlet Internal Linear Combination) (Basak & Delabrouille 2012, 2013)is an ILC algorithm working in needlet space. It exploits the properties of needlets to be localized both in real and harmonic space. This allows calibrating the weights to follow the irregularities in the distribution of contaminant emissions . As any ILC techniques it only provides estimates for a single component, the CMB. The advantages of using the needlet frame will be discussed in detail in the following chapters.

Commander (Eriksen et al. 2004; Eriksen et al. 2008) follows a totally different approach, performing a Bayesian fit of a parametric model of observations. It provides a full modellization of the sky signal, estimating all the significant sources contributing

to the total microwave sky brightness. The model parameters are estimated via Gibbs sampling on a real space map at low resolution (although a new version working in harmonic space, dubbed Commander2, was used in the last *Planck* release (Planck Collaboration et al. 2018e)). Its outputs are the most complete among the model tested in the challenge, it provides maps of all known components, therefore it is computationally very expensive.

SEVEM (Spectral Estimation Via Expectation Maximization) (Fernández-Cobos et al. 2012) is a multi-step algorithm that models the CMB signal only. In the first step it uses internal templates, obtained from the difference of adjacent channels, to clean the three central bands of *Planck*, *i.e.* the least contaminated by astrophysical emission. It then obtains estimates of the CMB power spectrum from these channel from a Expectation Maximization algorithm. Finally, the spectrum is used to Wiener filter the pre-cleaned maps, obtaining the final CMB map..

SMICA (Spectral matching Independent Component Analysis) (Cardoso et al. 2008) is an ICA technique working in harmonic space based on the evaluation of auto and cross-power spectra of the survey's channels. The condition of decorrelation of the CMB from other emission is enforced using localized correlations, *i.e.* looking at the binned spectral covariance matrix. A parametric model of the of the spectral covariance $C_\ell(\theta)$[3], is adjusted to the data minimizing, for each multipoles bin, the cost function:

$$\sum_{\ell_{min}}^{\ell_{max}} (2\ell+1) \left[\mathrm{Tr}(\hat{C}_\ell C_\ell(\theta)^{-1} + \log\det C_\ell(\theta) \right], \tag{5.7}$$

where $[\ell_{min}, \ell_{max}]$ are the multipole bin boundaries and $\hat{C}_\ell$ is the standard pseudo-C_ℓ estimator of power spectra of the channels. It gives estimates of the power spectra of the CMB and of the foregrounds included in the model $C_\ell(\theta)$, for a maximum of N_{chan} components considered.

[3]Notice that here C_ℓ contains all the harmonic correlations between the channels for the multipole ℓ, not the CMB power spectrum as usual

Chapter 6

Needlet Regression

Wavelets are functions with the property to be localized both in space (or time) and frequency. They are widely used in signal processing due to their ability to identify localized oscillatory signals in the data stream. The representation of signals as wavelet series is a valid alternative to the standard Fourier analysis, more suitable to investigate non-periodic features.

Thanks to their versatility, spherical wavelets are powerful tools for CMB data analysis. The reason is easy to understand: on one side, harmonic analysis is the ideal framework to test CMB models, that are almost uniquely developed and computed in frequency space; on the other side, many practical limitations, like incomplete sky coverage, foreground emission or missing pixels, makes the exact computation of the harmonic coefficients a challenging task. The tight space localization of wavelets is ideal to overcome these issues and at the same time to investigate the spectral behavior of CMB fluctuations. Some examples of the applications of wavelet in CMB analysis can be found in (Vielva et al. 2004; Hansen et al. 2006; Cabella et al. 2004; Moudden et al. 2005; McEwen et al. 2007).

Wavelets have the important property to give a sparse representation of smooth functions. Sparse means that the majority of the coefficients of the wavelet expansion of a given function are zero. In other words, the information content of a coherent signal, when projected on a wavelet basis, is concentrated in few wavelet coefficients. This property is very useful in signal representation and denoising.

The choice of the wavelet basis is not unique, the basis functions should be modeled in the optimal shape to better highlight the features of interest. In the context of signal analysis on the sphere, a number of possible bases exist in the literature (Antoine et al.

2010; Sanz et al. 2006; McEwen et al. 2007, 2006). In this work we resort to spherical needlets, developed as a functional analysis tool by (Narcowich et al. 2006) and applied to CMB analysis in (Baldi et al. 2006, 2007; Pietrobon et al. 2006).
Needlets have a number of mathematical features that make them ideal for the application in CMB analysis. In this chapter we will review the main characteristics and show the construction of the spherical needlet basis. In the following section, we will then discuss in detail the notion of sparseness and wavelet thresholding techniques.

6.1 Spherical Needlets

Spherical wavelets are usually constructed by relying on a local flat sky approximation. This means that the base function is defined on a flat tangent plane and then implemented on the sphere. The needlet basis, instead, is defined directly in harmonic space, with respect to spherical harmonics. As we will show in the following, this is a great advantage for the exact computation of the needlets coefficients. Furthermore, as their name suggest, needlets enjoy remarkable localization properties. In real space, by construction, they are quasi-exponentially localized, this means that the basis function decreases faster that any polynomial around its center. Their support in harmonic space instead is compact so that each layer is exactly limited to a log-constant multipole interval. Moreover, we will show that random needlet coefficients have noticeable uncorrelation properties, both in space and frequency space. Finally, as is the case for other wavelet systems, needlets are not a orthonormal basis but a tight frame, so that the basis contains redundant elements; this is a key element to exploit sparseness. In the following description of the construction of needlets we will follow extensively (Baldi et al. 2009; Marinucci & Peccati 2011; Marinucci et al. 2008).
Needlets are constructed in spherical harmonics space, so that we start recalling of some useful notions about spherical harmonics representations. The projector of a squared integrable function on the sphere $f \in L^2(S^2)$ into the space of spherical harmonics $\mathcal{H}_\ell$ is the kernel operator:

$$P_{\mathcal{H}_\ell} f(x) = \int_{S^2} \mathrm{d}x \; f(x) K_\ell(\langle x, y \rangle), \tag{6.1}$$

where the kernel K_ℓ is the Legendre polynomial of degree ℓ, we recall the relation with the spherical harmonics:

$$K_\ell(\langle x, y\rangle) = \frac{2\ell+1}{4\pi} P_\ell(\langle x, y\rangle) = \sum_m Y_\ell^m(x)\overline{Y}_\ell^m(y), \tag{6.2}$$

known as the spherical harmonics addition theorem. This kernel operator satisfies the reproducing property, the proof is immediate recalling the spherical harmonics orthogonality relation:

$$\int_{S^2} \mathrm{d}y\ K_\ell(\langle x, y\rangle) K_{\ell'}(\langle y, z\rangle) = \sum_{mm'} Y_\ell^m(x)\overline{Y}_{\ell'}^{m'}(z) \int_{S^2} \mathrm{d}y\ Y_{\ell'}^{m'}(y)\overline{Y}_\ell^m(y) = K_\ell(\langle x, z\rangle)\delta_\ell^{\ell'}. \tag{6.3}$$

The projector (6.2) represents the first block for the construction of the basis. The importance of this relation can be understood recalling that, under certain conditions, integrals like (6.2) can be exactly computed as a weighted summation. Calling $\mathcal{P}_\ell$ the restriction to the sphere S^2 of the polynomials $p(x)$ of degree $< \ell$, this space can be decomposed as the direct sum of the spaces of spherical harmonics of degree $\leq \ell$, formally $\mathcal{P}_\ell = \bigoplus_0^\ell \mathcal{H}_\ell$. It is a known results that there exists a finite set of points $\mathcal{X}_j \in S^2$ and weights $\lambda_{jk} \in \mathbb{R}^+$ so that:

$$\int_{S^2} \mathrm{d}x\ p(x) = \sum_{\xi_{jk} \in \mathcal{X}_j} \lambda_{jk} p(\xi_{jk}) \tag{6.4}$$

gives the exact result if $p \in \mathcal{P}_\ell$. Note that we have introduced here the indices j, k, the meaning and importance of which will become clear in the following.

The importance of this standard result is that integrals involving spherical harmonics can be exactly computed numerically (or approximated with a known tolerance). A number of numerical libraries providing the necessary cubature points and weights exists . Given that needlets will depend on these function by construction, they will inherit these properties.

The last ingredient is the window function $b(\ell, j)$ that sets the harmonic support for each needlet layer j. This function must satisfy three properties:

1. compact support: $b(\ell, j) > 0$ if $\ell_{min,j} \leq \ell \leq \ell_{max,j}$ and $b(\ell, j) = 0$ otherwise. This ensure that each needlet layers represents a fixed range of scales. Moreover, each layer j will have equal support in $\log(\ell)$.

2. partition to unity, so that $\sum_j b^2(\ell, j) = 1$.

3. smoothness, *i.e.* $b(\ell, j)$ is infinitely differentiable.

We follow the procedure presented in (Baldi et al. 2009) for the derivation of the proper $b(\ell, j)$ satisfying these conditions. First we consider the function:

$$\phi_1(t) = \begin{cases} e^{-\frac{1}{1-t^2}} & \text{if } -1 \leq t \leq 1 \\ 0 & \text{if } t < -1 \bigcup t > 1, \end{cases} \tag{6.5}$$

this function satisfies condition 1 and 3 since it is C^∞ and its support is compact over the interval $[-1, 1]$. We use then this function to define:

$$\phi_2(x) = \frac{\int_{-1}^{x} \mathrm{d}t\ \phi_1(t)}{\int_{-1}^{1} \mathrm{d}t\ \phi_1(t)}, \tag{6.6}$$

that is still infinitely differentiable; it is also a non decreasing function with $\phi_2(-1) = 0$ and $\phi_2(1) = 1$. For the next step we have first to introduce the needlet bandwidth parameter B that governs the multipole coverage of each layer. We then implement a change of variable in the last definition to obtain:

$$\phi_3(t) = \begin{cases} 1 & \text{if } \quad 0 \leq t < \frac{1}{B} \\ \phi_2\left(1 - \frac{2B}{B-1}(t - \frac{1}{b})\right) & \text{if } \quad \frac{1}{B} \leq t \leq 1 \\ 0 & \text{if } \quad t > 1. \end{cases} \tag{6.7}$$

This function is constant for t between 0 and B^{-1} and then decreases to zero between B^{-1} and 1. This can be easily checked noticing that its argument is equal to 1 if $t = B^{-1}$ and equal to -1 if $t = 1$, and making then a comparison with the properties of (6.6). This is the last step in the construction of the window function $b(\ell, j)$, which we define from its square to ensure the condition 2, that is we take the positive root of:

$$b^2(x) = \phi_3\left(\frac{x}{B}\right) - \phi_3(x), \tag{6.8}$$

note that the support of this function is (B^{-1}, B).

We have now all the ingredients to define the spherical needlet basis functions as:

$$\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_{\ell} b(\ell B^{-j}) \sum_{m=-\ell}^{\ell} Y_\ell^m(\xi_{jk}) \overline{Y}_\ell^m(x) \tag{6.9}$$

where j represents the needlets scale, λ_{jk} and ξ_{jk} are respectively the weights and the cubature points at the level j. Given the definition (6.8), the support of $b(\ell B^{-j})$ is $\ell \in (B^{j-1}, B^{j+1})$, this is the multipoles window spanned by the needlets layer j. It is immediate to recognize in the last factor the projector (6.2), so that we can define the needlet coefficient as:

$$\beta_{jk} = \int_{S^2} \mathrm{d}x \; f(x)\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_{\ell} b(\ell B^{-j}) \sum_{m=-\ell}^{\ell} a_{\ell m} Y_\ell^m(\xi_{jk}) \tag{6.10}$$

where we use:

$$a_{\ell m} = \int_{S^2} \mathrm{d}x \; f(x)\overline{Y}_\ell^m(x). \tag{6.11}$$

Equation (6.10) is the direct needlet transform, and β_{jk} are the needlet coefficients at scale j in the position defined by the cubature point ξ_{jk}. Thanks to the properties (6.4), the harmonic coefficients $a_{\ell m}$ can be computed numerically with a high level of precision. This makes the numerical implementation of needlets very convenient.

The inverse of equation (6.10) is:

$$f(x) = \sum_{jk} \beta_{jk}\psi_{jk}(x), \tag{6.12}$$

that can be derived from:

$$\sum_{jk} \beta_{jk}\psi_{jk}(x) = \sum_{j} \sum_{\ell m} b(\ell B^{-j}) b(\ell' B^{-j}) a_{\ell m} Y_\ell^m(x) \sum_{\ell' m'} \sum_{k} Y_\ell^m(\xi_{jk}) \overline{Y}_{\ell'}^{m'}(\xi_{jk}) \lambda_{jk}. \tag{6.13}$$

noting that the last summation over k in this relation is the spherical harmonic orthogonality relation, computed on the cubature points ξ_{jk} with weights λ_{jk}, yielding $\delta_\ell^{\ell'} \delta_m^{m'}$, we can write:

$$\sum_{jk} \beta_{jk}\psi_{jk}(x) = \sum_{j} \sum_{\ell m} b^2(\ell B^{-j}) a_{\ell m} Y_\ell^m(x). \tag{6.14}$$

recalling now the partition to unity property of the window function $\sum_j b\ell B^{-j} = 1$, we recover equation (6.12):

$$\sum_{jk} \beta_{jk}\psi_{jk}(x) = \sum_{\ell m} a_{\ell m} Y_\ell^m(x) = f(x). \tag{6.15}$$

As an additional remark, we highlight that at each layer the needlet coefficients have zero mean:

$$\sum_k \beta_{jk}\sqrt{\lambda_{jk}} = 0 \tag{6.16}$$

Relations (6.10) and (6.12) are direct consequences of the tight frame nature of the needlet representation.

