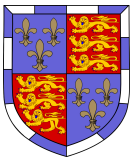
UNIVERSITY OF
CAMBRIDGE

Connecting Inflation to Observations Through the Bispectrum

Philip Clarke



St. John's

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the statement of contributions and specified in the text. It is not substantially the same as any that I have submitted, or am concurrently submitting, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

Philip Clarke
September, 2021

Statement of contributions

In this thesis I describe the research I performed during my PhD studies at the University of Cambridge, supervised by Professor Paul Shellard. Chapters 3, 4 and 5 describe original work, while chapters 1 and 2 are entirely introductory, and mostly follow standard textbook material, collected and written by myself. Some of chapter 2 is edited material from [1], which was co-authored with Paul Shellard.

Chapters 3 and 4 describe original research I performed in collaboration with Paul Shellard, and mostly consist of edited material that was originally published in [1]. I have updated the text and figures in chapter 3 to include improved basis sets that I have developed since the publication of [1]. Figure 4.2 is also new, and I am grateful to the anonymous referee of [1] for their comments which suggested it. Most of the results presented in these two chapters were obtained using the PRIMODAL code, which I designed and wrote, which implements the methods described in this work, and is validated in these sections.

Chapter 5 describes work done in collaboration with Wuhyun Sohn, James Fergusson and Paul Shellard. This work will be published in [2]. The methods that will be described in [2] have been implemented in the CMB-BEST code, which was designed and implemented by Wuhyun Sohn. My main contribution to [2] is described in chapter 5.

Abstract

Connecting Inflation to Observations Through the Bispectrum

Philip Clarke

Numerically calculating the full primordial bispectrum predicted by a model of inflation and comparing it to what we see in the CMB sky is very computationally intensive, necessitating layers of approximations and limiting the models which can be constrained. The inherent separability of the tree level in-in formalism provides a means by which to obviate some of these difficulties.

To exploit this property, one can expand in separable basis functions. The practicality of this method is then determined by the descriptive power of the basis chosen, i.e. by the range of scenarios for which that basis provides a convergent representation of the bispectrum. The central difficulty encountered in obtaining fast convergence is the effect of dominant non-physical k-configurations. A secondary difficulty encountered is accurately including the early-time contributions to the higher-order coefficients (which are necessary to capture feature effects, such as resonance non-Gaussianity).

In this thesis we develop this separable, template-free approach into a practical and efficient numerical methodology which can be applied to a much wider and more complicated range of bispectrum phenomenology than previous analyses. This is an important step forward towards observational pipelines which can fully exploit the information contained in the primordial bispectrum to directly confront specific models of inflation. We use our implementation of this pipeline to obtain a constraint on DBI inflation, and validate our implementation by comparing this constraint to an equivalent one obtained by the *Planck* collaboration.

Acknowledgements

Write this!

My acknowledgements...

We make particular use of [3–6] throughout.

The calculations described in this thesis were implemented using the Python scientific computing ecosystem, especially Numpy [7], Scipy [8], and Matplotlib [9].

I am very grateful to lots of people for reasons.

For the generosity of their financial support, I am very grateful to the Robert Gardiner Scholarship, the Cambridge Trust European Scholarship, St. John's College, and the Cambridge Philosophical Society. I also thank the UK Science and Technology Facilities Council, and the people of the United Kingdom.

CSD3.

I am grateful to my college for reasons.

Contents

Chapter 1

Introduction I: Cosmology

1.1 The standard cosmological history

1.1.1 Fundamentals

When we look out into the sky, we see light that has traveled for a long time to get to us. In that light we see images of the universe as it was when the light left its source. We see the moon as it was a second ago, or the Andromeda Galaxy as it was millions of years ago. To what extreme can we push this? How far back in time can we see? As we look further away, and further back in time, we see a universe that looks nothing like the one we know today. We see a universe that was extremely hot and dense, but also extremely uniform. Today we have distinct galaxies and the empty spaces in between, but looking far enough back we see a universe without that structure, instead simply uniformly hot and dense everywhere.

But this uniformity wasn't perfect. There were patches that were very slightly more dense than the average, and others that were very slightly less dense. These non-uniformities provided the seeds that would form structure—the over-dense regions had a slightly higher gravitational pull on their surroundings, and caused the nearby matter to begin to fall in. This process would eventually result in the structure we see in the universe today: the stars, the galaxies, and all the complexity on our own planet.

The precise details of this story have been fleshed out thanks to the work of many people. It is incredible that we can say so much about the radically different epochs of the universe that came before our own, between this hot, dense initial state and today. More work remains to be done, however—one

aspect we can question is the origin of the non-uniformities that provided the seeds of structure. What was the physics that generated them? The goal of the work described in this thesis is to develop methods to connect this physics to the sky we see today; more precisely, we aim to use the light from the early universe, the Cosmic Microwave Background (CMB), to constrain physical theories of the very early universe (the epoch of exponential expansion known as inflation) using a particular statistical description, known as the bispectrum. We aim to mature the connection between models of inflation and their predicted bispectra, so we can better extract the fundamental physics waiting to be found in the universe's youth.

We begin with a brief discussion of the fundamentals of cosmology. General relativity is the theory which describes how the evolution of the universe depends on its matter and energy content. It describes the path of light as a geodesic, the generalisation of a straight line to curved spacetime. The equations that we derive from this theory (the Einstein equations) are written in terms of a metric, which we denote by $g_{\mu\nu}$. This object, a symmetric rank-two tensor, defines our notion of distance, and thus encodes the causal structure of the universe—which events can influence which other events. The Einstein equations are

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

where $G_{\mu\nu}$ is a complicated function of $g_{\mu\nu}$ which will be presented in (1.14), $T_{\mu\nu}$ is a symmetric rank-two tensor that describes the matter and energy content of the universe, and G is Newton's gravitational constant.

Spatially homogeneous and isotropic universes (i.e. ones that at some given time, look the same everywhere and in every direction) define a class of solutions of the Einstein equations. The metric of such a homogeneous and isotropic universe is known as a Friedmann–Lemaître–Robertson–Walker metric (or more usually, an FLRW metric). In this context the Einstein equations reduce to the Friedmann equations, which we will soon present in (1.7). But first, we present the FLRW metric (we will always set the speed of light to unity)

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a(t)^2 dS_3^2, \tag{1.2}$$

where $a(t)$ has the interpretation as the scale factor which describes the evolution of this homogeneous and isotropic universe. The line element ds^2 then determines

the interval between two spacetime points. Defining conformal time τ such that

$$\frac{d\tau}{dt} = \frac{1}{a}, \tag{1.3}$$

we can then write the metric in the useful form

$$ds^2 = a(t)^2 \left(-d\tau^2 + dS_3^2\right). \tag{1.4}$$

The homogeneity, isotropy and curvature of space are encoded in dS_3^2, but a non-trivial time dependence of $a(t)$ breaks homogeneity in time. We will only consider a flat universe, so $dS_3^2 = dx^2 + dy^2 + dz^2$, which corresponds to Euclidean geometry. This makes (1.4) a function of time multiplied by the Minkowski metric, in conformal time. The spatial part of the metric is written in terms of comoving distances, which do not change with the expansion of the universe—to convert a comoving distance r to a physical distance r_{phys} one simply multiplies by the scale factor $r_{phys} = a(t)r$. This means we can calculate the physical velocity

$$v_{phys} = \frac{dr_{phys}}{dt} = a(t)\frac{dr}{dt} + Hr_{phys}. \tag{1.5}$$

The result is a combination of the motion of an object with respect to the comoving coordinates $\left(a(t)\frac{dr}{dt}\right)$ and a term due to the expansion of the universe, which we see is linear in r_{phys} with a constant of proportionality of H.

The local curvature of spacetime and its evolution is determined by the matter and energy present in the universe. Radiation is one of the cosmologically important components of the present universe, in the form of the photons in the cosmic microwave background, the CMB. Today these photons free-stream through the universe. Another component is matter, the primary part of which is *dark* matter. This is invisible to our telescopes (which depend on electromagnetic radiation) but can be mapped by the effect its mass has on the curvature of spacetime, which we see through effects such as the lensing of passing photons. A more familiar component is what cosmologists refer to as baryonic matter, which is the protons, electrons, neutrons and all the other particles which make up the visible galaxies, stars and humans. The last component of our universe that we will mention is dark energy. At present, dark energy makes up the largest contribution to the energy budget of the universe. We will take dark energy to contribute to the energy content of the universe through a cosmological constant

which does not dilute as the universe expands. This will be encoded in (1.1) through the stress-energy tensor $T_{\mu\nu} = M_{Pl}^2 \Lambda g_{\mu\nu}$[1]. The radiation is denoted by γ, the matter by m, and the cosmological constant by Λ.

The components of the matter and energy content of the universe enter (1.1) through their energy-momentum tensors $T^{\mu\nu}$, which in a homogeneous and isotropic universe we can approximate as that of a perfect fluid. Labeling the fluid X,

$$T_X^{\mu\nu} = (\rho_X + p_X) U_X^\mu U_X^\nu + p_X g^{\mu\nu}, \tag{1.6}$$

where the index structure ensures the fluid is isotropic in its rest frame. U_X^μ is the four-velocity of the fluid, with $U_X^0 = 1$ and $U_X^i = 0$ in the rest frame. These fluids are completely characterised by the time dependence of their rest-frame energy density $\rho_X(t)$ and their rest-frame pressure density $p_X(t)$, both of which have no dependence on position, to preserve the homogeneity of the solution. For the components of the universe that we wish to describe, taking the quantity $w_X = \frac{p_X}{\rho_X}$ as constant will be a good description. This equation of state quantity w takes the values 0, 1/3 and -1 for matter, radiation and the cosmological constant respectively.