Frames are a key concept in functional analysis and are very common in signal representation, since they allows to define redundant bases. Given an arbitrary function f on the space V with inner product $\langle ., . \rangle$, and two constants c, C, a frame is a set of functions $\{e_i : i \in \mathcal{I} = [1, 2, 3, ...]\}$ such that:

$$c||f||^2 \leq \sum_{\mathcal{I}} |\langle f, e_i \rangle|^2 \leq C||f||^2, \tag{6.17}$$

where c, C represent the frame bounds. A *tight frame* has $c = C$, needlets satisfies this condition with $C = c = 1$, on the space of the squared integrable functions on the sphere $L^2(S^2)$. Calling ψ_i the basis functions and β_i the projection of the function $f \in L^2(S^2)$ on the ith frame element (since these results are general, we do not strictly follow the needlets notation here), recalling then the inner product on $L^2(S^2)$ we will have:

$$\beta_i = \langle f, \psi_i \rangle_{L^2(S^2)} = \int_{S^2} \mathrm{d}\sigma f \psi_i, \tag{6.18}$$

where $\mathrm{d}\sigma$ is the area element on the sphere. Therefore, dropping the subscript on the inner product for simplicity of notation, we have the Parseval identity:

$$||f||^2 = \sum_i |\langle f, \psi_i \rangle|^2 = \sum_i \beta_i^2, \tag{6.19}$$

and :

$$f = \sum_{\mathcal{I}} \langle f, \psi_i \rangle \psi_i = \sum_{\mathcal{I}} \beta_i \psi_i; \tag{6.20}$$

this equation states trivially that the function f can be represented with the numbers β_i and the elements of the basis ψ_i, and it is equivalent to (6.12).

A key property of frames is that they can represent a function but they are not necessarily a basis. They can be *overcomplete*, in the sense that we can add elements to the set $\{\psi_i\}$ and still satisfy the condition set by equation (6.17). Moreover, the elements of the frame do not have to be normal, but just satisfy $||\psi_i|| \leq \sqrt{C}$. If $C = 1$

the frame is called normalized. Note that an orthonormal basis is always a normalized tight frame, but the opposite is not necessary.

It is easy to show that the needlet basis satisfies the Parseval identity. We compute:

$$\begin{aligned}\sum_{jk}\beta_{jk}^2 &= \sum_{jk}\lambda_{jk}\left[\sum_{\ell} b(\ell B^{-j})\sum_m a_{\ell m}Y_\ell^m(\xi_{jk})\right]^2 \\ &= \sum_j\sum_{\ell\ell'} b(\ell B^{-j})b(\ell' B^{-j})\sum_m a_{\ell m}\overline{a}_{\ell' m'}\sum_k Y_\ell^m(\xi_{jk})\overline{Y}_{\ell'}^{m'}(\xi_{jk})\lambda_{jk} \\ &= \sum_j\sum_{\ell} b^2(\ell B^{-j})\sum_m a_{\ell m}\overline{a}_{\ell' m'} \\ &= \sum_{\ell m}|a_{\ell m}|^2 = \sum_\ell (2\ell+1)\hat{C}_\ell = ||f(x)||^2, \end{aligned} \tag{6.21}$$

where we use the same relations exploited in the computation of (6.12) and in the last equality we resort to the definition of the power spectrum estimator, the pseudo-C_ℓ:

$$E[C_\ell] = \hat{C}_\ell = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2. \tag{6.22}$$

On the other side we have:

$$||f(x)||^2 = \int_{S^2}(d)x\ f^2(x) = \int_{S^2}(d)x\left(\sum_{\ell m} a_{\ell m}Y_\ell^m(x)\right) \tag{6.23}$$

$$= \sum_{\ell m\ell' m'} a_{\ell m}\overline{a}_{\ell' m'}\int_{S^2}(d)x\ Y_\ell^m(x)\overline{Y}_{\ell'}^{m'}(x)$$

$$= \sum_{\ell m}|a_{\ell m}|^2 = \sum_\ell (2\ell+1)\hat{C}_\ell = \sum_{jk}\beta_{jk}^2. \tag{6.24}$$

These two relations prove that the needlet basis is a tight frame with $C = c = 1$, and also that the transform conserves the "energy" of the signal.

After the definition and the construction of needlets, with these relations in hand we will now review the properties that make the needlets representation a good choice for CMB analysis. Needlets owe the name to their localization properties. As we explained in the previous section, the basis functions are localized quasi-exponentially around their centers, represented by the cubature points ξ_{jk}. In their seminal paper, (Narcowich et al. 2006) proved this statement showing that, for any point x on the sphere surface there exists a constant c_M so that:

$$|\psi_{jk}| \leq \frac{c_M B^j}{(1+B^j\arccos(\xi_{jk},x))^M}. \tag{6.25}$$

Note that the function arccos in the above formula represents the distance on the sphere; this states exactly that the function ψ_{jk} decrease faster than any power law. This property is of the utmost importance in CMB analysis, where the presence of missing observations poses a serious problem for the computation of harmonic coefficients. The consequence is the onset of spurious correlations between the harmonic coefficients $a_{\ell m}$. These correlations represent a great limitation for the evaluation of the power spectrum and the other cumulants, and must be corrected for, with high computational costs. As proven in (Marinucci et al. 2008) instead, thanks to their localization properties, the needlet coefficients are much less sensitive to gaps in the map. This is a precious feature for CMB analysis since working in needlet space allows to avoid to correct for missing observation.

Needlets are particularly well suited for the representation of random fields on the sphere, thanks to their uncorrelation properties. The fact that the needlet window function $b(\ell B^{-j})$ has compact support in (B^{j-1}, B^{j+1}) ensures that theoretical correlations between β_{jk} must cancel if the difference in levels is greater than 2, so that if $j - j' > 2$ we have:

$$\beta_{jk}\beta_{j'k'} = \sqrt{\lambda_{jk}\lambda_{j'k'}} \sum_{\ell\ell'} b(\ell B^{-j}) b(\ell' B^{-j'}) \sum_{mm'} a_{jk} a_{j'k'} Y_\ell^m(\xi_{jk}) Y\ell'^{m'}(\xi_{jk}) = 0, \quad (6.26)$$

as the simple consequence of the fact that the supports of the two basis functions do not overlap. Looking at the correlations within the same scale instead, it was proven in (Baldi et al. 2009) that the needlet representation of a Gaussian random field with smooth power spectrum satisfies:

$$|Corr(\beta_{jk}\beta_{jk'})| = \left| \frac{\beta_{jk}\beta_{jk'}}{\sqrt{beta_{jk}^2 \beta_{jk'}^2}} \right| \leq \frac{c_M}{(1 + B^j \arccos \xi_{jk}, \xi_{jk'})} \quad (6.27)$$

for any positive integer M and $c_M > 0$. So that the needlet coefficients at high frequency layers behaves as a sample of i.i.d. random variables. Moreover, the variance of the coefficients is directly correlated to the underlying power spectrum:

$$\langle \beta_{jk}\beta_{jk} \rangle = \lambda_{jk} \sum_\ell b^2(\ell B^{-j}) \sum_m \langle a_{jk} a_{jk} \rangle Y_\ell^m(\xi_{jk}) Y\ell^m(\xi_{jk}) \quad (6.28)$$

$$= \frac{1}{N_j} \sum_\ell b^2(\ell B^{-j}) \frac{2\ell + 1}{4\pi} C_\ell = \sigma_j^2 \quad (6.29)$$

where N_j is the number of coefficients in the layer, and in the derivation we use $P_\ell(\cos(0)) = 0$. The needlet coefficients of a Gaussian random field at the level j are thus extracted from a Gaussian with zero average and variance defined in (6.29).

6.2 Sparseness

Sparseness is a key property in wavelets regression methods, since it allows developing efficient techniques to disentagle coherent signals and stochastic noise. A signal is said sparse if, in a given basis or frame, it can be represented using only few of the basis elements. The property to give a sparse representations of functions is related both to the characteristics of the functions and of the basis itself.

In the case of wavelet regression, and of space-frequency representations in general, it derives from the property to identify discontinuities, coupled with the multi-resolution nature of the basis. Empirically this can be understood from the fact that, in general, coherent signals are smooth functions, since in order to be coherent they cannot have too many discontinuities. Since the wavelet basis functions correlate with fluctuations and discontinuities, a smooth signal will be represented by a few coefficients, encompassing the characteristic behavior of the original function. Random noise, on the other hand, is neither a coherent signal nor a smooth function, but a series of fluctuations without a characteristic scale or position. As a consequence, the noise signal will be spread among all the coefficients. In this hand-waving example, the result of the projection on a wavelet basis of a coherent signal plus a random noise component will be a set of only few coefficients with high signal to noise while all the others will be noise dominated. This is the intuition behind wavelet denoising methods based on thresholding techniques. Note however that the smoothness of the function is not a necessary condition to obtain a sparse representation in wavelets space. Functions with strong discontinuities can also be sparse provided that, for example, the jumps are well localized. This implies that sparseness is a more general property than smoothness.

Wavelets are not the only kind of basis that provides a sparse representation. In some cases, it is possible to combine different bases together, or add elements to the original system, to obtain overcomplete bases. As we saw in the previous section, an overcomplete basis is a frame. In signal processing literature, such systems are also called dictionaries. An example is the system obtained coupling a periodic basis with a localized one. So that certain coefficients encompass the periodic features while others identify narrow

discontinuities. Another example can be obtained describing a data-set defined on a two dimensional space using more than two dimensions, *i.e.* more that two basis vector. Intuitively, to use more basis elements implies an higher probability that the information about the peculiar features of the signal collapses on a single elements. As a rule of thumb, we can state that more the system used is rich, more the representation is sparse.

We emphasize the fact that the concepts treated here are general, but their application on a specific case requires a careful modeling of the system used with respect to the purpose of the analysis.

In our case, the interesting properties of the needlet system described in the previous sections, give us good reasons to consider them as a very well suited system for our investigation.

6.3 Thresholding

Thresholding algorithms are filtering techniques aiming at the separation of the signal from the noise by exploiting the sparseness of wavelet representations. Following the ideas presented in the previous section, a thresholding algorithm will set to zero all the coefficients under a certain threshold. The basic assumption is precisely that, in a mixture of a random noise and a coherent signal, the significant part of the information (*i.e. the signal*) collapses into few high S/N coefficient. It is, in spirit, similar to principal component analysis, that aim to reduce the complexity of a multidimensional data-set identifying the most significant components.

The straightforward application is called hard thresholding (HT), the effects of the hard thresholding operator on the needlet coefficients are simply:

$$HT(\beta_{jk}) = \begin{cases} 0 & \text{if} \quad |\beta_{jk}| < \lambda \\ \beta_{jk} & \text{if} \quad |\beta_{jk}| \geq \lambda \end{cases} \tag{6.30}$$

where λ is a given threshold. In the case of a coherent signal, only few significant coefficients survives this operation, providing an optimal representation as well as an efficient data compression.

Another option is provided by the soft thresholding (ST). The ST operator differs slightly from the hard thresholding one, in the sense that the significant coefficients are now

rescaled proportionally to the chosen threshold:

$$ST(\beta_{jk}) = \mathrm{sgn}(\beta_{jk})(|\beta_{jk}| - \lambda)_+ \begin{cases} \beta_{jk} + \lambda & \text{if} \quad \beta_{jk} \leq \lambda \\ 0 & \text{if} \quad |\beta_{jk}| < \lambda \\ \beta_{jk} - \lambda & \text{if} \quad \beta_{jk} \geq \lambda \end{cases} \tag{6.31}$$

where the operator $(*)_+$ stands for the positive part of the argument.

It is interesting to notice that the soft thresholding solution can be interpreted from a Bayesian perspective as a maximum posterior estimator from a Gaussian Likelihood with a leptokurtic prior on the parameters, that in this case are the needlet coefficients. Assume that we observe a signal θ, that we know to be sparse in a given basis, from a noisy data-set with known noise variance σ^2, so that each data point $x = \theta + n$ will be extracted from a Gaussian distribution with mean θ and (known) scale σ. Assume also that the scale parameter $1/\lambda$ of the Laplace prior on θ is known, so that we can write:

$$P(\theta|x) \propto \mathcal{L}(x|\theta)P(\theta) = N(x;\theta,\sigma)L(\theta;\lambda,0), \tag{6.32}$$

$$-\log P(\theta|x) \propto \frac{(x-\theta)^2}{2\sigma^2} + \lambda|\theta| + const. \tag{6.33}$$

note that we use $N(*;\mu,\sigma)$ and $L(*;\mu,\lambda)$ to define respectively Normal and Laplace distributions. The maximum posterior estimator (MPE) is obtained by minimizing equation (6.33). We will show now that this problem can be reduced to the soft thresholding.

We start taking the derivative with respect θ (that we denote with ∂_θ):

$$\partial_\theta(\log P(\theta|x)) = -\frac{(x-\theta)}{\sigma} + \lambda\partial_\theta|\theta| = 0, \tag{6.34}$$

$$\hat{\theta} = x - \sigma^2\lambda\partial_\theta|\hat{\theta}|, \tag{6.35}$$

since the absolute value is not differentiable around zero (and equivalently the L1 norm $||\theta||$ dealing with multidimensional data), we should take the subgradient, so that we

have:

$$\partial_\theta ||\theta|| = \begin{cases} 1 & \text{if } \theta > 0 \\ -1 & \text{if } \theta < 0 \\ [-1, 1] & \text{if } \theta = 0, \end{cases} \tag{6.36}$$

note that for $\theta = 0$ the subgradient is actually an interval of values. We can understand the soft thresholding solution applying the conditions (6.36) at equation (6.35). First notice that the "correction" term $\sigma^2\lambda\partial_\theta|\hat{\theta}|$ can only take values in the interval $[-\sigma^2\lambda, \sigma^2\lambda]$. So that, looking at the case $|x| \geq \sigma^2\lambda$, θ must be $\neq 0$ since this "correction" term cannot account for signals higher than $\sigma^2\lambda$. Moreover we must have $\text{sgn}(\hat{\theta}) = \text{sgn}(x)$. The solution in this case is so $\hat{\theta} = x - \text{sgn}(x)\sigma^2\lambda$. On the contrary, if $|x| < \sigma^2\lambda$, we take $\hat{\theta} = 0$ and we can consider all the signal in x as coming from the subgradient term. After these considerations it is clear that the solution coincides with the soft thresholding operator, that in this case is:

$$ST(x) = \text{sgn}(x)(|x| - \sigma^2\lambda)_+. \tag{6.37}$$

The crucial part of any thresholding method is the choice of the proper threshold. A proper threshold must ensure that all the significant information is conserved while the major part of spurious noise is removed. If the noise is Gaussian with known variance σ^2, a common technique is the $K\sigma$ clipping, where K is an integer. In this case the threshold works similarly to an hypothesis test (the HT in particular works exactly in the same way). Basically, a coefficient is selected if its significance level with respect to the null hypothesis of zero signal plus noise is high enough in terms of standard deviations.
If the noise is properly modeled, the application is trivial, otherwise an estimate of the noise shall be performed directly from the data. Usually in a wavelets analysis, estimates of the noise are obtained looking at the high frequency layers, that in general are noise dominated. If the noise is known to be not stationary in frequency space, the threshold can be modulated scale by scale for an optimal shrinkage. A widely used solution, called universal threshold, has been introduced by (Donoho & Johnstone 1994), it sets as threshold:

$$\lambda = \hat{\sigma}\sqrt{\frac{2\log(n)}{n}} \tag{6.38}$$

where n in the number of coefficients in the layer and $\hat{\sigma}$ is an estimate of the noise standard deviation. The standard deviation is often computed from the median absolute deviation:

$$\hat{\sigma} = \sqrt{n}\frac{\text{median}(|\beta_{jk} - \text{median}(\beta_{jk})|)}{0.6745}. \tag{6.39}$$

A large number of different techniques, developed to deal with different kinds of noise, exist in the literature. However, since a proper threshold must be tailored on the data of interest, we will refer to the existing literature (see *e.g.* (Wasserman 2006; Starck & Murtagh 2006)) for a complete review.

Chapter 7

Foreground template fitting with Needlet thresholding

In this chapter we apply the techniques just presented to the component separation problem in CMB surveys. The final results of this investigation will be soon submitted for publication as (Oppizzi et al. 2018b). The intuition behind this investigation is that foreground signals and the CMB fluctuations disentangle when the data are represented in a proper basis, frame or dictionary. Given the properties presented in the previous section, needlets seem an ideal choice for this purpose. It is interesting to stress from the start that the method we are going to describe does not rely on multi-frequency information. It can therefore represent a useful tool, especially for experiments with limited frequency coverage.

Foreground emission comes in the larger part from coherent sources concentrated around the galactic plane and in few large structures that extend at higher galactic latitudes. As we saw in the previous sections, space-frequency representations of coherent signals naturally tend to be sparse. Needlets are clearly not an exception: as a matter of fact, their good localization properties, coupled to their tight frame nature are the exact requirements to induce sparseness when describing smooth functions on the sphere. We will thus expect that the contribution from galactic foreground will be concentrated in few large coefficients that can be identified and fitted with a needlet thresholding technique.

On the other hand, CMB has very different features, since it is a random field and not a coherent signal. Thanks to their uncorrelation properties, needlet are very well suited to represent random signal; the needlet coefficients representing a Gaussian random field

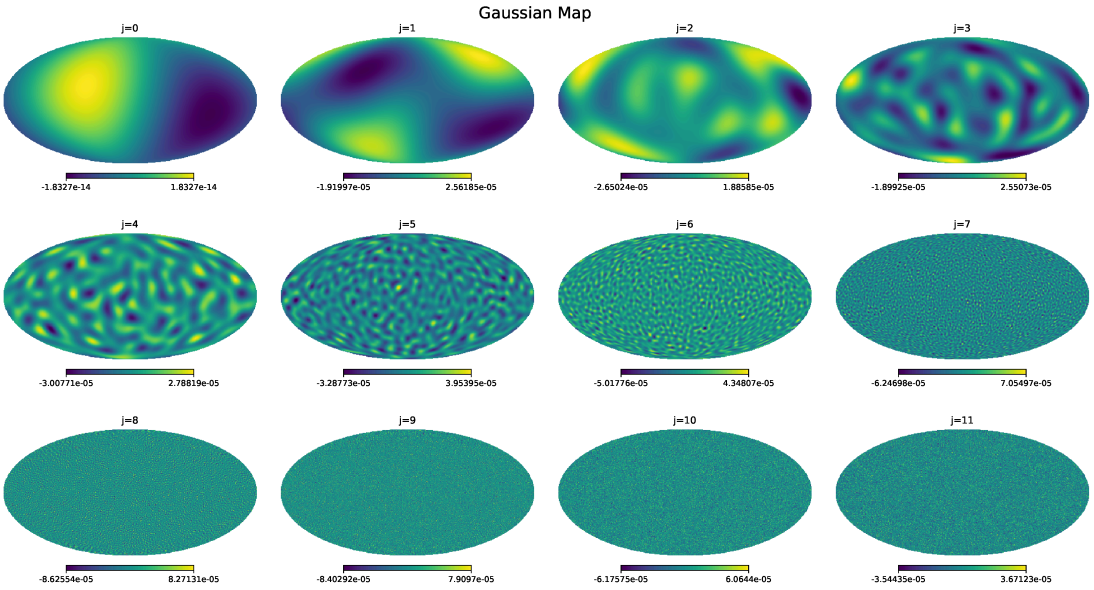

FIGURE 7.1: Needlet coefficients of a CMB realization for each frequency layer.

are *i.i.d.* Gaussian random variables. We also expect that, unlike foregrounds, CMB needlet space representation will not be sparse. The reason is that the CMB does not form coherent structures at all, but it is instead a uniform fluctuations field at any scale.

From this considerations, we assume that the needlet frame is the ideal system to separate the underlying Gaussian fields from the foreground emission. In the next section we will investigate the behavior of needlet representation of foreground emissions and CMB, we will then present our technique and some preliminary results.

Our algorithm are based on the HEALPix packages, so that the cubature points where the coefficients are computed will follow the HEALPix pixelation scheme.

7.1 Needlet representation of CMB and foregrounds

In this section we will investigate the principal properties of foregrounds and CMB representation in the needlet frame. Our main interest is in particular to verify the hypothesis that the foreground signal is sparse and correlated, while the CMB signal is not. Moreover, since, as we saw, the assumption of sparsity is related to an underlying non-Gaussian, leptokurtyc distribution of the signal coefficients, we expect the foreground β_{jk} to follow such a distribution. The CMB coefficients must instead follow an asymptotically (with the scale) Gaussian distribution. We will confront random CMB realizations with foreground reconstructions both from the Planck legacy archive[1]

[1]Based on observations obtained with Planck (http://www.esa.int/Planck), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada.

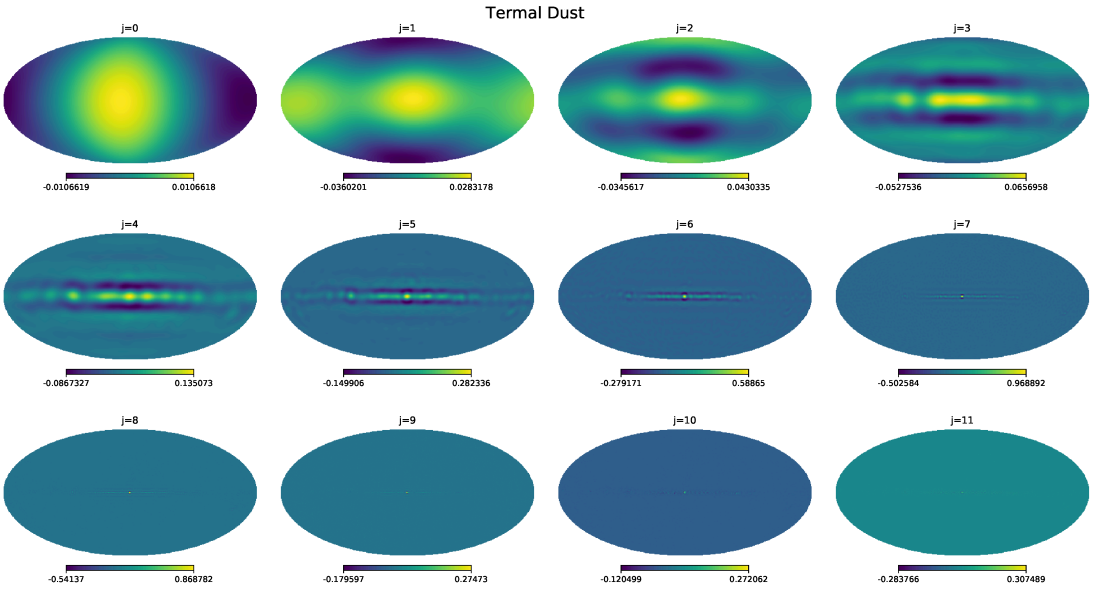

FIGURE 7.2: Needlet coefficients of a thermal dust emission template for each frequency layer

and computed from the Planck sky model. The foregrounds template from the Planck archive have been obtained with Commander (Eriksen et al. 2004), while the foregrounds simulations are computed with the Python Sky Model (PySM) algorithm described in (Thorne et al. 2017).

We show in figure 7.1 the needlet decomposition of a random CMB realization. It is evident that the signal is spread over all the needlet coefficients, as expected, given its stochastic nature. Furthermore, looking at the levels corresponding to the largest scales, it is easy to notice that only adjacent layers shows some level of correlation. Needlets actually split a continuous field in several independent realizations, the layers, each one covering a limited range of frequencies.

In figure 7.2 we show instead the same decomposition for a thermal dust template. The differences are glaring. First we see how the information is actually concentrated only in few coefficients located near the galactic plane. The lowest layers trace the diffuse emission while the higher frequency levels contain only few small scale corrections. Moreover, it is clear that all scales are highly correlated at any distance in frequency level, and this is expected since the signal is coherent. This shows how needlet are very efficient in minimizing spurious scale-scale correlations and highlighting the real ones.

Another interesting test is to look at the distribution of the needlet coefficients of these different signals. We expect that a sparse signal like the thermal dust template will present a distribution that tends to a Laplacian behavior, *i.e.* a double exponential.

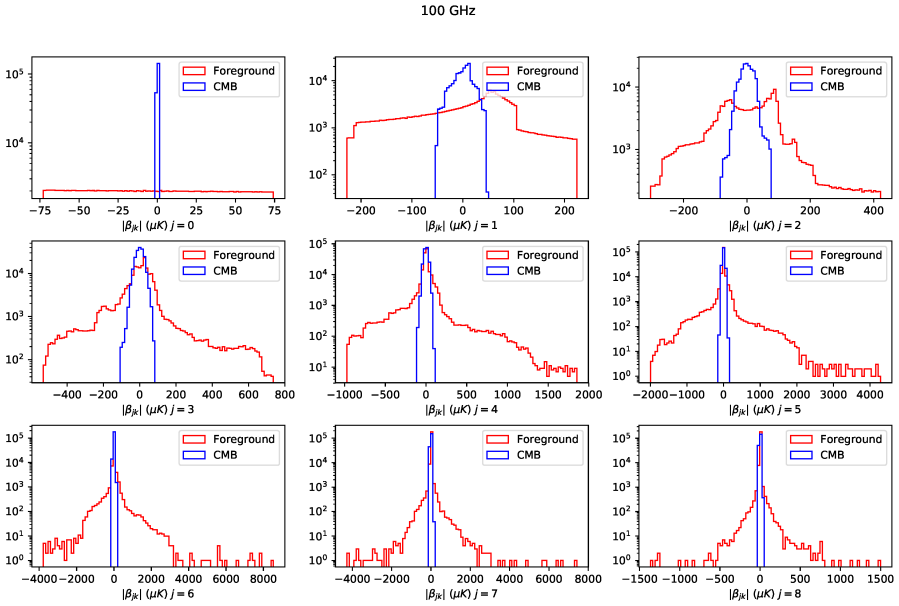

FIGURE 7.3: Histograms of the needlet coefficients of a CMB realization (Blue) and a template of the total foreground emission at 100Ghz (Red) for different needlet layers j.

The Gaussian field coefficients must instead tend to a normal distribution. This features are well captured by figure 7.3. These plots represent the histograms of CMB realizations and of a template representing the total foreground emission in the 100 Ghz channel of Planck.