Fortunately, when $g_{\mu\nu}$ takes the form (1.2), these equations simplify to the promised Friedmann equations

$$H^2 = \frac{8\pi G \rho}{3}, \tag{1.7}$$

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

where $H = \frac{\dot{a}}{a}$, and ρ and p are respectively the sum of the energy densities and of the pressure densities of the components of the universe. We use a dot to refer to a derivative with respect to coordinate time t. We also have the continuity equation

$$\dot{\rho}_X + 3\frac{\dot{a}}{a} (\rho_X + p_X) = 0. \tag{1.9}$$

This can be shown to hold for each component separately using the first law of thermodynamics for an adiabatic process. This describes the effect the expansion of the universe has on the energy density of the fluid.

[1] M_{Pl} is the reduced Planck mass, $M_{Pl}^2 = \frac{1}{8\pi G}$, in a system of units with $c = \hbar = 1$. In practice we will usually also set $M_{Pl} = 1$.

Given the equations of state and the densities of the different components of the universe, one can then calculate the time dependence of the scale factor $a(t)$ and of the densities ρ_X and p_X of each component X. In particular, using (1.9) and $\rho = wp$ we find

$$\rho_X \propto a^{-3(1+w_X)} \tag{1.10}$$

for each component X, which can then be used with (1.7) to determine the time dependence of the scale factor $a(t)$. This is summarised in Table 1.1. This gives us the homogeneous background evolution of the universe—evolution which has observable effects in physical phenomena such as red-shifted photons arriving from distant galaxies.

As we have mentioned, free particles travel on geodesics. This means that they follow the path which extremises the proper time, the time experienced by an observer travelling along with the particle. This principle allows the derivation of the geodesic equation

$$P^\nu \partial_\nu P^\mu + \Gamma^\mu_{\nu\sigma} P^\nu P^\sigma = 0 \tag{1.11}$$

where P^μ is the particle's 4-momentum. The Christoffel symbols $\Gamma^\mu_{\nu\sigma}$ are functions of the metric, and are used to build $G_{\mu\nu}$, which appeared in the Einstein equations (1.1)

$$\Gamma^\mu_{\nu\sigma} = \frac{1}{2} g^{\mu\gamma} \left(\frac{\partial g_{\gamma\nu}}{\partial x^\sigma} + \frac{\partial g_{\gamma\sigma}}{\partial x^\nu} - \frac{\partial g_{\nu\sigma}}{\partial x^\gamma} \right) \tag{1.12}$$

$$R_{\mu\nu} = \partial_\mu \Gamma^\mu_{\nu\sigma} - \partial_j \Gamma^\mu_{\sigma\mu} + \Gamma^\mu_{\mu\tau} \Gamma^\tau_{\nu\sigma} - \Gamma^\mu_{\nu\tau} \Gamma^\tau_{\mu\sigma} \tag{1.13}$$

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \tag{1.14}$$

where $R_{\mu\nu}$ is known as the Ricci curvature tensor, and R is its trace. By homogeneity and isotropy, (1.11) can be rewritten as

$$P^0 \frac{\partial P^\mu}{\partial t} + \Gamma^\mu_{\nu\sigma} P^\nu P^\sigma = 0. \tag{1.15}$$

The magnitude of the physical 3-momentum is $p^2 \equiv a^2 \delta_{ij} P^i P^j$. Evaluating (1.12) for the metric (1.2) and using $P_\mu P^\mu = -m^2 \implies EdE = pdp$ we get that

$$p \propto \frac{1}{a}. \tag{1.16}$$

The interpretation of this result is that as the universe expands, the particle's momentum decreases, in a manner determined by the time dependence of $a(t)$. For a photon, this implies that its energy decreases and its wavelength increases as it free streams through the universe.

Using (1.9) and the equation of state (w, summarised in Table 1.1) for matter, radiation and the cosmological constant Λ we can solve for the dependence of ρ_X on a, as we saw in (1.10). For matter, we obtain the simple result that $\rho_m \propto a^{-3}$, i.e. that it dilutes with the expansion of the universe. For radiation we find $\rho_r \propto a^{-4}$, i.e. that it dilutes with the expansion of the universe, and also loses energy as it redshifts, as we saw in (1.16). For Λ, we find $\rho_\Lambda \propto a^0$, by design.

Epoch	w	$a(t)$	$a(\tau)$
Radiation	$\frac{1}{3}$	$t^{\frac{1}{2}}$	τ
Matter	0	$t^{\frac{2}{3}}$	τ^2
Λ	-1	e^{Ht}	$-\frac{1}{\tau}$

Table 1.1: How the scale factor, $a(t)$, evolves in the different epochs of the universe.

A subscript of 0 will denote a quantity evaluated today. The critical density $\rho_{crit} \equiv \frac{3H^2}{8\pi G}$ is defined as as the (time-dependent) density for which the universe is flat. Using this, and choosing to scale our spatial coordinates such that $a_0 = 1$, we can rewrite the Friedmann equation (1.7) as

$$H^2 = H_0^2 \left(\frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \Omega_\Lambda \right) \tag{1.17}$$

where we define the fractional density

$$\Omega_X = \frac{\rho_X}{\rho_{crit,0}} \tag{1.18}$$

for X being radiation, matter or the cosmological constant. For our flat universe $\sum \Omega_{i,0} = 1$. The measured values for these quantities, as presented in [10], are

$$\Omega_{m,0} = 0.3111 \pm 0.0056, \tag{1.19}$$

$$\Omega_\Lambda = 0.6889 \pm 0.0056. \tag{1.20}$$

The temperature of the blackbody CMB radiation has also been found to be $T_{\gamma,0} = 2.725 \pm 0.001K$ [11].

Given these numbers, and (1.17), we can make some statements about the past, present and future of the universe. The numerical evolution of the components is illustrated in Figure 1.1, with the recent past highlighted in Figure 1.2. Firstly, the present—we see that the cosmological constant, Λ, is dominant now. We also see that the energy density of radiation is very small, $\Omega_{\gamma,0} \ll \Omega_{m,0}$. Looking to the past, we see that the transition from matter domination to Λ domination happened relatively recently in the history of the universe, at $a_{m\Lambda} = \left(\frac{\Omega_{m,0}}{\Omega_\Lambda}\right)^{\frac{1}{3}} \approx 0.77$. We can see that this domination will continue as the matter and radiation content of the universe dilute and redshift away. Running the clock backwards, that is towards $a = 0$, we see that the matter and radiation had higher energy densities in the early universe, which is due to their scaling as a^{-3} and a^{-4} respectively. Running the clock back far enough, we see that despite $\Omega_{\gamma,0} \ll \Omega_{m,0}$ today, at one stage it was the case that $\Omega_\gamma \gg \Omega_m$, and the universe was dominated by radiation. This is due to the relative scaling $\frac{\Omega_\gamma}{\Omega_m} \propto a^{-1}$. This epoch is the limit on how far back in time we can see. At sufficiently early times, the density of free electrons was so high that the mean free path of photons was very short. Once the temperature dropped enough that neutral atoms formed, the mean free path of photons became long enough that they could travel cosmological distances, and eventually fall into our telescopes—this is the origin of the CMB.

The standard cosmological model of the history of the universe, the ΛCDM (Λ, cold dark matter) model, takes this era of radiation domination as its starting point. In the distant past, the components of the universe (all the different types of matter and radiation) were in thermal equilibrium with each other—that is, they were interacting with each other, and the interaction was sufficiently efficient that their temperatures matched and the net flow of thermal energy between them was zero.

The evidence for this story is strong. It includes observable predictions for the relative abundances of the lightest elements [12] (which were forged in the first few minutes of the universe) and baryon acoustic oscillations [13] in the CMB and large scale structure (LSS). The latter is not evidence for thermal equilibrium, but can be directly linked to the end of the coupling between the baryons and the acoustic waves in the radiation, before the release of the CMB. There are however unanswered questions within the ΛCDM model. One is the Hubble tension, an apparent tension between different measurements of the precise rate of expansion of the universe today—see for example [14, 15]. Despite

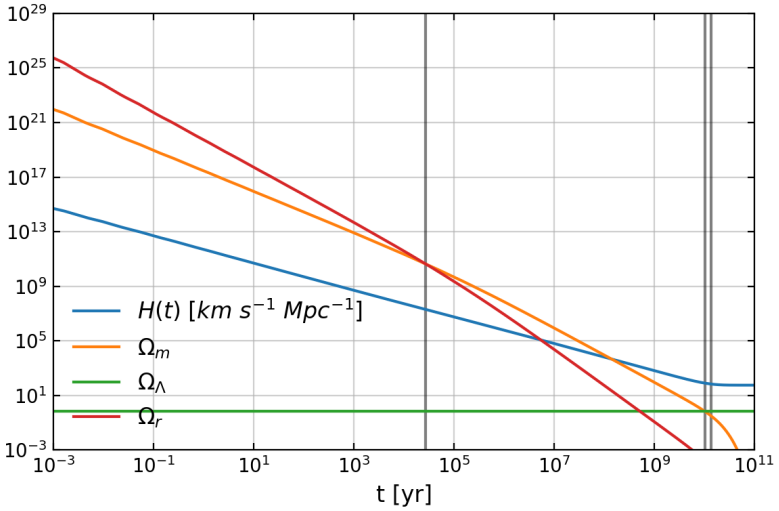


Figure 1.1: The evolution of the components of the universe up to the present, and slightly beyond. The vertical grey lines mark matter-radiation equality, matter-dark energy equality, and the present day, respectively. These quantities are evolved using (1.17) and (1.9). In the past, densities of matter and radiation were far higher, and the radiation energy density dominated over the matter. During this high density epoch, the expansion of the universe ($H(t)$) was far stronger. In the future, as Λ comes to dominate, $H(t)$ will become constant.

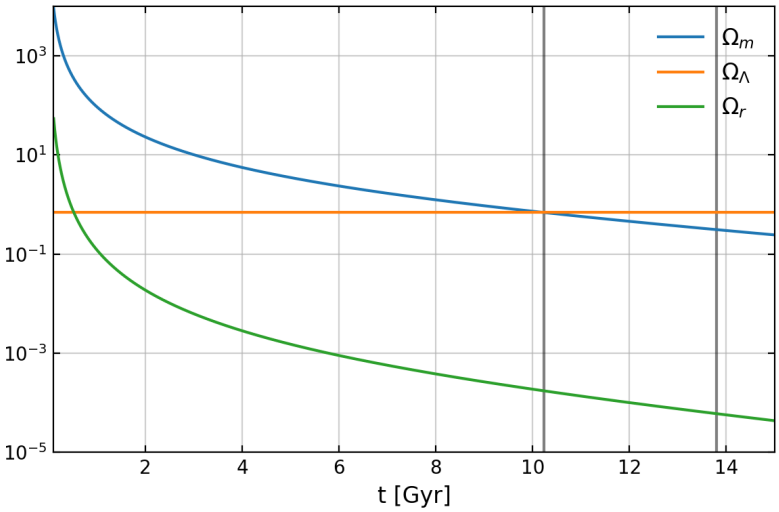


Figure 1.2: The evolution of the components of the universe, zoomed in to more clearly show the matter-dark energy transition. The first vertical grey line is matter-dark energy equality, the second is the present day.

the remaining questions, it is incredible that only a century ago the cutting edge debate was whether the universe we know had a beginning or not—compared to now, where the debate rages over the second and third significant figures of its age.

Before focusing our whole attention on the past in the rest of this thesis, let us briefly look toward the future of the universe. What will we see as the universe evolves? Eventually a will be sufficiently large that Ω_Λ will the dominant contribution to the energy budget of the universe. The beginning of this epoch can be seen in Figure 1.3 and Figure 1.4, which show the evolution of $a(t)$ and $\dot{a}(t)$.

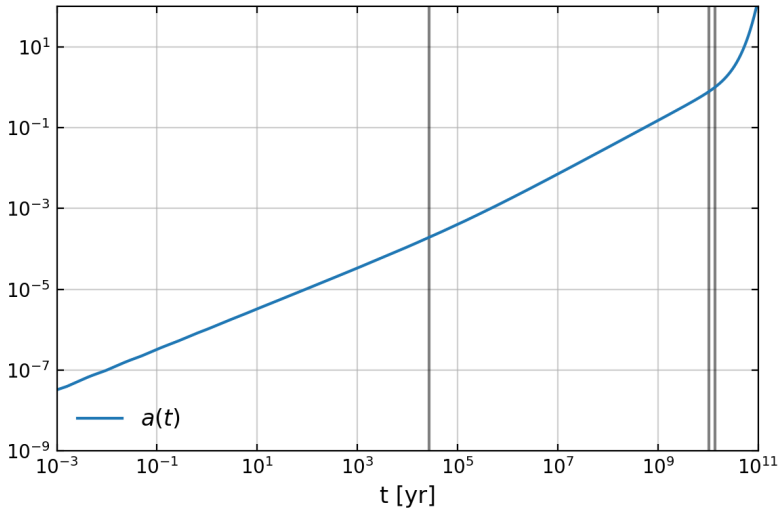


Figure 1.3: The evolution of the scale factor. For most of the ΛCDM history it evolves as some power of t (see Table 1.1), however as Λ comes to dominate it will begin to grow exponentially. The vertical grey lines mark matter-radiation equality, matter-dark energy equality, and the present day, respectively.

In this Λ dominated epoch we can use (1.7) to find

$$H^2 = H_0^2 \Omega_\Lambda \tag{1.21}$$

$$\implies \dot{a} = \pm H_0 \sqrt{\Omega_\Lambda} a \tag{1.22}$$

from which the initial conditions pick out the exponentially expanding solution, $a(t) = a_0 e^{H_0 \sqrt{\Omega_\Lambda}(t-t_0)}$. Using the Friedmann equation for sufficiently far in the future we can rewrite this as

$$a(t) = a_0 e^{H(t-t_0)} \tag{1.23}$$

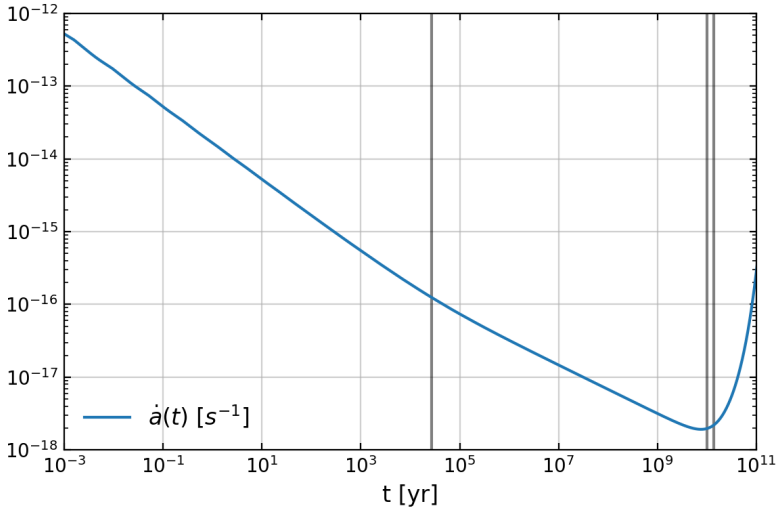


Figure 1.4: During the radiation and matter dominated eras, the evolution of $a(t)$ as been slowing—this is decelerating expansion. However, as Λ comes to dominate, $\dot{a}$ will begin to increase, and the universe will enter an epoch of accelerated expansion. The vertical grey lines mark matter-radiation equality, matter-dark energy equality, and the present day, respectively.

where H is a constant, having frozen in to a factor $\sqrt{\Omega_\Lambda}$ smaller than its value today. One question we can ask, to obtain some physical meaning for this evolution, is how far light will travel in the lifetime of the universe. Since $ds^2 = 0$ for a photon, in terms of conformal time τ we find that for the photon travelling in the x direction, $dx = d\tau$. This means that on a plot of x vs. τ, the path of a photon will be a line of a slope of unity, even in an expanding universe. Integrating, we then find that

$$\tau(t) = x(t) = (Ha_0)^{-1} \left(1 - e^{-H(t-t_0)}\right) + \tau_0 \tag{1.24}$$

where as usual a 0 subscript denotes a quantity evaluated today. We see that the comoving distance the photon will travel is finite, even though the physical distance (and the time taken, $\int dt$) diverges. Numerically integrating (1.17) (including radiation, matter and Λ) we obtain the full evolution of τ, illustrated in Figure 1.5. We find that the points that we see in today's CMB are a comoving distance of 14Gpc away. Since $a_0 = 1$, the present physical distance to these points is equal to the comoving distance, but this will increase exponentially in the future. We find the asymptotic value of τ to be 19Gpc, indicating that we will never see the light that was emitted from the CMB at comoving distances greater than this.

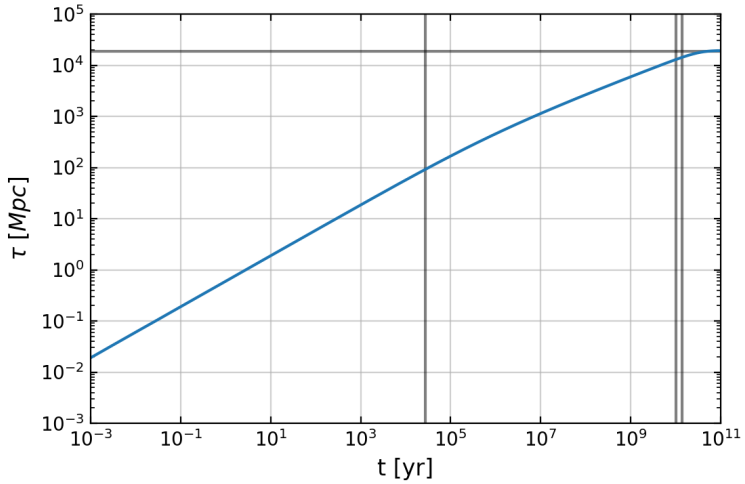


Figure 1.5: The evolution of the conformal time τ during ΛCDM. We see that due to the eventual Λ dominance, τ will asymptote to a constant, denoted here by the horizontal grey line at $\tau = \tau_0 + (H_0 a_0)^{-1}$. Thus, there is a maximum comoving distance that we can ever expect to receive CMB photons from. The vertical grey lines mark matter-radiation equality, matter-dark energy equality, and the present day, respectively.