We then compute the "energy scaling" of a needlet space map with respect to the number of coefficients considered. First we define the "map energy" as $\sum_{jk} \beta_{jk}^2$. We then iteratively remove coefficients from the summation, starting from the smallest one (in absolute value), and for each step we compute the residual energy. The scaling of this quantity with respect to the fraction of deleted coefficients provides an estimates of the sparseness. The energy of a sparse signal will come from the few percent largest coefficient, and it will show an initially constant trend, followed by an abrupt decay, corresponding to the removal of the informative coefficients. A non sparse signal will instead show a more uniform scaling behaviour everywhere. The results of this test on two foreground templates and a CMB map is showed in figure 7.4. The left panel shows the scaling for a Gaussian field and for the reconstruction of the foreground contribution in the 353 GHz channel of Planck, while the right panel represents the scaling for the foreground contributions at 30 Ghz. Both these channels are foreground dominated, the former by thermal dust emission and the latter by synchrotron and bremsstrahlung. We

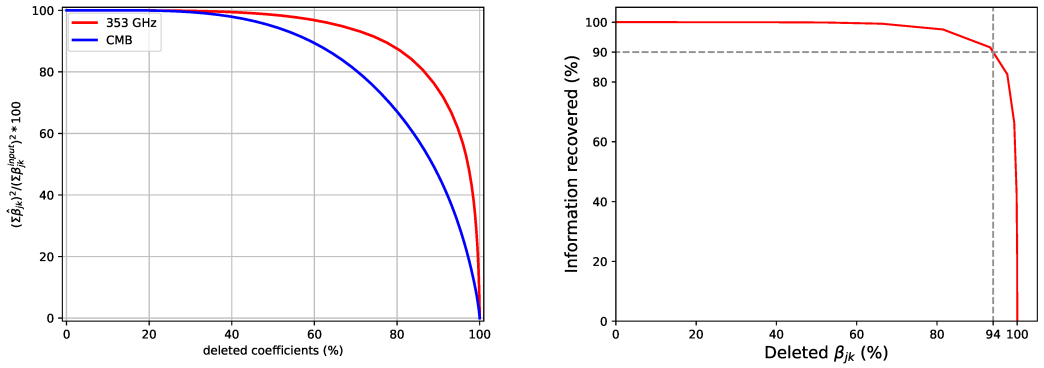

FIGURE 7.4: Right Panel: energy scaling of a template of foreground emission at 353 GHz (Red) and of a CMB realization (Blue). Left panel: energy scale of a 30Ghz foreground template (Red).

see how the CMB curve remains well below the others, showing a more regular scaling. The two foreground templates instead, especially at 30 GHz, clearly present an initial constant trend and then a fast decay, a clear indication of sparseness. As pointed out in the left panel, 90% of the template energy is stored in just 6% of the β_{jk}

The results of these tests are in line with the expectations from our theoretical considerations about CMB and foreground representations behavior in light of the characteristics of the needlet frame. An interesting check of the needlet properties is to look at the coefficient histograms before and after a real space masking. Figure 7.5 show the histograms of the β_{jk} of the same maps showed in 7.3 after the removal of the galactic plane with a real space masking. Even after the removal of the most contaminated regions in real space, still just a few large spurious coefficients remain. This suggests that a needlet representation of the signal works better than a real space one to isolate the spurious emission in few coefficients.

7.2 Thresholding Implementation

In this section we show a preliminary investigation of the performance of the needlet thresholding technique in the separation of foreground and CMB signals. Notice that, given the nature of the method, it is more correct to say that it separates the Gaussian and the strongly non-Gaussian components of the signals. To evaluate the results we look at the reconstruction of the power spectrum of the CMB underlying field from simulated observations. We also investigate the synergy of this separation method with other techniques such a simple ILC algorithm. At this stage, our interest is in assessing

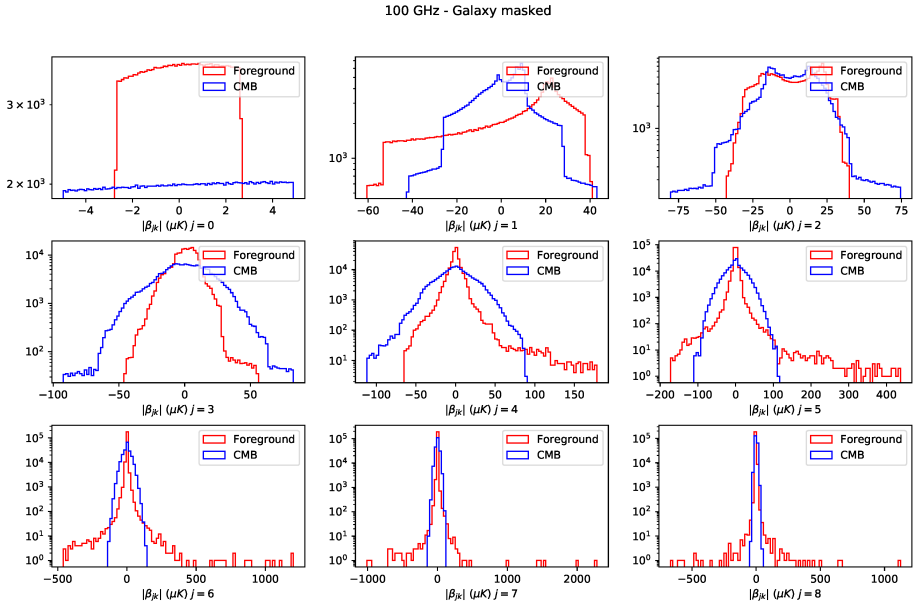

FIGURE 7.5: Histograms of the needlet coefficients of a CMB realization (Blue) and a template of the total foreground emission at 100Ghz (Red) for different needlet layers j after the masking of the Galactic plane ($f_{sky} = 0.8$).

the impact of the thresholding rather than developing a complete component separation pipeline.

The pipeline adopted in these test is quite straightforward: we first build up a suite of low resolution simulations of the microwave sky at the Planck frequencies. We then apply a needlet thresholding algorithm to each map to obtain template of the foregrounds emission. These templates are fitted and then removed from the input data to obtain the cleaned map. We compute then the power spectra from each reconstructed map and then we compare it with the spectra of the input CMB realizations. We test many different solutions and algorithm, here we show the results obtained with an hard thresholding algorithm. We use the universal threshold estimator (Donoho & Johnstone 1994) to compute the threshold adaptively for each layer.

We show in figure 7.6 the reconstruction of a foreground template power spectrum from a simulated mixture of CMB thermal dust and noise. In this example we use an hard thresholding algorithm with the threshold selected scale by scale with the universal threshold estimator. The thresholded power spectrum is reconstructed using $\sim 10\%$ of the needlet coefficients. In figure 7.7 we show instead the recovered CMB. These results are very good in terms of information compression, and they suggest that this technique is very effective at high multipoles but present some limitation in separate the signals

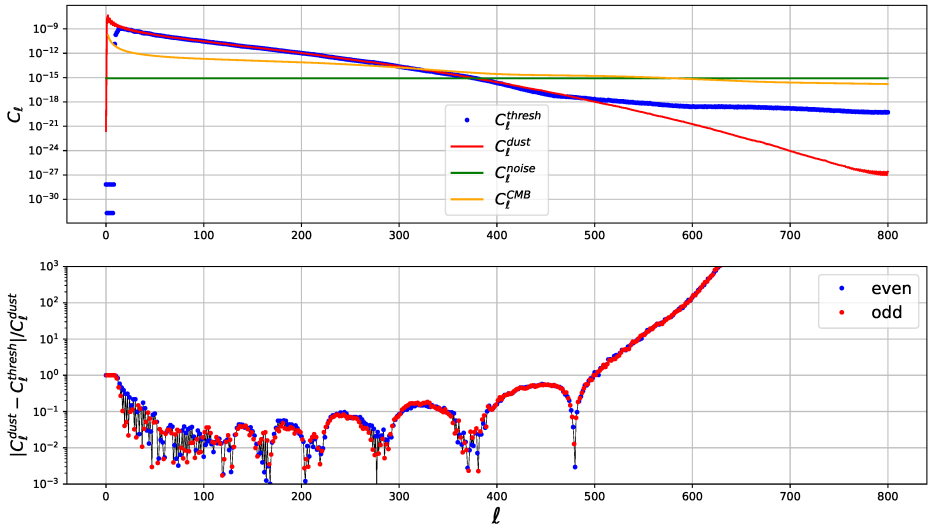

FIGURE 7.6: Power Spectra comparison for the HT method with multiple thresholds λ with maximum needlet scale $jmax = 25$. Top: angular power spectra of the thresholded map (blue dots), of the dust template (red line), CMB (yellow) and noise (green). Bottom: residuals between the dust power spectrum and the reconstructed template, blue and red dots refer respectively to even ad odd multipoles.

at the larger scales. To better highlight these feature we repeat our test on a set of low resolution simulation.

In figure 7.8 we show the confrontation between the power spectra reconstruction after the thresholding and after the real space masking of the 100 GHz channel. We chose this frequency because it lies in the region less contaminated by foreground emissions. The mask used covers the galactic plane and it removes the 20% of the sky from the analysis. The plot represents the mean and the standard deviation of the power spectra obtained from 150 simulations. Being a region with a low level of foreground emission, the galactic mask alone already removes the largest part of the contaminated regions. However, some residual power still remain. The thresholding method, instead, gives opposite results at the high and low end of the spectrum. At high frequencies, the reconstruction seems very good even in this single channel case. At the lowest multipoles instead, the method cannot separate the two components and a lot of residual power remains in the final reconstruction. This is an evidence of the fact that, at low multipoles there are not enough modes to efficiently disentangle the two components. This can be interpreted as a "cosmic variance" limit of this technique. Since it relies on data from a single map, this feature is unavoidable and must be complemented by resorting to other techniques.

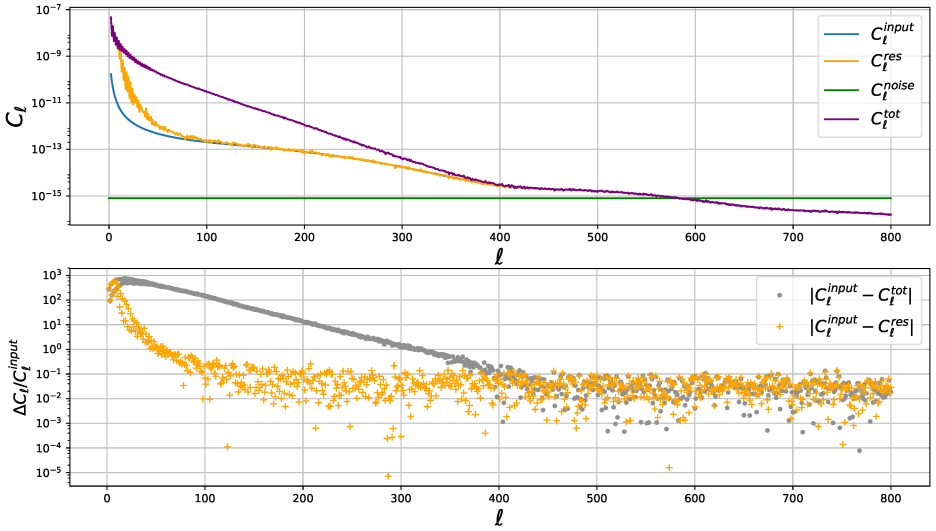

FIGURE 7.7: Power Spectra comparison for the HT method with multiple λ with $jmax = 25$. Top: angular power spectra of CMB input signal (blue line), of the residual map (orange line), and noise (green). Bottom orange: residuals between the power spectra of the input CMB signal and of the residual map (input map - thresholded map). Bottom grey: residuals between the power spectra of the input CMB signal and of the total map (CMB +signal +noise).

To evaluate the synergies with other methodologies, we test then the implementation of needlet thresholding in combination with a simple ILC algorithm. For the test presented we do not refine the ILC method and use it in its standard formulation. The idea is to start with thresholding-based "pre-cleaning" approach, for each channel, followed by a standard ILC, combining all frequencies to get the final map. Our aim here is to check whether the "pre-thresholding" can improve the performance or, on the contrary, the combined implementation of several techniques could eventually even distort the results. Since we are interested in checking the performance in experiments with few frequency channels, we used only the 3 Planck central bands for this test (*i.e.* 70,100,143 GHz). The results are in figure 7.9. In the first panel we show the outcome from a real space ILC performed over masked maps, while in the second panel the ILC is run over the map recovered after the removal of the template obtained from the thresholding. Also in this case we found that the needlet thresholding not only works better than a real space masking, but also provides a remarkable improvement in the performance of a simple ILC. The results of this preliminary investigation motivates us to study further developments of this technique.

7.3 Bayesian Treatment

In order to get a better insight of the thresholding procedure introduced in the previous section, we now illustrate and justify it within a general Bayesian framework. As shown in e.g., (Vansyngel et al. 2016), a Bayesian approach provide a way to described different component separation techniques within a unified, general formalism. This approach shows in particular how different common component separation methods amount to different choices of the prior hypothesis and sampled or marginalized parameters.

We start as usual by assuming to have observations from K channels, with a mixture of N components and M elements (pixels) in each channel (by "pixel" we mean real space pixels, $a_{\ell m}$, needlet coefficients or the elements of whichever basis is adopted to represent the signal). We recall then the linear mixture model:

$$d_i = As_i + n_i \tag{7.1}$$

where d_i is a vector of K elements representing the observations at the pixel i, s_i is a vector of N elements representing the contribution of the components at the pixel i, A is the *mixing matrix* of dimension $N \times K$ that weights the contribution of the components among the channels, and n_i is the noise on the pixel i.

The Bayesian formulation of the component separation problem aims to solve:

$$P(A, s|d) \propto \mathcal{L}(d|A, s)P(A, S) \tag{7.2}$$

as shown e.g., in (Vansyngel et al. 2016). With specific assumptions on the priors and eventually variables to marginalize over, the formulation above can be used to define typically adopted component separation techniques, such as ILC and SMICA. In our case, we want to introduce the sparsity assumption on the foreground templates. A similar hypothesis is also at the basis of the development of the GMCA algorithm in (Bobin et al. 2008). We will clarify the differences between the approach discussed here and GMCA in due course. The usual way to enforce sparsity in a Bayesian context is to assume leptokurtic priors. A common choice is the Laplace distribution:

$$L(x; \lambda, \mu) = \frac{\lambda}{2} e^{-\lambda|x-\mu|} \tag{7.3}$$

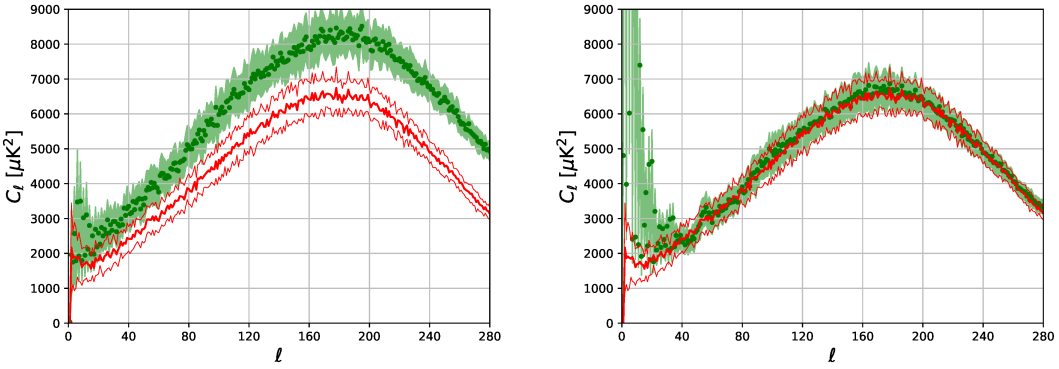

FIGURE 7.8: CMB input power spectrum and its reconstruction after a real space masking of the galactic plane ($f_{sky} = 0.8$) (Left) and the thresholding (Right) measured from 100 simulations. Red thick line: sample average of the input power spectrum, the thin red lines indicate the 1σ region. Green: sample average of the reconstructed signal, the shaded green region represent the 1σ region.