1.2 Initial conditions for ΛCDM

1.2.1 Motivations for inflation

The ΛCDM model relies upon an early epoch in which the components that we have discussed are in thermal equilibrium, at an incredibly high temperature. These components were distributed almost perfectly uniformly, but with some tiny perturbations. This simple scenario then evolves under the expansion of the universe and the gravitational collapse of perturbations (and other interactions) and forms the universe we know. These special characteristics of this early epoch then become clues to the epoch of the universe that preceded ΛCDM.

One such clue that we have is known as the horizon problem. Roughly speaking, it is the problem that the universe is more homogeneous on large scales than we would expect, given the history of ΛCDM. Why is the temperature on one side of the sky so close to the temperature on the other side of the sky?

During the matter dominated era, $p = 0$ and $a(\tau) \propto \tau^2$. Therefore we can approximate the total comoving distance a photon could have traveled between

the singularity at $a_i = 0$ and recombination at $a_{rec} = 1/1100$. We find

$$x_{rec} = \int_0^{\tau_{rec}} d\tau = \int_0^{a_{rec}} \frac{d\tau}{da} da \propto \sqrt{a_{rec}}. \tag{1.25}$$

We can see that this is finite. Including the radiation dominated period does not change this conclusion—performing the integration numerically we find the comoving distance $x_{rec} = 280$Mpc. We can also calculate that the comoving distance that a photon could have freely streamed through the universe since last scattering is $x_0 = 14000$Mpc. This matches the physical distance since $a_0 = 1$. This means that we would expect the homogeneous patches at recombination to span at most an angle of

$$\tan^{-1}\left(\frac{2x_{rec}}{x_0 - x_{rec}}\right) \approx 2^\circ \tag{1.26}$$

on the CMB sky today. We can also obtain an order of magnitude approximation of the number of disconnected patches we should see, by simply dividing the area of the CMB sky by the apparent area that could be causally affected by a single point at $a = 0$

$$\frac{4\pi(x_0 - x_{rec})^2}{\pi(x_{rec} - 0)^2} \approx 10^5. \tag{1.27}$$

This is completely at odds with observations—the CMB is in fact very close to homogeneous *across the entire sky*, clearly a feature of our universe that requires an explanation.

While in the basic ΛCDM model there has not been enough conformal time since $a = 0$ for opposite sides of the CMB to come into thermal equilibrium, this can be remedied by adding an epoch that causes the horizon (the characteristic conformal length scale $(aH)^{-1}$) to shrink—this would give large scales enough time to come into thermal equilibrium, before they leave the horizon. The usual ΛCDM evolution would then come into play, bringing these scales back inside the horizon where we can observe them. One way of doing this is via an effective cosmological constant—in that case the horizon $1/(aH)$ shrinks exponentially. We then find that $\tau = -1/(aH)$, so in the limit $a \to 0$ we see there is infinite conformal time in the past, to allow the entire observable universe to come into causal contact. We refer to this period as the epoch of inflation.

We will mention three more features of the ΛCDM initial conditions that we would like an explanation for. Adiabaticity is the statement that for each of

the components (denoted here by i) of the universe, whose energy densities we can write as $\rho_i(t) = \bar{\rho}_i(t) + \delta\rho_i(t, \mathbf{x})$, their initial background values $\bar{\rho}_i$ and their initial perturbations $\delta\rho_i$ were related such that

$$\frac{\delta\rho_i}{\dot{\bar{\rho}}_i} = \frac{\delta\rho_j}{\dot{\bar{\rho}}_j}. \tag{1.28}$$

To phrase this another way, we can choose a local reparametrisation of time $t \to t + \delta t(x)$ with $\delta t(x) = \frac{\delta\rho_i}{\dot{\bar{\rho}}_i}$, so that

$$\bar{\rho}_i(t) + \delta\rho_i(t, x) = \bar{\rho}_i(t) + \delta t(x)\dot{\bar{\rho}}_i \approx \bar{\rho}_i(t + \delta t(x)) \tag{1.29}$$

and thus the reparametrisation can absorb the linear perturbations in *all* of the components. An implication we can draw from this is that a single degree of freedom would be sufficient to lay down these initial perturbations.

The Gaussianity of the initial conditions is the statement that their statistics are completely described by their two-point correlations. We will define this, and discuss in more detail in Section 1.3.1, where we will also define scale invariance, the statement that correlations do not depend on the scale at which they are measured. It turns out that observations show that the CMB appears close to Gaussian, although scale invariance is not precisely respected. The deviation from perfect scale invariance has been measured to high confidence by the *Planck* satellite, with earlier measurements made by the WMAP satellite [16]. These are some of the properties of the early universe that the epoch of inflation must explain.

1.2.2 Driving inflation with a scalar field

These features of the early universe can be explained by a period of near-exponential expansion preceding the usual ΛCDM evolution. One simple way one could imagine driving this expansion is through a single scalar field. At the end of inflation, this scalar field is usually assumed to decay, eventually leaving the universe filled with the standard components of ΛCDM, in thermal equilibrium. The mechanism for this process, known as reheating, is unknown.

Consider an inflaton action of the form

$$S = \int d^4x\sqrt{-g}P(X, \phi) \tag{1.30}$$

with $\phi = \phi(t)$ and $X \equiv \frac{1}{2}\dot{\phi}^2$. We will work with the number of e-folds, N, as our time variable, defined by $\frac{a}{a_0} \equiv e^{N-N_0}$. We then have $x' \equiv \frac{dx}{dN} = H^{-1}\frac{dx}{dt}$.

The energy-momentum tensor is defined as

$$T_{\mu\nu} = -2\frac{\partial \mathcal{L}_m}{\partial g^{\mu\nu}} + g_{\mu\nu}\mathcal{L}_m \tag{1.31}$$

where $\mathcal{L}_m$ is the non-gravitational part of the Lagrangian density $\sqrt{-g}\mathcal{L} = \sqrt{-g}\left(\frac{R}{2} + \mathcal{L}_m\right)$. The energy and pressure densities can then be identified as

$$\rho = 2XP_{,X} - P \tag{1.32}$$

$$p = P, \tag{1.33}$$

so then the continuity equation (1.9) implies

$$\rho' = -6XP_{,X}\,. \tag{1.34}$$

The "sound speed" for the adiabatic perturbations is defined as [17]

$$c_s^2 = \left.\frac{\partial P}{\partial \rho}\right|_S = \frac{\dot{P}}{\dot{\rho}} = \frac{P_{,X}}{\rho_{,X}}. \tag{1.35}$$

When we allow perturbations about the background this will have the interpretation of a sound speed of those perturbations. Since we also have (1.32), we see that

$$c_s^2 = \frac{P_{,X}}{P_X + 2XP_{XX}}. \tag{1.36}$$

For canonical inflation models, $P(X,\phi) = X - V(\phi)$ so $P_{XX} = 0$ and $c_s = 1$. For non-canonical models with more complicated kinetic terms, the sound speed can deviate from unity. One such example we will follow in detail in this work is Dirac-Born-Infield (DBI) inflation [18].

For reasons we will discuss in Section 2.1.1 we will define a quantity τ_s in analogy with the usual τ defined in (1.3):

$$\tau_s' \equiv \frac{c_s}{aH}. \tag{1.37}$$

We also define the standard "slow-roll" parameters:

$$\epsilon \equiv -\frac{d\ln H}{dN} \tag{1.38}$$

$$\eta \equiv +\frac{d\ln \varepsilon}{dN} \tag{1.39}$$

$$\epsilon_s \equiv +\frac{d\ln c_s}{dN}. \tag{1.40}$$

From the continuity equation and the Friedmann equation we find the useful relation

$$\varepsilon = \frac{1}{2}\phi'^2 P_{,X}\,. \tag{1.41}$$

Using (1.34), we see that the equation of motion for ϕ is [19]

$$\phi'' + (3c_s^2 - \varepsilon)\phi' + H^{-2}\frac{\rho_\phi}{\rho_X} = 0. \tag{1.42}$$

We will now briefly discuss the special case of canonical models, with $P(X,\phi) = X - V(\phi)$. In this case (1.41) simplifies to $\varepsilon = \frac{1}{2}\phi'^2$. We also have

$$\rho_\phi = \frac{1}{2}\dot\phi^2 + V(\phi) \tag{1.43}$$

$$P_\phi = \frac{1}{2}\dot\phi^2 - V(\phi) \tag{1.44}$$

$$\implies H^2 = \frac{V(\phi)}{3-\varepsilon} \tag{1.45}$$

where the last equality used the Friedmann (1.7) equation and has not assumed ε is small. For successful inflation with the potential of the scalar field acting as a cosmological constant, we want $w \approx -1$, i.e. that $V(\phi) \gg \frac{1}{2}\dot\phi^2 = \varepsilon H^2$. Since the kinetic term is required to be small, this is referred to as *slow-roll* inflation. For a canonical model in slow-roll, the Friedmann equations then become

$$H^2 \approx \frac{1}{3}V(\phi), \tag{1.46}$$

$$\frac{\ddot a}{a} \approx -\frac{1}{6}\left(-2V(\phi)\right). \tag{1.47}$$

For a model with a canonical kinetic term, (1.42) simplifies to

$$\phi'' + (3-\varepsilon)\phi' + H^{-2}V_\phi(\phi) = 0. \tag{1.48}$$

In the slow-roll approximation we assume that ϕ' is not just transiently small, i.e. that $\phi'' \ll \phi'$, and so

$$\phi' \approx \frac{V_\phi(\phi)}{3H^2} \approx \frac{V_\phi(\phi)}{V(\phi)}. \tag{1.49}$$

From this, we see that demanding that ε be small places a constraint on the flatness of the potential in a canonical model. Requiring inflation to last sufficiently long also constrains $\eta \ll 1$.

For inflation to solve the horizon problem, it must result in a shrinking comoving Hubble radius, disconnecting regions that had previously been in thermal contact. This implies that

$$\frac{d}{dt}(aH)^{-1} = -\frac{\ddot{a}}{\dot{a}^2} < 0 \tag{1.50}$$

so $\ddot{a} > 0$, i.e. the expansion is accelerating. Another way of writing this is

$$\frac{d}{dt}(aH)^{-1} = -\frac{1}{(aH)^2}\left(\dot{a}H + a\dot{H}\right) \tag{1.51}$$

$$= -\frac{1}{a}(1-\varepsilon) \tag{1.52}$$

and so we need $\varepsilon < 1$ to successfully inflate the universe. However we see that for near-exponential expansion we have already demanded $\varepsilon \equiv -\dot{H}/H^2 \ll 1$ so this must be satisfied.

1.3 Statistical observables

1.3.1 From probabilities to observations

The discussion until this point has focused on quantities with definite values. This will no longer be true when we discuss quantities that have a quantum origin, as the prediction of a fundamental quantum theory is a statistical one. The true prediction is of a distribution from which our observation will be drawn; as such, to link the sky we see to some fundamental theory, we must talk in terms of probabilities.

We can define the expectation value of a functional F of a field configuration

$f(x)$ as

$$\langle F[f]\rangle = \int \mathcal{D}f \; F[f]\, P[f] \tag{1.53}$$

where the functional integral $\int \mathcal{D}f$ over f is over all the field configurations that $f(x)$ can take, and $P[f]$ is the corresponding probability distribution. In this thesis, we will be working with the expectation value of functions of field configurations, so will be working with quantities defined in terms of (1.53). In particular, the quantum mechanical nature of the inflationary origin of structure means that such predictions of inflationary theories will necessarily be statistical. That is to say, the predictions through which we will test these theories will be predictions of expectation values of these primordial perturbations.

The ensemble average (1.53) of $f(\mathbf{x})f(\mathbf{x}')$ for some function f is known as the two-point correlator, or the two-point function: $\langle f(\mathbf{x})f(\mathbf{x}')\rangle$. Note the brackets $\langle \ldots \rangle$ refer to the expected value over the ensemble, over all the possible realisations that could be drawn from the probability distribution, and not a spatial average. Higher order correlators are the expectation values of products the field evaluated at more points.

There are two assumptions we wish to immediately impose on such functions. Those conditions are statistical homogeneity and statistical isotropy, the requirement that the correlators are invariant under translations and rotations. This does not mean that some realisation of the ensemble is forbidden from breaking these symmetries, of course—fields are still allowed to be inhomogeneous, and thus have special points such as local extrema. The statement of statistical homogeneity is that, when averaged over the entire ensemble of possible realisations, these special points are equally likely to appear anywhere in space. All observations of the universe so far appear to respect these two symmetries of the statistics.

These conditions on the correlators of $f(\mathbf{x})$ place useful constraints on the correlators of the Fourier transform, $f(\mathbf{k})$. They demand the form

$$\langle f(\mathbf{k_1})f(\mathbf{k_2})\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k_1}+\mathbf{k_2})P(k_1), \tag{1.54}$$

$$\langle f(\mathbf{k_1})f(\mathbf{k_2})f(\mathbf{k_3})\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})B(k_1,k_2,k_3), \tag{1.55}$$

where $P(k)$ is known as the power spectrum and $B(k_1,k_2,k_3)$ is the bispectrum. The delta functions come from statistical homogeneity, while the fact that $P(k)$ and $B(k_1,k_2,k_3)$ depend only on the magnitudes of the vectors ($k = |\mathbf{k}|$) is

a result of demanding statistical isotropy. Note that $\langle f(\mathbf{k_1})f(\mathbf{k_2})f(\mathbf{k_3})\rangle$ has 9 apparent degrees of freedom, but $B(k_1, k_2, k_3)$ has only three. This is due to the delta function demanding that the sum of the $\mathbf{k_i}$ vanishes, thus forming a triangle. Since the orientation of this triangle cannot matter, three parameters will suffice. The conventional choice is the magnitudes of the vectors. Similar statements can of course be made for higher point correlators.

Our focus in this thesis will be on the predictions that models of inflation make for the bispectrum $B(k_1, k_2, k_3)$. This is one characteristic property of the initial conditions for the ΛCDM evolution. From this, one can calculate statistical properties of the CMB sky that we see. If the bispectrum has the form

$$B(k_1, k_2, k_3) = X(k_1)Y(k_2)Z(k_3), \tag{1.56}$$

or can be expressed as a sum of such terms, it is called separable. This property allows useful simplifications to such calculations, that will be discussed in Chapter 2.

Another symmetry one could imagine is statistical scale invariance. Scale invariance is the statement that

$$\langle f(\mathbf{x})f(\mathbf{x}')\rangle = \langle f(\lambda\mathbf{x})f(\lambda\mathbf{x}')\rangle\,. \tag{1.57}$$

One can show that this symmetry demands $P(k) \propto k^{-3}$. Motivated by this, it is usual to define the dimensionless power spectrum

$$\mathcal{P}(k) = \frac{k^3}{2\pi^2}P(k). \tag{1.58}$$

Evidence from the cosmological scales we can observe suggests that the symmetry (1.57) is only approximately respected by the probability distribution from which our universe was drawn. In fact, it has been estimated that $P(k) \propto k^{n_s-4}$, with $n_s = 0.9649 \pm 0.0042$ at 68% CL [20, 21]. Thus we can make the statement that the odds of a universe like ours being drawn from a perfectly scale-invariant probability distribution are negligibly small.

1.3.2 From observations to probabilities

Instead of describing observations generated from known probabilities, we usually wish to do the opposite. We wish to take a series of observations of some random variable and use these observations to understand the probability distribution

that they were drawn from. It is reasonable to wonder how we could ever make contact with a theory that only makes these kind of predictions, given that we only have access to one universe. Are we doomed from the start? Thankfully, it turns out, that is not the case.

Inflation is the theory we use to define the probability distribution that our universe may have been drawn from. While we may not have access to other universes drawn from the same distribution, one can ask how much information we can glean from the one we do have. For example, if we magically had access to the entirety of one spatially infinite universe then what would our measured spatial averages tell us? Would those averages match the ensemble averages? If the distribution is such that the correlations between the field evaluated at distant points goes to zero sufficiently fast, then we could intuitively imagine that the answer is yes—since distant regions are uncorrelated, they are effectively separate draws, and so a spatially infinite universe will give us enough information to measure the true ensemble average. This intuition is made precise by the ergodic theorem (see for example appendix D of [22]). If the assumptions of the ergodic theorem are met, then spatial averages and ensemble averages match—despite the fact we only have one universe, if we had the whole spatial infinity of that universe we would still be able to measure the ensemble average. Of course, we do not have access to spatial infinity, thus our spatial averages will only ever be an approximation of the true ensemble averages—this is known as cosmic variance. To summarise, there are two ways to beat this cosmic variance (to measure an ensemble average $\langle f(\mathbf{x})^2 \rangle$ at some fixed $\mathbf{x}$, say). We could stay in the same spatial region, but magically peek into all the other universes. Or, we could stay in the same universe, but magically peek into every spatial region. The ergodic theorem tells us that these two methods will give the same answer, thus allowing us to make contact between spatial measurements made in our one universe and the ensemble predictions of inflation.

In practice, even in a limited spatial region we can have access to multiple samples drawn independently from one distribution. One such situation we will see in Section 2.2, with the multipole moments of the CMB temperature fluctuations. For a given large angular scale there will be multiple coefficients drawn from the same distribution due to statistical isotropy, improving the precision with which the parameters of the distribution can be estimated.

1.4 Thesis outline

Given the context we have outlined, we will now give a description of the goals, methods and results presented in this thesis. The main goal discussed here is to develop an efficient numerical pipeline for connecting inflation models directly to observations through the CMB bispectrum. The concrete results of such a goal are constraints on the parameters of inflation models, not constraints on phenomenological templates or summary f_{NL} parameters. This allows the use of the full bispectrum shape information from an inflation scenario, not point samples or a limit. The novelty of our results comes from our separable approach and numerical methods granting access to more accurate, and new, bispectrum shapes.

To achieve these goals we have developed methods to preserve the separability that is built in to the tree-level in-in formalism—these terms will be defined in Chapter 2. We do this through an expansion in a primordial basis. This will link scenarios of inflation to the CMB, through an estimator that will be presented in [2]. The calculations involved in that estimator are expensive, but need only be done once per primordial basis. This motivates the desire for a basis that converges quickly for a broad range of inflation models, and as such we have explored this topic in detail.