We want to study the implementation of this kind of prior in a Bayesian framework, to give a Bayesian interpretation of our estimators.

First we must rewrite the linear mixture model in a form more suited to our scope, which is ultimately that of recovering the CMB signal. Our assumption is that the foreground signal is sparse in the needlet domain, but the CMB is not (this is a first difference with the GMCA algorithm, which instead implicitly assumes sparsity also for the CMB component). We thus rewrite the linear mixture model assuming that components with a "Gaussian" and a "non-Gaussian" prior probability coexist in the data. We assume that the only "Gaussian" component is the CMB, thus we have, in the single pixel:

$$d_i = As_i + \mathbf{e}c_i + n_i = f_i + \mathbf{e}c_i + n_i \tag{7.4}$$

where we separate the CMB from the other component and we denote it with c_i, times the vector of ones $\mathbf{e}$ of length K, since it is constant between channels. Furthermore $f_i = A's_i$, where A' is the mixing matrix with the row corresponding to the CMB set to zero. We rewrite the model in this way because at this stage we are not interested in the mixing matrix A, so that instead of estimating the underlying N templates, we consider the linear combination, f_i, of all these templates in each of the K channels. We assume that a Laplace distribution is a proper prior also for this combination.

In this model, the stochasticity in the data d comes from the instrumental noise n. This is usually well modeled, independent between channel and Gaussian, but in certain basis its covariance matrix C_n is dense and can be difficult to treat. We will deal with

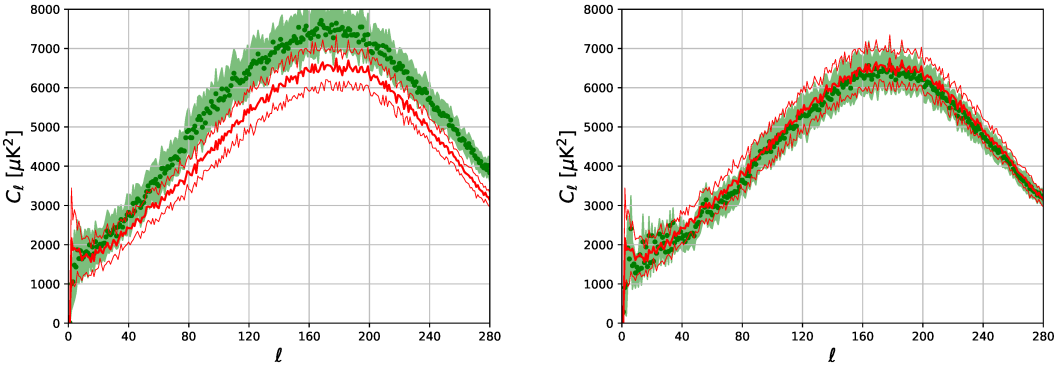

FIGURE 7.9: CMB input power spectrum and its reconstruction with an ILC of the 70, 100 and 143 GHz channel of *Planck* after a real space masking of the galactic plane ($f_{sky} = 0.8$) (Left) and the thresholding (Right) measured from 100 simulations. Red thick line: sample average of the input power spectrum, the thin red lines indicate the 1σ region. Green: sample average of the reconstructed signal, the shaded green region represent the 1σ region.

this issue later on.

We will split the problem as:

$$P(f, c|d) = P(c|f, d)P(f|d)$$

$$P(f|d) \propto P(f) \int_\infty \mathrm{d}c\; P(d|f, c)P(c) \tag{7.5}$$

$$P(f, c|d) \propto P(c|f, d)P(f) \int_\infty \mathrm{d}c\; P(d|f, c)P(c) \tag{7.6}$$

then we will solve the problem in two steps, first finding first the value $\hat{f}$ which maximize 7.5, followed by maximizing 7.6 , with this value.

To do that, we have to marginalize the likelihood with respect to c. In this step we can exploit the explicit separation between channels in this notation. Since the noise is uncorrelated and the CMB is constant, this is a good assumption. The thresholding is performed channel by channel, thus we consider d, f and c as maps and template at a certain frequency (*i.e.* vectors vector of length M). We have:

$$\mathcal{L}(d|c, f) = N(d; c + f, C_n) \propto \exp\left[-\frac{1}{2}(d - f - c)^\intercal C_n^{-1}(d - f - c)\right] \tag{7.7}$$

$$P(c) = N(c; 0, C_c) \propto \exp\left[-\frac{1}{2}c^\intercal C_c^{-1} c\right], \tag{7.8}$$

where C_n and C_c are respectively the noise and cmb covariance matrices. Therefore we can write:

$$\mathcal{L}(d|c,f)P(c) \propto \exp\left[-\frac{1}{2}d^{\mathbf{T}}C_n^{-1}d + f^{\mathbf{T}}C_n^{-1}d - \frac{1}{2}f^{\mathbf{T}}C_n^{-1}f\right] \times \\ \times \exp\left[-\frac{1}{2}c^{\mathbf{T}}\left(C_n^{-1}+C_c^{-1}\right)c + c^{\mathbf{T}}\left(C_n^{-1}d - C_n^{-1}f\right)\right], \tag{7.9}$$

to integrate out c, first we note that the only terms involved in the integral are in the second line, then we remember:

$$\int \exp\left[-\frac{1}{2}\vec{x}^T\mathbf{R}\vec{x} + \vec{B}^T\vec{x}\right]d^nx = \sqrt{\frac{(2\pi)^n}{\det R}}\exp\left[\frac{1}{2}\vec{B}^T\mathbf{R}^{-1}\vec{B}\right] \tag{7.10}$$

so that we have:

$$\mathcal{L}(d|f) = \int_\infty \mathrm{d}c\ \mathcal{L}(d|f,c)P(c) \propto \exp\left[-\frac{1}{2}d^{\mathbf{T}}C_n^{-1}d + f^{\mathbf{T}}C_n^{-1}d - \frac{1}{2}f^{\mathbf{T}}C_n^{-1}f\right] \times \\ \times \exp\left[\frac{1}{2}\left(C_n^{-1}d - C_n^{-1}f\right)^{\mathbf{T}}\left(C_n^{-1}+C_c^{-1}\right)^{-1}\left(C_n^{-1}d - C_n^{-1}f\right)\right] \tag{7.11}$$

we then implement the Laplacian prior, again assuming uncorrelation between the channels, thus we have $P(f) \propto \exp -\lambda||f||$ where $||*||$ is the L1 norm.
Adding this to 7.11 and setting $R = \left(C_n^{-1}+C_c^{-1}\right)$ for simplicity of notation, we compute the derivative of the Log Likelihood and we find:

$$\partial_f(-\log(P(f,d)) = -C_n^{-1}d + C_n^{-1}f + 2C_n^{-1}R^{-1}C_n^{-1}d - 2C_n^{-1}R^{-1}C_n^{-1}f + \lambda\partial_f||f|| \tag{7.12}$$

and then:

$$\hat{f} = d + \left(C_n^{-1} - 2C_n^{-1}R^{-1}C_n^{-1}\right)^{-1}\lambda\partial_f||f||, \tag{7.13}$$

and as we derived in the previous section 6.3, the solution to this problem is the soft thresholding operator with threshold $\left(C_n^{-1} - 2C_n^{-1}R^{-1}C_n^{-1}\right)^{-1}\lambda$.

7.3.1 ILC-like implementation

We now show how this result can be implemented within a ILC algorithm. So that we want to maximize $P(c|\hat{f},d)$, after deriving $\hat{f}$ from the thresholding (in other words, we "pre-clean" each channel with thresholding, followed by combining the channel via ILC). We now consider all the channels together, so that we reintroduce the vector $\mathbf{e}$

and the index i representing the single pixel (or multipole, or needlet coefficient). We use this notation here since our aim is to present general results, that do not depend from the representation of the dataset. On the other hand, the structure of the data covariance changes significantly from one basis to another, so that the optimal approach can change.

$$P(c_i|d_i,\hat{f}_i) \propto \mathcal{L}(d_i|c_i,\hat{f}_i)P(c_i) = N(d_i; c_i+\hat{f}_i, C_N)N(c_i;0,C_{c_i})), \tag{7.14}$$

note that C_N is the covariance matrix of the noise between channels and c_i, C_{c_i} are numbers. Now again, by differentiating the log-Posterior with respect c_i we find:

$$\begin{aligned} -\log(P(c_i|d_i,\hat{f}_i)) &\propto \frac{1}{2}(d_i-\hat{f}_i-\mathbf{e}c_i)^{\mathbf{T}}C_N^{-1}(d_i-\hat{f}_i-\mathbf{e}c_i)+\frac{1}{2}c_i^2C_{c_i}^{-1} \\ \partial_c\left(-\log(P(c_i|d,\hat{f}_i))\right) &\propto (\mathbf{e}^{\mathbf{T}}C_N^{-1}\mathbf{e}+C_{c_i}^{-1})c_i+\mathbf{e}^{\mathbf{T}}C_N^{-1}(\hat{f}_i-d_i)=0 \end{aligned} \tag{7.15}$$

and in conclusion we can find the new Bayesian ILC formula as:

$$c_i = \frac{\mathbf{e}^{\mathbf{T}}C_N^{-1}}{\mathbf{e}^{\mathbf{T}}C_N^{-1}\mathbf{e}+C_{c_i}^{-1}}(d_i-\hat{f}_i) \tag{7.16}$$

or, looking instead at the maximum likelihood the solution is more similar to the standard ILC

$$c_i = (\mathbf{e}^{\mathbf{T}}C_N^{-1}\mathbf{e})^{-1}\mathbf{e}^{\mathbf{T}}C_N^{-1}(d_i-\hat{f}_i). \tag{7.17}$$

This is equivalent to applying soft thresholding in different channels, and then subtracting the results from the data-set before performing ILC, as we did in the test presented before. Note that a similar approach can of course be adopted to find algorithms which combine thresholding with a variety of other methods, such as template fitting. Working in needlet space, thresholding could be a straightforward implementation in others needled based techniques as NILC. The thresholding can obtain CMB cleaned foreground templates from one channel. This would be very useful when it is impossible to exploit multi-frequency information, or the channels are to few to rely in adjacent channels differences.

As a further comment, we point out that even if here we consider a single Gaussian component (the CMB), we also can also assume a mixture of different Gaussian signals, under the hypothesis that foreground emission has a Gaussian residual. So that we would express $c = As'$, where s' are the templates after the removal of the strongly

non-Gaussian part. This would not change the ILC solution just computed (at least the maximum likelihood solution), but can represent an improvement for methods relying on Gaussian prior on the foreground templates.

7.4 Threshold Selection

The main problem of thresholding algorithm is the selection of the threshold. We presented some of the standard techniques in the previous chapter, but we also study here some solutions suited for our specific goal. The first problem is that the procedure for the selection of a proper λ cannot be computationally expensive, since it is only a pre-processing step.

Since the CMB is well known to be a perfectly isotropic and Gaussian field, natural solutions can be based on the minimization of the non-Gaussianity or the anisotropy of the cleaned map. Under the assumption of isotropy, the CMB needlet coefficients are equally distributed with variance:

$$\sigma_j^2 = \frac{1}{N_j} \sum_{\ell=B^{j-1}}^{B^{j+1}} \frac{2\ell+1}{4\pi} b^2 \left(\frac{\ell}{B^j}\right) C_\ell. \tag{7.18}$$

This cannot be the case if we have significant foreground contamination. We can thus look for the threshold that minimize the following "anisotropy indicator":

$$\Delta_j = \frac{1}{N_j} \sum_k \left[\frac{\beta_{jk}^2}{\sigma_j^2} - 1\right]^2 \tag{7.19}$$

where the variance can be estimated from the coefficients itself.

In similar fashion, one can seek to minimize the non-Gaussianity of the coefficients. A simple way to do that by setting the threshold that minimize the coefficients sample skewness (of course this could be extended to higher order cumulants, or to include spatial information, e.g., the full bispectrum/trispectrum, accounting for mode-mode coupling):

$$\hat{\gamma} = \frac{1}{N_j} \frac{\sum_k (\beta_{jk})^3}{\sigma_j^{3/2}} \tag{7.20}$$

where again, N_j is the number of coefficient at the level j. Both these conditions can be implemented in an iterative threshold selection algorithm.