The first line of research presented in this thesis is the determination of the feasibility of the overall method. The initial result here was identifying the contributions of the non-physical configurations as a novel problem to this formalism, and identifying basis choice as the key method of overcoming this difficulty. In this work we describe multiple basis sets and make quantitative comparisons of their convergence on realistic and interesting models, including ones with features. These basis sets can overcome the difficulty of the dominant non-physical k-configurations, converging far more efficiently than the basic basis sets, for physically interesting models.

The second line of research presented is the development of basis-independent methods that allow the fast and accurate calculation of higher order coefficients of the basis expansion of the tree-level in-in formalism. To this end we set up the separable formalism and describe the methods used to overcome the difficulties encountered. These difficulties include accurately including the highly oscillatory early-time contributions, and avoiding difficult numerical cancellations between terms in the interaction Hamiltonian (and corrections from field redefinitions). A careful and comprehensive validation of our methods is presented, showing that

the convergence problems can be overcome, the oscillatory contributions captured efficiently. At the primordial level this is done by validating our results using three distinct tests. Validation is done on established templates with non-trivial features; the squeezed-limit consistency condition is verified; our full bispectrum results are compared to previous codes through point-tests.

As a side-effect of this desire for a separable primordial bispectrum, the primordial calculation of the inflationary bispectrum presented in this work is much more efficient than previous methods, in the sense that it converges far faster in modes than previous methods would in point samples. These methods are implemented in the PRIMODAL code, a Python code which can quickly and precisely produce a full feature bispectrum, despite minimal optimisation. It is presently implemented for single-field models of inflation.

The final highlighted result, obtained in collaboration, is the connection of an inflationary scenario to the CMB through the CMB-BEST code[2]. This allowed in particular a constraint to be placed on the sound speed of DBI inflation, c_s. At the level of the CMB, validation of the method as a whole (when applied to the *Planck* data) is achieved by comparison with the *Planck* constraint on the DBI sound speed. This constraint is translated into a constraint on a fundamental parameter in the context of the inflationary parameter scan.

Chapter 2 of this thesis is an introduction to the bispectrum as an observable. The various parts of the pipeline that connects inflation scenarios to observations through the bispectrum are reviewed. The usual paradigm of bispectrum estimation in the CMB is outlined, along with the motivation for separable bispectra. The in-in formalism for calculating the tree level bispectrum for a given model of inflation is presented in detail, and the point at which the separable formalism diverges is highlighted. Reviews of $P(X, \phi)$ models of inflation, and discussions of previous numerical codes for calculating the primordial bispectrum k-configuration by k-configuration are presented. Also reviewed is the previous work in achieving separability through modal expansions in [23].

Chapter 3 describes the first line of research mentioned above—the discussion covers the need to take the non-physical configurations into account and how the basis sets presented in this work are built. Also presented is a quantitative comparisons of the convergence of each basis set to relevant examples of bispectrum shapes.

Since the paradigm presented here is only viable if one can find a basis that

[2]Developed by Wuhyun Sohn.

can efficiently represent a wide variety of bispectra, the majority of Chapter 3 is devoted to presenting this exploration. This topic is distinct and separate from our numerical methods, but nonetheless vital.

Chapter 4 details the second line of research mentioned above—a discussion of the details of recasting the in-in formalism in an explicitly separable form is presented, making explicit each step of the calculation. Methods for efficiently dealing with the early time contributions to the integrals are discussed, as well as other numerical issues that require care and attention. Also presented are validation tests on a very broad range of types of non-Gaussianity.

Chapter 5 presents a constraint on the sound speed of the DBI model, which validates our pipeline against a constraint from the *Planck* analysis. This chapter will describe the parametrisation of the scan, and pay careful attention to the convergence of the result with respect to the basis size.

Chapter 6 discusses possible avenues of future work, and presents our conclusions.

Chapter 2

Introduction II: Probing Inflation

In this chapter we will review details of the bispectrum, including how it is predicted by an inflationary scenario and connected to the CMB sky. We will point out where our methods diverge from previous work, briefly discuss the computational complexity of the CMB calculation, and the approximations that have previously been used to make this tractable.

The primordial bispectrum is one of the main characteristics used to distinguish between models of inflation. While it is well known that the physics of inflation must have been extremely close to linear, and the initial seeds of structure it laid down drawn from a probability distribution that was very close to Gaussian, there is expected to have been some level of coupling between the Fourier modes of the perturbations. In the simplest example of an inflation model this is expected to be unobservable [24], but the possibility remains that inflation was driven by more complex physics that may have left an observable imprint on our universe today. Some models of inflation have interactions that predict non-Gaussian correlations at observable levels. This can happen through self-interactions [5, 18], interactions between multiple fields [25–37], sharp features [38–40] and periodic features [41–50]. Another observable consequence of primordial non-Gaussianity that has been studied, through which it could possibly be probed, is the generation of primordial black holes [51–54].

We will focus on non-Gaussianity from a single field in this thesis, however inflation scenarios with more than one field are an active field of research. In these models, which are motivated by string theory considerations [27], non-linearities of interactions between fields during inflation can generate observably

large non-Gaussian signals. This is in part because they can evade the squeezed limit consistency condition [55]. The extra degrees of freedom in these scenarios can also generate bispectrum shapes not usually seen in single-field models (for example, see recent work in [56, 57]). Multi-field models are a prime target for future work—the methods outlined in this thesis have been implemented and tested for single-field models, though they are expected to generalise to this case.

If the fundamental physical model of inflation is a single field, then the non-linearities of the self-interactions of that field may have produced observably large non-Gaussian signatures [36, 58]. Detecting these signatures in the CMB bispectrum requires an understanding of their form [59], which motivates the study of these dynamics. Models of inflation that can produce observably large non-Gaussian signatures, and are therefore constrained by the current experimental bounds [60], include string theory-inspired models such as DBI inflation [61] and axion-monodromy [62, 63] (which gives large non-Gaussianity through the resonance mechanism).

There is an extensive literature on the details of the calculation of bispectra from models of inflation [5, 38, 41, 64–69]. Multi-field models can produce large correlations between modes of very different scales; non-canonical kinetic terms can reduce the sound speed of the perturbations, boosting both the smooth non-Gaussian correlations, and any features which may be present [18, 61, 70–74]; effectively single-field models with imaginary sound speeds can generate a bispectrum mostly orthogonal to the usual equilateral and local templates [75].

However, understanding the inflationary prediction is only the first part of the process. The computational challenge of using these predictions to translate CMB data into constraints on specific inflation scenarios has been tackled by various methods. Much progress has been made by coarse-graining the model space, using a small number of approximate representative templates, and leveraging the simplifying characteristic of separability with respect to the three parameters of the bispectrum [59, 76].

The primordial bispectrum is the Fourier equivalent of the three-point correlator of the primordial curvature perturbation. If this field is Gaussian, the bispectrum vanishes, so it is a valuable measure of the interactions in play during inflation. If some inflation model predicts a bispectrum that is sufficiently well approximated by the standard separable templates, the constraints on those standard templates can be translated into constraints on the parameters of the model. The fact that the amplitudes of all primordial templates estimated thus far from the CMB are consistent with zero has already provided such constraints

in certain scenarios [60, 77]. With the high-precision *Planck* data, and data from forthcoming experiments such as the Simons Observatory (SO) [78] and CMB-S4 [79], robust pipelines must be developed to extract the maximum amount of information possible, by bypassing the approximations and limitations of previous methods.

Constraining an arbitrary template using previous methods has been difficult due to the nature of bispectrum estimation in the CMB and LSS [80–83]. Our aim in this work is to develop the inflationary part of a pipeline to allow to efficiently test a much broader range of models. In this work, we explore shapes arising from tree-level effects in single field models. We do this numerically, allowing quantitative results for a broad range of models, and avoiding extra approximations. Our general aim is to apply the Modal philosophy of [84–86] to calculating primordial bispectra. This Modal philosophy is a flexible method that has broadened the range of constrained bispectrum templates, by expanding them in a carefully chosen basis. The Modal estimator is thus capable of constraining non-separable templates, while the *KSW* estimator [59], for example, cannot.

To do this, we will exploit the intrinsic separability of the tree-level in-in formalism to apply the Modal methods at the level of inflation. Expressing the primordial bispectrum in a separable basis expansion turns out to lead to an increase in the efficiency of the calculation at the primordial level. However, the main advantage is still that expressing the primordial shape function in this way reduces the process of bispectrum estimation in the CMB to a cost which (while large) need only be paid once per basis, not per scenario. The details of the bispectrum estimation part of the pipeline will be detailed in [2].

A proof of concept of this approach at the primordial level was presented in [23]. We go beyond the work of [23] both in developing the choice of basis (the feasibility of the method depending vitally on the chosen basis achieving sufficiently fast convergence in a broad range of interesting models) and in the methods we use to allow us to go to much higher order in our primordial expansion, allowing us to apply the method to feature bispectra for the first time.

2.1 Perturbations about the background

We begin by discussing the details of the perturbations. The action for the perturbations is derived by splitting the relevant quantities into a background part and a small perturbation, which we can parameterise for the inflaton ϕ and

the metric $g_{\mu\nu}$ as

$$\phi(t,\mathbf{x}) = \bar{\phi}(t) + \delta\phi(t,\mathbf{x}), \qquad g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu} \tag{2.1}$$

where $\bar{g}_{\mu\nu}$ is the unperturbed metric defined in (1.2). In the uniform density gauge we have (neglecting tensor modes)

$$\delta\phi = 0, \qquad h_{ij} = a^2 e^{2\zeta}\delta_{ij} \tag{2.2}$$

where h_{ij} is the spatial part of the metric. The second order action for the perturbations in this gauge can be found to be

$$S_2 = \int d^3x dt\ \frac{a\varepsilon}{c_s^2}\left(a^2\dot{\zeta}^2 - c_s^2\left(\partial\zeta\right)^2\right) \tag{2.3}$$

Defining $v = -z\zeta$ and switching to conformal time we can rewrite (2.3) as

$$S_2 = \frac{1}{2}\int d^3x d\tau \left[\left(\frac{dv}{d\tau}\right)^2 - c_s^2\left(\partial v\right)^2 + \frac{1}{z}\frac{d^2z}{d\tau^2}v^2\right] \tag{2.4}$$

where

$$z^2 = \frac{2a^2\varepsilon}{c_s^2} = a^2\frac{\rho + P}{\left(c_s H\right)^2}. \tag{2.5}$$

In terms of v, we can recognise the dynamics as that of a canonically normalised scalar field, oscillating with a frequency that changes as the universe expands. The sound speed c_s (1.35) is the propagation speed of the perturbations. This defines a sound horizon, c_s/H—as the modes cross this sound horizon they freeze. Expanding the mass, we find

$$\frac{1}{z}\frac{d^2z}{d\tau^2} \approx 2\left(aH\right)^2 \tag{2.6}$$

where we have neglected terms that are slow-roll suppressed. The evolution of the perturbations is therefore well approximated using $\frac{1}{z}\frac{d^2z}{d\tau^2} = 2\left(aH\right)^2$, unless the evolution deviates from the slow-roll behaviour.

2.1.1 Evolution of the perturbations

Varying the second order action (2.3), we can calculate the equation of motion of the perturbations. In terms of derivatives with respect to N, we obtain

$$\zeta_k{}'' + (3 - \varepsilon + \eta - 2\varepsilon_s)\zeta_k{}' + \left(\frac{kc_s}{aH}\right)^2 \zeta_k = 0. \tag{2.7}$$

We can see that at late times, when $kc_s \ll aH$, ζ_k will freeze. This conservation of ζ outside the horizon [87] is an important property[1]. It is useful as it means that ζ_k will then be frozen during the transition from inflation to the usual radiation dominated period, and we need only pick up the evolution again once it re-enters the horizon in the well-understood radiation era.

From (2.4), in conformal time, we obtain the Mukhanov-Sasaki equation

$$\frac{\partial^2 v_k}{\partial \tau^2} + \left(c_s^2 k^2 - \frac{1}{z}\frac{d^2 z}{d\tau^2}\right) v_k = 0 \tag{2.8}$$

however, this is not ideal as c_s depends on time. Instead, let us use τ_s (1.37) as our time parameter. If we follow [19] and define

$$y_k = \sqrt{2kc_s} v_k, \tag{2.9}$$

we obtain

$$\frac{\partial^2 y_k}{\partial \tau_s^2} + \left(k^2 - 2\left(\frac{aH}{c_s}\right)^2 (1 + \ldots)\right) y_k = 0 \tag{2.10}$$

where "..." represents terms which are slow-roll suppressed.

At early times, well before crossing the sound horizon at $c_s k = aH$, we have $k \gg aH/c_s$. We can see that the equation of motion becomes

$$\frac{\partial^2 y_k}{\partial \tau_s^2} + k^2 y_k \approx 0 \tag{2.11}$$

and we can clearly see that the solutions look like $y_k \propto e^{\pm ik\tau_s}$.

[1] Systems with multiple degrees of freedom have ζ evolving even outside the horizon. See [17] for a discussion on how this relates to non-adiabatic pressure in inflation models with non-canonical kinetic terms.

2.1.2 Early time behaviour

We now wish to quantize the perturbations. We expand v as

$$v(\tau, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \left[v_k(\tau)\hat{a}_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}} + v_k^*(\tau)\hat{a}_\mathbf{k}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}} \right] \tag{2.12}$$

$$= \int \frac{d^3\mathbf{k}}{(2\pi)^3} \left[v_k(\tau)\hat{a}_\mathbf{k} + v_k^*(\tau)\hat{a}_{-\mathbf{k}}^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.13}$$

where $v_k(\tau)$ and $v_k^*(\tau)$ satisfy (2.8). The creation and annihilation operators satisfy

$$\left[\hat{a}_\mathbf{k}, \hat{a}_{\mathbf{k}'}^\dagger \right] = (2\pi)^3 \delta^{(3)}(\mathbf{k} - \mathbf{k}'). \tag{2.14}$$

Taking the commutator of v and its conjugate momentum $\pi = \frac{dv}{d\tau}$ we can obtain the canonical commutator

$$[v(\tau, \mathbf{x}), \pi(\tau, \mathbf{y})] = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \left(v_k \pi_k^* - \pi_k v_k^* \right) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \tag{2.15}$$

$$= i\delta^{(3)}(\mathbf{x} - \mathbf{y}) \tag{2.16}$$

if we demand that $v_k\pi_k^* - \pi_k v_k^* = i$. We notice that if we quantize y (defined in (2.9)), its Fourier modes should behave like a free field in Minkowski space with respect to τ_s when they are deep in the horizon. Demanding this early time behaviour of y_k gives us the Bunch-Davies vacuum

$$v_k = \frac{1}{\sqrt{2c_s k}} e^{-ik\tau_s} \tag{2.17}$$

when $\frac{c_s k}{aH} \to \infty$.

2.1.3 Late time behaviour

Taking the slow-roll parameters as constant, we can obtain a simple approximate solution to (2.7)

$$\zeta_k(\tau) = \frac{H}{\sqrt{4\varepsilon c_s k^3}} \left(1 - ik\tau_s\right) e^{ik\tau_s}. \tag{2.18}$$

Treating the solution more carefully, one finds

$$P_\zeta(k) = \frac{H^2}{4\varepsilon c_s k^3} \tag{2.19}$$

where the right-hand side is evaluated at $c_s k = aH$, defining the scale dependence of the result. This is enhanced by a factor of c_s compared to the usual canonical power spectrum. For a review and discussion of corrections see [5].

The power spectrum is typically parameterised as

$$P_\zeta(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}. \tag{2.20}$$

The scalar spectral index n_s can then be found to be

$$n_s - 1 = -2\varepsilon - \eta - \varepsilon_s \tag{2.21}$$

at leading order in slow-roll.

For completeness, we note some possible points of notational confusion. In [5] the definition $s^{there} = d\ln c_s/dN$ is used, however we refer to this quantity (as we defined in (1.38)) as ε_s. In [19], the quantity referred to as s is the integral of the sound speed with respect to conformal time—we refer to this quantity as τ_s, which we defined in (1.37). In [71], τ_s^{here} is also referred to as s, but ε_s^{here} is referred to as σ_1. This is summarised in Table 2.1.

Here	Definition	[5]	[19]	[71]
τ_s	(1.37)		s	s
ε_s	(1.38)	s		σ_1
z	(2.5)	ε/c_s^2		

Table 2.1: Notation summary and comparison for some inflationary parameters used in the literature.

2.2 The power spectrum

The fractional temperature anisotropies of the CMB can be decomposed in spherical harmonics Y_{lm} as follows

$$\frac{\Delta T}{T}(\hat{\mathbf{n}}) = \sum_{l,m} a_{lm} Y_{lm}(\hat{\mathbf{n}}). \tag{2.22}$$

If we demand the statistics are invariant under rotation, it can be shown that

$$\langle a_{lm} a^*_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l. \tag{2.23}$$

We can project the three dimensional inhomogeneities onto the CMB sky, using the Rayleigh plane wave expansion and the addition theorem for spherical harmonics. Writing $\mathbf{x} = r\hat{\mathbf{n}}$,

$$e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_l i^l (2l+1) j_l(kr) P_l(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}}) \tag{2.24}$$

$$= 4\pi \sum_{l,m} i^l j_l(kr) Y^*_{lm}(\hat{\mathbf{k}}) Y_{lm}(\hat{\mathbf{n}}) \tag{2.25}$$

where $j_l(x)$ is the spherical Bessel function. Expanding the fractional temperature fluctuations $\frac{\Delta T}{T}(\mathbf{x})$ in Fourier modes, and then expanding the factor $e^{i\mathbf{k}\cdot\mathbf{x}}$ using (2.25), we find

$$a_{lm} = 4\pi i^l \int \frac{d^3k}{(2\pi)^3} \frac{\Delta T}{T}(\mathbf{k}) j_l(k\chi_*) Y^*_{lm}(\hat{\mathbf{k}}) \tag{2.26}$$

where χ_* the comoving distance to last scattering. Then on large scales, we find

$$C_l \approx 4\pi \int d\ln k \; \frac{\mathcal{P}_\zeta(k)}{25} [j_l(k\chi_*)]^2. \tag{2.27}$$

The factor of 1/5 comes from the relation between the fractional temperature fluctuations and the primordial curvature perturbations on large scales. These are the scales that are sufficiently large that we can neglect acoustic oscillations during the radiation dominated epoch, leaving inhomogeneous gravitational redshifting (the Sachs-Wolfe effect) as the dominant contribution. The spherical Bessel function encodes the fact that a three dimensional wave of a given wavelength intersects the sphere at multiple different angular separations.