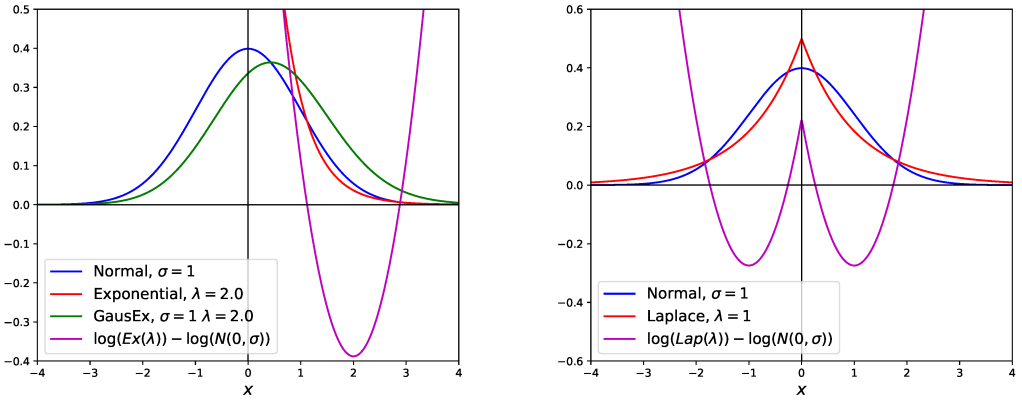

FIGURE 7.10: Some examples of the distributions described in the text, in magenta we show the log-ratio between the Gaussian an the exponential (Left) and the Laplace distribution (Right) for these particular configuration.

An alternative is to build a "classifier algorithm", defining the signal as a Gaussian/Laplace or a Gaussian/exponential mixture and then compute the probability that each coefficient is extracted from one of the two distributions. To show this procedure, we now an proceed to illustrate an example, assuming a mixture of a Gaussian variable with mean $\mu = 0$ and variance σ^2 and an exponentially distributed variable with factor λ. We choose to use the exponential/Gaussian mixture here for the sake of clarity. In this particular case, the resulting variable has a well-known distribution, the so called Exponentially modified Gaussian, so that the procedure can be exactly visualized in therms of such distribution. However the same procedure can be applied to any mixture, recalling that the cumulants of the sum of two *i.i.d.* random variables are the sum of the cumulants of the original distributions. Furthermore, if one of the variable is Gaussian with zero mean, the cumulants above the 2nd of the distribution of the mixture variable depend only on the non-Gaussian component. Therefore, even assuming fairly complicated NG distributions, it should be possible to obtain estimates of the parameters, as long as there is a known relation with the cumulants.

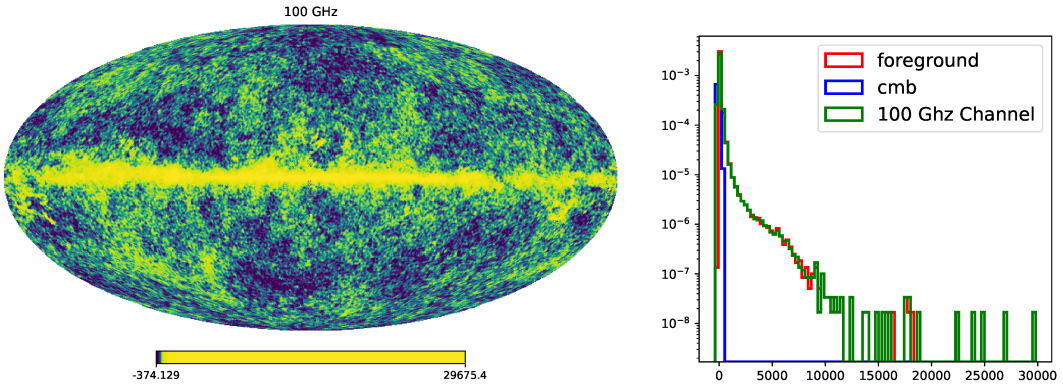

FIGURE 7.11: Left: Simulated map of a 100 GHz observation. Right: Histograms of pixels of the underlying templates and the map

That said, naming x_i the Gaussian camponent, y_i the NG one, and d_i the resulting data-set, where i run over the pixels, for the exponential case we have:

$$\begin{aligned} d_i &= x_i + y_i \\ P_X(x_i|\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x_i^2}{2\sigma^2}} \\ P_Y(y_i|\lambda) &= \lambda e^{-\lambda y_i} \\ P_B(d_i|\sigma,\lambda) &= 0.5\lambda e^{-0.5\lambda\left(2d_i - \lambda\sigma^2\right)} \left[1 + \text{erf.}\left(\frac{d_i - \lambda\sigma^2}{\sqrt{2}\sigma}\right)\right] \end{aligned} \tag{7.21}$$

The last expression is the Exponentially modified Gaussian Distribution, named also exGaussian or GausEx. The relations between the moments of this distribution and the parameters of the underlying Normal and Exponential distributions, are:

$$\begin{aligned} m &= \mu + 1/\lambda, \\ s^2 &= \sigma^2 + 1/\lambda^2, \\ \gamma &= \frac{2}{\lambda^3 s^3}. \end{aligned} \tag{7.22}$$

where m is the sample mean, s the sample standard deviation and γ the sample skewness. We can use equations 7.22 to estimate λ and σ and use them to compute the threshold, looking at the ratio of the two distribution.

We search for a threshold τ[2] for which:

$$P_Y(y=\tau) > P_X(x=\tau)$$
$$\ln\left(\frac{P_Y(\tau)}{P_X(\tau)}\right) = \ln\left(P_Y(\tau)\right) - \ln\left(P_X(\tau)\right) > 0. \tag{7.23}$$
$$\tag{7.24}$$

Inserting the proper distributions with the parameters estimated from the previous formula we have to solve the following equation in τ:

$$\ln(\lambda) - \lambda\tau + 0.5\ln\left(2\pi\sigma^2\right) + \frac{(\tau)^2}{2\sigma^2} = 0$$
$$\tau_{1,2} = \sigma^2\left(\lambda \pm \sqrt{\lambda^2 - \frac{1}{\sigma^2}\ln(2\pi\sigma^2\lambda^2)}\right), \tag{7.25}$$
$$\tau_0 = \min\left\{\ln\left(\frac{P_Y(\tau)}{P_X(\tau)}\right)\right\} = \lambda\sigma^2 \tag{7.26}$$

so that, in this exponential case, the solution reduce to a second order equation. The parameter $\tau_{1,2,0}$ can be interpreted looking at figure 7.10 as the intersections of the log-ratio with the x-axis and its minimum respectively.

In addition to its simple treatment, the Gaussian exponential mixture has a straightforward application in the masking of a real space map. The distribution of foreground template pixels is more similar to an exponential. Note that, of course, it is *not* strictly exponential, but this simple parametrization can still be useful to better account for tail contributions than simple Gaussian modeling.

The data used are shown in figure 7.11, it is a low resolution reconstruction of a 100 Ghz survey (Temperature), assuming a $30'$ beam and nside=128, based on the Planck foregrounds templates and a random CMB realization.

With the method explained in the previous section, we can retrieve 3 notable values: the two roots $\tau_{1,2}$, representing the bounds of the interval where the Gaussian term dominates, and τ_0 representing the minimum of the log ratio of the inferred distribution.

[2]note that in the previous section we called the threshold λ, we change here to avoid confusion with the exponential scale parameter

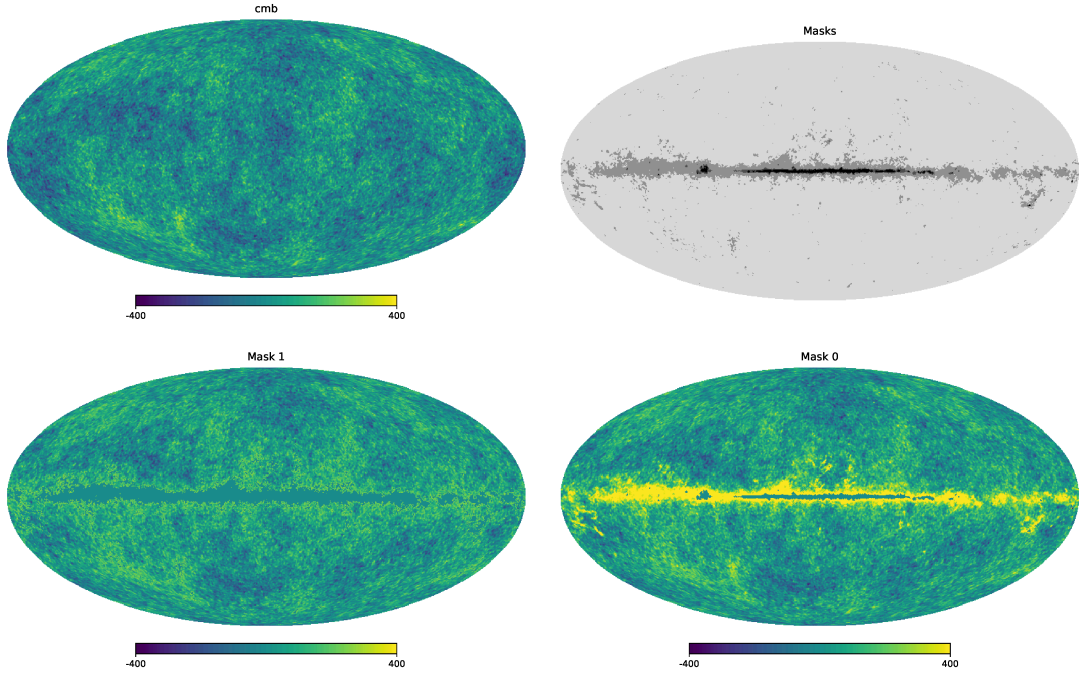

FIGURE 7.12: Up left: Input CMB. Up right: masks. Low left: simulated map with the conservative mask. Low right: simulated map with the "medium mask"

Fore these data we found:

$$\tau_1 = 263\ \mu K$$
$$\tau_0 = 1489\ \mu K$$
$$\tau_2 = 2715\ \mu K.$$

Being in real space, we cannot exploit sparsity, as we do when working with a needlet basis. Therefore these values can only be used to build a mask, but not to recover the foreground template. However, it is interesting to notice that pixels with values $> \tau_2$, *i.e.* the 0.4% of the sky, represent the 75% of the foreground template energy (defined as $\sum_i T_i^2$ where T is the template and i run over the pixels). This is a promising aspect in view of a needlet template fitting application.

Figure 7.12 shows the results of this masking procedure: the original CMB field, the masks obtained from the three threshold.

7.5 Conclusions

The search for an optimal solution of the component separation problem in microwave surveys is one of the most challenging field of research in observational Cosmology.

Foreground systematics represent nowadays one of the main, if not the main limitation in CMB analysis. This is especially true in the pursuit for polarization B-modes anisotropies from primordial gravitational waves, that are expected to produce a very low signal. In the last decades, during which CMB surveys played a major role in cosmological observations, a great number of different methodologies was developed to disentagle the various emissions. These techniques exploit peculiar features of the components to identify and then separate them, or to reconstruct a particular source of interest.

In this work we present a preliminary investigation on a technique relying on the assumption that the foreground signal is "sparse" in a proper representation, *i.e.* the majority of the signal is concentrated in few expansion coefficients. The spherical needlet frame enjoys the right characteristic to induce sparseness in the representation of coherent signals: it provide a space-frequency representation with remarkable localization properties both in real and frequency space; moreover, it has proven to be perfectly suited to describe random fields on the sphere due to the good uncorrelation properties of its coefficients. So that we expect that foreground signal information collapses in few coefficients while the CMB's does not.

We check this hypothesis on a wide set of simulations of the microwave sky, and we found promising confirmation to it. We indeed observe that the foreground templates can be faithfully represented with just a few modes, when working in the needlet domain. On the other hand, the CMB signal is actually spread among all the coefficients, as expected from its stochastic nature.

A very natural way to deal with sparse signals in space-frequency representations is a shrinkage technique called wavelet thresholding. We analyze the potential of a needlet thresholding algorithm to identify and then remove the foreground contributions. This method does not rely on multi-channel information, so that it could be a useful tool in surveys with few observational frequencies.

We found that, in the small scale regime, this technique is quite effective in isolating the foreground components, that actually collapse into few coefficients. On the other hand, on the largest scales (approximately below $\ell = 20$ in our tests), there are too few modes (*i.e.* uncorrelated basis functions, in our representation) to efficiently disentangle the different components. We then investigate the synergy of this technique with other methods, and we found that the performance of a simple ILC algorithm significantly improves when the map is pre-cleaned with a needlet thresholding.

We also provide a complete Bayesian treatment of our techniques, in order to gain better insight and include the implementation of thresholding as a pre-cleaning procedure within a general, multi-frequency, component separation pipeline. Furthermore, we search for different threshold selection criteria, tailored around our specific problem, based on the identification and minimization of non-Gaussian and anisotropic biases induced by foreground emission.

In conclusion, in this preliminary investigation we found that needlet thresholding is a promising tool for component separation, especially in surveys with limited frequency coverage. Further developments are still ongoing for the implementation of this technique, exploring the possibility of combining it with different multi-frequency component separation approaches. Besides the thresholding-ILC combination, discussed here, another interesting possibility could consist in using thresholding to extract foreground templates at specific frequencies, followed by template-fitting/marginalization.

The needlet frame allows for a natural and straightforward extension to polarization surveys and partial sky maps, which will also be explored. A natural and potentially interesting field of application is indeed its use on current generation ground-based B-mode surveys, indeed characterized by small frequency coverage.

Finally, we briefly comment on the similarities and differences with the GMCA pipeline (Bobin et al. 2008), which also is characterized by a built-in thresholding procedure. As already pointed out earlier, one important difference is that our approach is not assuming sparsity of the CMB signal. Besides being more realistic, this has also the important consequence of fully preserving the isotropy of the CMB signal, if our thresholding is *e.g.* integrated within a template-fitting procedure. Moreover, GMCA aims at recursively finding the component of the mixing matrix, while the current method implicitly marginalize over them, aiming, as we already stressed, at the construction of a general, single frequency pre-analysis procedure, that can be freely and flexibly combined with a variety of other pipelines.

Chapter 8

Conclusions

The pursuit for non-Gaussian signatures in the CMB has received great attention within the cosmological community over the last decades. The detection and the measurement of such a signal would provide an unprecedented insight into the Physics of the Early Universe. The most promising theory to solve the outstanding questions related to initial conditions in the Standard Cosmological Model is the *Theory of Inflation*. However, information obtained under the Gaussian assumption, *i.e.* from the power spectrum, has little power to discriminate between the huge number of different Inflationary scenarios proposed in the literature. On the contrary, primordial non-Gaussianity is expected to be in most cases a small, but non-negligible and strongly model-dependent effect. For this reason, primordial non-Gaussianity is expected to be one of the most informative fingerprints of the origin of structure in the Universe, probing physics at extremely high energy scales inaccessible to laboratory experiments.

Actual calculations in the Inflationary framework show that the statistic most sensitive to the departure from Gaussianity is the tree-point function, or rather its harmonic space counterpart, the primordial *Bispectrum*. The CMB Bispectrum, being linearly linked to the primordial field, provides an almost direct probe of PNG. Unfortunately, due to the high numerical complexity the full Bispectrum cannot be computed brute force. Moreover, single bispectrum configurations have very low signal-to-noise and carry therefore negligible statistical weight. The general approach to extract relevant information from the bispectrum is therefore that of fitting specific, theoretical motivated templates to the data, in search for peculiar NG features. To date, a large number of different Bispectrum shapes has been tested on the *Planck* data-set, but no significant hints of a primordial NG signal has been detected. On the other hand, very tight

constraints have been obtained on many primordial NG scenarios.

In the first part of this thesis, we present the analysis of several scale-dependent bispectrum shapes, both using the WMAP 9 year and the *Planck* data-sets. Two of the three models considered have been constrained for the first time in this work, and none of them has ever been tested on the *Planck* data-set.

Their analysis required the development of new, specifically tailored estimators. More in detail, we develop a complete set of tools for CMB forecasting, simulation and estimation of primordial running bispectra, arising from a variety of curvaton and single-field (DBI) models of Inflation. Constraining the running of the primordial NG parameter $f_{\rm NL}$ can provide valuable extra-information on the Physics of Inflation, allowing for better discrimination between different scenarios. The constraints on the running parameter $n_{\rm NG}$ that we obtain from the WMAP analysis are: $-0.6 < n_{\rm NG} < 1.4$, $-0.3 < n_{\rm NG} < 1.2$, $-1.1 < n_{\rm NG} < 0.7$ for the one-field curvaton, two-field curvaton and DBI scenarios, respectively. In the analysis of *Planck* data we have to improve our estimator to deal with secondary biases (ISW-lensing) and to prevent the possible insurgence of statistical artifacts due to the lower significance of the estimated primordial NG amplitudes. We resort to different prior choices, implementing an uninformative Jeffreys prior, and different approaches, as likelihood profiling. We run our SD estimation pipeline on the *Planck* temperature maps and simulations, while we also test the standard local, equilateral and orthogonal templates with the KSW estimator (that is a byproduct of our pipeline) on both temperature and E polarization maps. We refer to the final *Planck* release for the results on the scale invariant templates. We do not find evidence in favour of scale-dependent models for any of the templates tested. Furthermore, we found that at the current level of detection it is not meaningful to put constraints on the additional scale-dependence parameter n_{NG}. We conclude observing that in this analysis we reach the maximum theoretical constraining power on direct bispectrum estimation for these templates, so that future test shall rely on different probes.

In the second part of this thesis, we still deal with the study of non-Gaussian and anisotropic signals in microwave sky maps, but with a totally different aim, namely that of studying and separating foreground components. One of the major limitations in modern CMB analysis indeed comes from foregrounds systematics. The detection and minimization of the foreground contribution ("component separation") in microwave surveys is therefore nowadays one of the most active fields in observational cosmology. Component separation algorithms exploit characteristic features of foreground emission

to disentangle them from the background radiation. Since the foreground is strongly non-Gaussian, NG signatures represent an ideal gauge to unearth the spurious components in microwave surveys. Following this rationale, we develop a technique relying on the hypothesis that the foreground are "sparse" in a proper representation. Sparseness can be interpreted in the general framework of non-Gaussianity. In particular it is usually formalized resorting to the Laplace distribution, or other similar leptokurtic, "heavy-tailed" distributions. A signal is said to be "sparse" in a given basis if its information appear compressed in very few basis elements with high S/N. We found that spherical needlets represent an ideal frame for our purposes, since it provides sparse representation of coherent signals while at the same time it enjoys a number of mathematical properties very well suited to threat Gaussian random fields. Therefore, our assumption is that foreground signal information, in virtue of their coherence, anisotropy and non-Gaussianity, collapses in few coefficients, while the CMB component is spread among all the coefficients.

We verify this assumption with a number of tests on different simulations of microwave sky maps. All the analyses performed provide confirmation to our hypothesis. Therefore, the following step has been the implementation of a thresholding technique, a method widely used in wavelet regression of sparse signals. This method has the advantage of not relying on multi-channel information, so that it can be a useful tool in surveys with few observational bands, as is the case in current generation, ground-based surveys. On the single map, it provides good results at high spatial frequencies, whereas at low frequencies the low number of modes makes a sparse representation impossible.

Therefore, we provide a complete Bayesian treatment of needlet thresholding in CMB surveys, in order to check for possible synergies with other, multi-channel techniques. We find that thresholding is indeed very effective in improving the performance of a simple ILC algorithm, as a frequency-by-frequency pre-cleaning procedure. We then study ad-hoc solutions for the selection of the optimal threshold for our problem, considering the minimization of typical foreground signatures as non-Gaussianity and anisotropy.

We judge the results of this preliminary investigation very promising; future prospects of this work will be the development of a complete component separation pipeline combining thresholding and multi-frequency separation algorithms, as well as the extension to polarization surveys and partial sky coverage.

Appendix A

Useful Formulae of General Relativity

In general relativity in the partial derivative operation we have to take into account the curvature of the space time. So the *covariant derivative* operator is introduced:

$$\nabla_\mu V^\mu = \partial_\mu V^\mu + \Gamma^\nu_{\mu\lambda} V^\lambda \tag{A.1}$$

where V^μ is a generic 4-vector and $\Gamma^\nu_{\mu\lambda}$ is the connections coefficients, also known as *Christoffel symbol*. The Christoffel symbol describe geodesic in non-trivial coordinates system.

A geodesic is the generalization in a curved space of a straight line.

The Christoffel symbol is defined from the metric as:

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\mu\nu} \left[\partial_\alpha g_{\beta\nu} + \partial_\beta g_{\alpha\nu} - \partial_\nu g_{\alpha\beta} \right]. \tag{A.2}$$

The *geodesic equation*, expressing the Newton law with no forces in general relativity is :

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda} = 0. \tag{A.3}$$

In a curved space-time, the parallel transport of a vector V^σ on a closed loop defined by two infinitesimal vector A^μ and B^μ results in a modification δV^ρon the vector. This change is quantified by the curvature tensor also know as *Riemann tensor*, that is:

$$\delta V^\rho = R^\rho_{\sigma\mu\nu} A^\mu B^\nu V^\sigma. \tag{A.4}$$

The definition of the Riemann tensor is

$$R^{\rho}_{\sigma\mu\nu} = \partial_{\mu}\Gamma^{\rho}_{\nu\sigma} + \partial_{\nu}\Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} + \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma} \tag{A.5}$$

The contraction of the Riemann tensor give the *Ricci tensor* $R_{\mu}\nu$, its definition follow from (A.5):

$$R^{\mu\nu} = \partial_{\lambda}\Gamma^{\lambda}_{\mu\nu} + \partial_{\nu}\Gamma^{\lambda}_{\mu\lambda} + \Gamma^{\lambda}_{\lambda\beta}\Gamma^{\beta}_{\mu\nu} + \Gamma^{\lambda}_{\beta\nu}\Gamma^{\beta}_{\mu\lambda}. \tag{A.6}$$

The track of the Ricci tensor define the Ricci scalar:

$$\begin{aligned} R =& g^{\mu\nu}R_{\mu\nu} \\ \nabla^{\mu}R_{\mu\nu} =& \frac{1}{2}\nabla_{\nu}R \end{aligned} \tag{A.7}$$

The Einstein equation is:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} = G_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{A.8}$$

were $G_{\mu\nu}$ is the *Einstein tensor*.

Appendix B

Special Functions

B.1 Spherical harmonics

Spherical harmonics are the angular portion of a set of solutions to Laplacian. Usually we refer to Laplace's spherical harmonics Y_ℓ^m that are a specific set of spherical harmonics forming an orthogonal system. They are eigenfunction of the equation

$$\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]Y_\ell^m(\theta,\phi)=-\ell(\ell+1)Y_\ell^m(\theta,\phi). \tag{B.1}$$

We used them to describe a distribution on a spherical surface with harmonics behavior; this is the analogue of the Fourier decomposition in flat space. The orthogonal relation states:

$$\int \mathrm{d}\Omega\,\overline{Y}_\ell^m\Omega Y_{\ell'}^{m'}(\Omega)=\delta_{\ell\ell'}\delta_{mm'}. \tag{B.2}$$

B.1.1 Legendre Polynomials

The Legendre polynomials are particular solutions of Legendre's equations the is useful to solve the Laplace's equation in spherical coordinates. For our aim is sufficient to know their orthogonality relation:

$$\int_{-1}^{1}\mathcal{P}_\ell(x)\mathcal{P}_{\ell'}(x)=\delta_{\ell\ell'}\frac{2}{2\ell+1}. \tag{B.3}$$

They can be expressed by a sum of products of spherical harmonics:

$$\mathcal{P}_\ell(\hat{x}\cdot\hat{x}')=\frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell}Y_\ell^m(\hat{x})\overline{Y}_\ell^m(\hat{x'}). \tag{B.4}$$

B.2 Spherical Bessel Functions

The spherical Bessel function $j_\ell(x)$ satisfy the differential equation:

$$\frac{\mathrm{d}^2 j_\ell(x)}{\mathrm{d}x^2} + \frac{2}{x}\frac{\mathrm{d}j_\ell(x)}{\mathrm{d}x} + \left[1 - \frac{\ell(\ell-1)}{x^2}\right] j_\ell(x) = 0. \tag{B.5}$$

They are crucial in the study of the CMB because, leading the radial evolution, they project the inhomogeneities at last scattering into the actual sky.
These function obey to the so called closure relation:

$$\int_0^\infty \mathrm{d}x\, x^2 j_\ell(ax) j_\ell(bx) = \frac{\pi}{2a^2}\delta_D(a-b) \tag{B.6}$$

They are related to Legendre polynomials by expressions:

$$\frac{1}{2}\int_{-1}^{1} \mathrm{d}\mu \mathcal{P}_\ell(\mu) e^{ix\mu} = \frac{j_\ell(x)}{(-i)^\ell} \tag{B.7}$$

The inversion of this equations gives a useful expansion for Fourier basis functions:

$$e^{i\vec{k}\cdot\vec{x}} = \sum_{\ell=0}^{\infty} i^\ell (2\ell+1) j_\ell(kx) \mathcal{P}_\ell(\hat{k}\cdot\hat{x}). \tag{B.8}$$

Replacing formula (B.4) in this equation we obtain the spherical wave expansion of vector plane wave used in chapter 2.2 equation (2.30):

$$e^{i\vec{k}\cdot\vec{x}} = (4\pi)\sum_{\ell=0}^{\infty} i^\ell j_\ell(kr) \sum_{m=-\ell}^{\ell} Y_\ell^m(\hat{x})\overline{Y}_\ell^m(\hat{k}) \tag{B.9}$$

Another important relation allows to express the integral containing squared Bessel functions in terms of Euler Gamma functions:

$$\int_0^\infty \mathrm{d}x\, x^{n-2} j_\ell^2(x) = 2^{n-4}\pi \frac{\Gamma(\ell + \frac{n}{2} - \frac{1}{2})\Gamma(3-n)}{\Gamma(\ell + \frac{5}{2} - \frac{n}{2})\Gamma^2(2 - \frac{n}{2})} \tag{B.10}$$

B.3 Wigner 3-j symbols

The Wigner 3-j symbols are related to the Clebsh–Gordan coefficients which describe coupling of two angular momenta, they are defined as a 2×3 matrix

$$\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

more specifically a symbol of this kind describes three angular momenta forming a triangle, for this reason they satisfies the triangle conditions $L_1 + L_2 + L_3 = 0$, from this derives:

$$|\ell_i - \ell_j| \leq \ell_k \leq \ell_i + \ell_j$$
$$m_1 + m_2 + m_3 = 0$$
$$\ell_1 + \ell_2 + \ell_3 = even.$$

The special case with $\ell_1 = \ell_2$ and $m_3 = 0$ is useful in the calculation of the variance of the bispectrum:

$$\begin{pmatrix} \ell & \ell & \ell' \\ m & -m & 0 \end{pmatrix} = \frac{(-1)^\ell}{\sqrt{2\ell+1}} \delta_{\ell' 0} \tag{B.11}$$

Other important proprieties are the orthogonality relations we used in the calculation of angle average bispectrum

$$\sum_{\ell_3 m_3} (2\ell_3 + 1) \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1' & m_2' & m_3 \end{pmatrix} = \delta_{m_1 m_1'} \delta_{m_2 m_2'} \tag{B.12}$$

$$\sum_{m_1 m_2} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3' \\ m_1 & m_2 & m_3' \end{pmatrix} = \delta_{m_1 m_1'} \delta_{m_2 m_2'} \tag{B.13}$$

$$\sum m_1 m_2 m_3 \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}^2 = 1. \tag{B.14}$$

An analytical expression for the symbol $\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}$ (Verde et al. 2000), that we used in the computations of Fisher matrix is:

$$\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} = (-1)^L \sqrt{\frac{(L+2\ell_1)!(L+2\ell_2)!(L+2\ell_3)!}{(L+1)!}} \frac{(L/2)!}{(L/2-\ell_1)!(L/2-\ell_1)!(L/2-\ell_1)!} \tag{B.15}$$

where $L = \ell_1 + \ell_2 + \ell_3$. Since problems arise in the numerical evaluation of the factorial for high ℓ, this formula can be replaced appealing to the Gosper version of the Stirling factorial approximation (Giovi et al. 2003):

$$n! = \left(\frac{n}{e}\right)^n \sqrt{\pi \left(2n + \frac{1}{3}\right)} \tag{B.16}$$

following this approximation the Wigner symbol can be rewrote in a computational simpler form:

$$\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \simeq \left(-\frac{L}{L+1}\right)^{L/2} \frac{1}{(6L+7)^{1/4}} \sqrt{\frac{3e(3L+1)}{\pi(L+1)}} \prod_{i=1}^{3} \frac{(6L-12\ell_i+1)^{1/4})}{\sqrt{3L-6\ell_i+1}} \tag{B.17}$$

Bibliography

Ade, P. A. R., Aghanim, N., Alves, M. I. R., et al. 2014a, A&A, 571, A13

Ade, P. A. R., Aghanim, N., Alves, M. I. R., et al. 2016, A&A, 594, A25

Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., et al. 2014b, A&A, 571, A14

Albrecht, A. & Steinhardt, P. J. 1982, Phys. Rev. Lett., 48, 1220

Antoine, J.-P., Roşca, D., & Vandergheynst, P. 2010, Applied and Computational Harmonic Analysis, 28, 189 , special Issue on Continuous Wavelet Transform in Memory of Jean Morlet, Part I

Baldi, P., Kerkyacharian, G., Marinucci, D., & Picard, D. 2006, ArXiv Mathematics e-prints [math/0606599]

Baldi, P., Kerkyacharian, G., Marinucci, D., & Picard, D. 2007, ArXiv e-prints [[arXiv]0706.4169]

Baldi, P., Kerkyacharian, G., Marinucci, D., & Picard, D. 2009, Ann. Statist., 37, 1150

Bardeen, J. M., Steinhardt, P. J., & Turner, M. S. 1983, Phys. Rev. D, 28, 679

Bartolo, N., Fasiello, M., Matarrese, S., & Riotto, A. 2010, JCAP, 1012, 026

Basak, S. & Delabrouille, J. 2012, MNRAS, 419, 1163

Basak, S. & Delabrouille, J. 2013, MNRAS, 435, 18

Baumann, D. 2011, in Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009, 523–686

Baumann, D. & Peiris, H. V. 2009, Adv. Sci. Lett., 2, 105

Becker, A. & Huterer, D. 2012, Phys. Rev. Lett., 109, 121302

Bennett, C. L., Bay, M., Halpern, M., et al. 2003, ApJ, 583, 1

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20

Biagetti, M., Perrier, H., Riotto, A., & Desjacques, V. 2013, Phys.Rev., D87, 063521

Bierman, E. M., Matsumura, T., Dowell, C. D., et al. 2011, ApJ, 741, 81

Bobin, J., Moudden, Y., Starck, J. L., Fadili, J., & Aghanim, N. 2008, Stat. Meth., 5, 307

Brandenberger, R., Kahn, R., & Press, W. H. 1983, Phys. Rev. D, 28, 1809

Bucher, M., Racine, B., & van Tent, B. 2016, JCAP, 1605, 055

Bucher, M., van Tent, B., & Carvalho, C. S. 2010, MNRAS, 407, 2193

Byrnes, C., Gerstenlauer, M., Nurmi, S., Tasinato, G., & Wands, D. 2010, Journal of Cosmology and Astroparticle Physics, 2010, 004

Cabella, P., Hansen, F., Marinucci, D., Pagano, D., & Vittorio, N. 2004, Phys. Rev. D, 69, 063007

Cardoso, J.-F., Martin, M., Delabrouille, J., Betoule, M., & Patanchon, G. 2008, ArXiv e-prints [[arXiv]0803.1814]

Chen, X. 2005, Phys. Rev. D, 72, 123518

Chen, X. 2010, Advances in Astronomy, 2010, 638979

Chen, X., Huang, M.-x., Kachru, S., & Shiu, G. 2007, JCAP, 0701, 002

Cooray, A. 2005, Mon. Not. Roy. Astron. Soc., 363, 1049

Cooray, A. & Melchiorri, A. 2006, J. Cosmology Astropart. Phys., 1, 018

CORE Collaboration, Finelli, F., Bucher, M., et al. 2016, ArXiv e-prints [[arXiv]1612.08270]

Creminelli, P., Nicolis, A., Senatore, L., Tegmark, M., & Zaldarriaga, M. 2006, Journal of Cosmology and Astroparticle Physics, 2006, 004

Creminelli, P., Senatore, L., & Zaldarriaga, M. 2007, Journal of Cosmology and Astroparticle Physics, 3, 019

Delabrouille, J. & Cardoso, J.-F. 2009, Diffuse Source Separation in CMB Observations, ed. V. J. Martinez, E. Saar, E. M. Gonzales, & M. J. Pons-Borderia (Berlin, Heidelberg: Springer Berlin Heidelberg), 159–205

Dodelson, S. 2003, Modern Cosmology (Amsterdam: Academic Press)

Donoho, D. & Johnstone, J. 1994, Biometrika, 81, 425

Emami, R., Dimastrogiovanni, E., Chluba, J., & Kamionkowski, M. 2015, Phys. Rev., D91, 123531

Eriksen, H. K., Jewell, J. B., Dickinson, C., et al. 2008, The Astrophysical Journal, 676, 10

Eriksen, H. K., O'Dwyer, I. J., Jewell, J. B., et al. 2004, ApJS, 155, 227

Falk, T., Rangarajan, R., & Srednicki, M. 1993, ApJ, 403, L1

Fergusson, J. R., Liguori, M., & Shellard, E. P. S. 2010, Phys. Rev. D, 82, 023502

Fergusson, J. R., Liguori, M., & Shellard, E. P. S. 2012, J. Cosmology Astropart. Phys., 12, 032

Fernández-Cobos, R., Vielva, P., Barreiro, R. B., & Martínez-González, E. 2012, MNRAS, 420, 2162

Gangui, A., Lucchin, F., Matarrese, S., & Mollerach, S. 1994, ApJ, 430, 447

Gangui, A. & Martin, J. 2000, MNRAS, 313, 323

Génova-Santos, R., Rubiño-Martín, J. A., Peláez-Santos, A., et al. 2017, MNRAS, 464, 4107

Génova-Santos, R., Rubiño-Martín, J. A., Rebolo, R., et al. 2015, MNRAS, 452, 4169

Giovi, F., Baccigalupi, C., & Perrotta, F. 2003, Phys. Rev. D, 68, 123002

Gruetjen, H. F., Fergusson, J. R., Liguori, M., & Shellard, E. P. S. 2017, Phys. Rev., D95, 043532

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Guth, A. H. & Pi, S.-Y. 1982, Phys. Rev. Lett., 49, 1110

Hansen, F. K., Banday, A. J., Eriksen, H. K., Gorski, K. M., & Lilje, P. B. 2006, Astrophys. J., 648, 784

Hanson, D., Smith, K. M., Challinor, A., & Liguori, M. 2009, Phys. Rev. D, 80, 083004

Hasselfield, M. et al. 2013, Astrophys. J. Suppl., 209, 17

Hilton, M. et al. 2013, Mon. Not. Roy. Astron. Soc., 435, 3469

Hu, W. 2000, Phys. Rev. D, 62, 043007

Jaynes, E. T. 1968, Transactions On Systems Science and Cybernetics, 4

Jeffreys, H. 1946, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 186, 453

Kass, R. E. & Wasserman, L. 1996, Journal of the American Statistical Association, 91, 1343

Khatri, R. & Sunyaev, R. 2015, J. Cosmology Astropart. Phys., 9, 026

Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330

Komatsu, E. & Spergel, D. N. 2001, Phys. Rev. D, 63, 063002

Komatsu, E., Spergel, D. N., & Wandelt, B. D. 2005, The Astrophysical Journal, 634, 14

Leach, S. M., Cardoso, J.-F., Baccigalupi, C., et al. 2008, A&A, 491, 597

Lewis, A., Challinor, A., & Hanson, D. 2011, J. Cosmology Astropart. Phys., 3, 018

Liguori, M., Sefusatti, E., Fergusson, J. R., & Shellard, E. P. S. 2010, Advances in Astronomy, 2010, 980523

Linde, A. 1982, Physics Letters B, 108, 389

Mangilli, A. & Verde, L. 2009, Phys. Rev. D, 80, 123007

Mangilli, A., Wandelt, B., Elsner, F., & Liguori, M. 2013, A&A, 555, A82

Marinucci, D. & Peccati, G. 2011, Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications, London Mathematical Society Lecture Note Series (Cambridge University Press)

Marinucci, D., Pietrobon, D., Balbi, A., et al. 2008, MNRAS, 383, 539

McEwen, J. D., Hobson, M. P., Lasenby, A. N., & Mortlock, D. J. 2006, Mon. Not. Roy. Astron. Soc., 371, L50

McEwen, J. D., Vielva, P., Hobson, M. P., Martinez-Gonzalez, E., & Lasenby, A. N. 2007, Mon. Not. Roy. Astron. Soc., 376, 1211

McEwen, J. D., Vielva, P., Wiaux, Y., et al. 2007, Journal of Fourier Analysis and Applications, 13, 495

Moudden, Y., Cardoso, J.-F., Starck, J.-L., & Delabrouille, J. 2005, EURASIP Journal on Applied Signal Processing, 2005, 484606

Munshi, D. & Heavens, A. 2010, Monthly Notices of the Royal Astronomical Society, 401, 2406

Narcowich, F. J., Petrushev, P., & Ward, J. D. 2006, SIAM J. Math. Analysis, 38, 574

Oppizzi, F., Liguori, M., Renzi, A., Arroja, F., & Bartolo, N. 2018a, J. Cosmology Astropart. Phys., 5, 045

Oppizzi, F., Liguori, M., Renzi, A., et al. 2018b, 1

Peskin, M. E. & Schroeder, D. V. 1995, An Introduction to quantum field theory (Reading, USA: Addison-Wesley)

Pietrobon, D., Balbi, A., & Marinucci, D. 2006, Phys. Rev. D, 74, 043524

Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2016a, A&A, 594, A10

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014a, A&A, 565, A103

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014b, A&A, 571, A24

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016b, A&A, 594, A17

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016c, A&A, 594, A26

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016d, A&A, 594, A28

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018a, ArXiv e-prints [[arXiv]1807.06212]

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018b, ArXiv e-prints [[arXiv]1807.06207]

Planck Collaboration, Akrami, Y., Argüeso, F., et al. 2018c, ArXiv e-prints [[arXiv]1807.06206]

Planck Collaboration, Akrami, Y., Arroja, F., et al. 2018d, ArXiv e-prints [[arXiv]1807.06205]

Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2018e, ArXiv e-prints [[arXiv]1807.06208]

Raccanelli, A., Doré, O., & Dalal, N. 2015, J. Cosmology Astropart. Phys., 8, 034

Ravenni, A., Liguori, M., Bartolo, N., & Shiraishi, M. 2017, J. Cosmology Astropart. Phys., 9, 042

Sanz, J. L., Herranz, D., Lopez-Caniego, M., & Argueso, F. 2006, in 14th European Signal Processing Conference (EUSPICO 2006) Florence, Italy, September 4-8, 2006

Sefusatti, E., Liguori, M., Yadav, A. P., Jackson, M. G., & Pajer, E. 2009, Journal of Cosmology and Astroparticle Physics, 2009, 022

Sefusatti, E., Liguori, M., Yadav, A. P. S., Jackson, M. G., & Pajer, E. 2009, J. Cosmology Astropart. Phys., 12, 022

Senatore, L., Smith, K. M., & Zaldarriaga, M. 2010, J. Cosmology Astropart. Phys., 1, 028

Serra, P. & Cooray, A. 2008, Phys. Rev. D, 77, 107305

Shandera, S., Dalal, N., & Huterer, D. 2011, J. Cosmology Astropart. Phys., 3, 017

Smith, K. M. & Zaldarriaga, M. 2011, Monthly Notices of the Royal Astronomical Society, 417, 2

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJ, 396, L1

Smoot, G. F., Gorenstein, M. V., & Muller, R. A. 1977, Phys. Rev. Lett., 39, 898

Starck, J.-L. & Murtagh, F. 2006, Astronomical Image and Data Analysis (Springer-Verlag Berlin Heidelberg)

Starobinsky, A. 1982, Physics Letters B, 117, 175

Struve, F. G. W. 1847, Etudes d'Astronomie Stellaire: Sur la voie lactee et sur la distance des etoiles fixes

Taruya, A., Koyama, K., & Matsubara, T. 2008, Phys. Rev., D78, 123534

Thorne, B., Dunkley, J., Alonso, D., & Naess, S. 2017, Mon. Not. Roy. Astron. Soc., 469, 2821

Trumpler, R. J. 1930a, PASP, 42, 214

Trumpler, R. J. 1930b, Lick Observatory Bulletin, 14, 154

Trumpler, R. J. 1930c, PASP, 42, 267

Tzavara, E. & van Tent, B. 2013, J. Cosmology Astropart. Phys., 6, 001

Vansyngel, F., Wandelt, B. D., Cardoso, J.-F., & Benabed, K. 2016, A&A, 588, A113

Verde, L., Heavens, A. F., & Matarrese, S. 2000, MNRAS, 318, 584

Vielva, P., Martínez-González, E., Barreiro, R. B., Sanz, J. L., & Cayón, L. 2004, ApJ, 609, 22

Wang, L. & Kamionkowski, M. 2000, Phys. Rev. D, 61, 063504

Wasserman, L. 2006, All of Nonparametric Statistics (Springer-Verlag New York)

Whittet, D. C. B. 1992, Dust in the galactic environment