For large scales, where there was not enough time for oscillations before

recombination, the term $(j_l(k\chi_*)/5)^2$ acts as a transfer function that simply encodes the projection effects. For smaller scales where the physics is more complicated this would be encoded in some more general transfer function $\Delta_l(k)$

$$C_l = 4\pi \int d\ln k \; \mathcal{P}_\zeta(k)\Delta_l(k). \tag{2.28}$$

We can write down an estimator for this angular power spectrum. Since C_l depends only on l, we can use all the values of m for each l

$$\hat{C}_l = \frac{1}{2l+1}\sum_{m=-l}^{l} a_{lm}a_{lm}^*. \tag{2.29}$$

The variance of this estimator is given by

$$var\left(\hat{C}_l\right) = \frac{2}{2l+1}C_l^2, \tag{2.30}$$

which reflects the fact that multipoles with higher l have more independently drawn modes that we can observe.

For $\mathcal{P}_\zeta(k) = A_s$, i.e. a perfectly scale invariant power spectrum, we can use that

$$\int_0^\infty d\ln x \; j_l^2(x) = \frac{1}{2l(l+1)} \tag{2.31}$$

to evaluate the expression for the large scale power spectrum (2.27). We find that $l(l+1)C_l \propto$ const, i.e. that on large scales (small l) this quantity is independent of l.

2.3 The primordial bispectrum

2.3.1 The shape function

The primordial bispectrum is usually written as:

$$\langle \zeta_{\mathbf{k_1}}\zeta_{\mathbf{k_2}}\zeta_{\mathbf{k_3}}\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})B(k_1,k_2,k_3) \tag{2.32}$$

The delta function comes from demanding statistical homogeneity; demanding statistical isotropy restricts the remaining dependence to the magnitudes of the vectors. We denote the magnitude of $\mathbf{k_i}$ as k_i. This leaves us with a function

of three parameters, k_1, k_2, k_3. It is useful to define the dimensionless shape function

$$S(k_1, k_2, k_3) = (k_1 k_2 k_3)^2 B(k_1, k_2, k_3). \tag{2.33}$$

As we saw in Section 1.3.1 the bispectrum is only defined where the triangle condition

$$\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3} = 0, \tag{2.34}$$

is satisfied, which implies that the triangle inequality must hold

$$k_1 + k_2 \geq k_3 \text{ and cyclic perms.} \tag{2.35}$$

Thus, we will occasionally speak of the bispectrum as a function of types of triangles—for example, when the three magnitudes are equal they form an equilateral triangle, and when one is much smaller than the other two, a squeezed triangle. The space of configurations we are interested in is therefore reduced from the full cube $[k_{\min}, k_{\max}]^3$ to a *tetrapyd* (illustrated in Figure 2.1), which is defined as the intersection of that cube with the tetrahedron that satisfies (2.35). This has important implications that we will explore in Chapter 3.

The amplitude of a bispectrum shape is usually quoted in terms of some amplitude parameter f_{NL}^F. We can schematically define f_{NL}^F for some template F as follows:

$$B^F(k_1, k_2, k_3) = f_{NL}^F \times F(k_1, k_2, k_3) \tag{2.36}$$

where F contains the dependence on the k-configuration. This will be discussed further in later sections, here we will merely note that a distinction is sometimes made between the "scale dependence" of the bispectrum, meaning the dependence on $k_1 + k_2 + k_3$, and the "shape dependence", meaning the dependence on each k_i separately, for fixed $k_1 + k_2 + k_3$.

2.4 Models

We consider the same basic models as in [23]. Firstly, a quadratic potential

$$V_{\phi^2}(\phi) = \frac{1}{2} m^2 \phi^2 \tag{2.37}$$

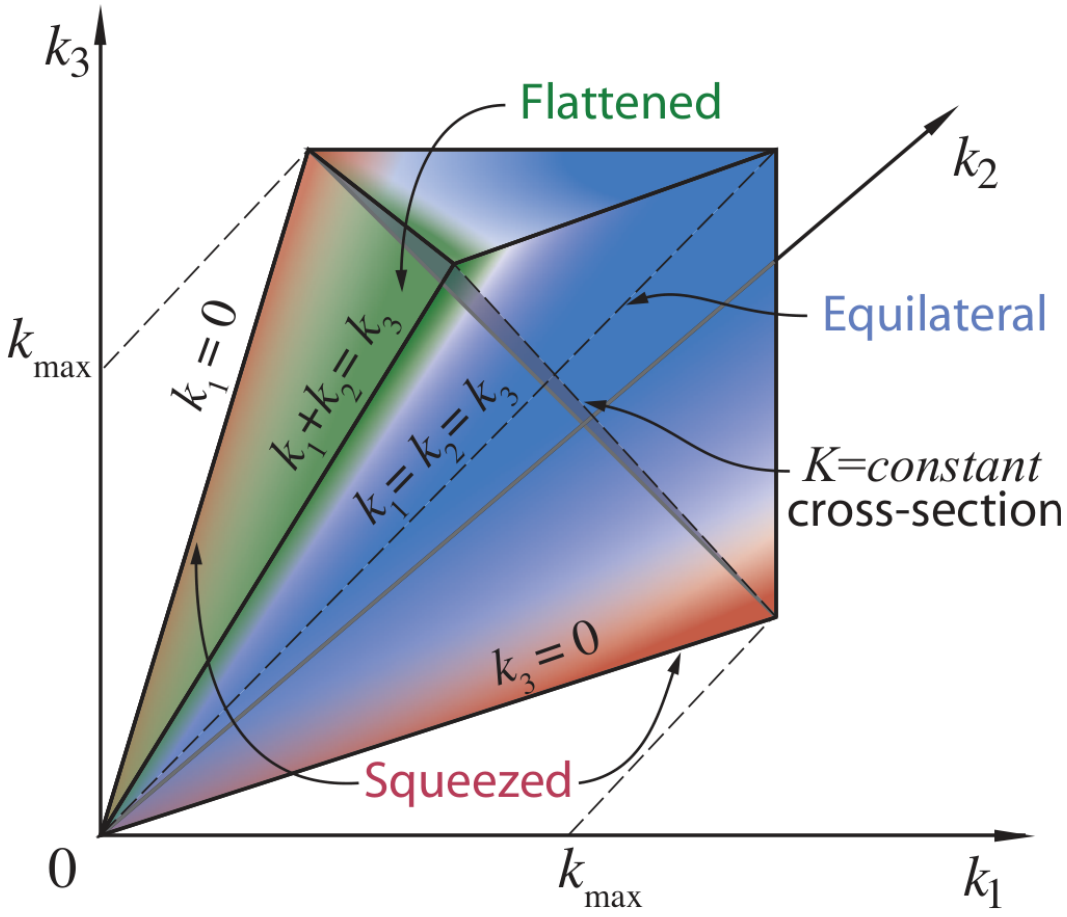


Figure 2.1: Half of the tetrapydal region on which the bispectrum is defined, along with the various limits commonly discussed in the literature. We thank Paul Shellard for providing this figure.

with a canonical kinetic term. We will also consider a non-canonical model in detail, the DBI model. The DBI model has kinetic term

$$S_{\rm DBI} = \int d^4x\sqrt{-g}\left(-\frac{1}{f(\phi)}\left((1+f(\phi)\partial_\mu\phi\partial^\mu\phi)^{\frac{1}{2}}-1\right)-V(\phi)\right), \qquad (2.38)$$

and we choose

$$f(\phi)=\frac{\lambda_{\rm DBI}}{\phi^4}, \qquad V(\phi)=V_0-\frac{1}{2}m^2\phi^2, \qquad m=\sqrt{\beta_{IR}}H_i \qquad (2.39)$$

where $\lambda_{\rm DBI}$, β_{IR} and V_0 are constants which (along with the initial value of ϕ, ϕ_0) define the scenario in question. H_i is the Hubble parameter evaluated at the initial time. This is in line with [88, 89].

The energy-momentum tensor in (1.1), in terms of the matter Lagrangian

$\mathcal{L}_m$, is given in (1.31). Following [17, 90], we can then use this to calculate the useful quantities

$$P(X,\phi) = -\frac{1}{f(\phi)}\left(\sqrt{1-2f(\phi)X}-1\right) - V(\phi) \tag{2.40}$$

$$\rho(X,\phi) = \frac{1}{f(\phi)}\left(\frac{1}{\sqrt{1-2f(\phi)X}}-1\right) + V(\phi) \tag{2.41}$$

$$c_s^2(X,\phi) = \frac{P_{,X}}{\rho_{,X}} \tag{2.42}$$

$$= 1 - 2f(\phi)X \tag{2.43}$$

We see that, as expected, in the limit $\lambda_{\rm DBI} \to 0$, the sound speed tends to unity. From these quantities, we can calculate the ϕ equation of motion (1.42)

$$\phi'' = -(3c_s^2 - \varepsilon)\phi' - \frac{3f_\phi(\phi)}{2f(\phi)}\phi'^2 + \frac{f_\phi(\phi)}{H^2 f(\phi)^2} - \left(V_\phi(\phi) + \frac{f_\phi(\phi)}{f(\phi)^2}\right)\frac{c_s^3}{H^2}. \tag{2.44}$$

See [61, 71] for further discussion, and also [91] for a brief discussion on setting consistent initial conditions.

2.5 A field guide to f_{NL}

We present here some of the ways f_{NL} is defined in the literature, for convenience. For the local form, one way of generating a non-Gaussian field is to write

$$\zeta(\mathbf{x}) = \zeta_G(\mathbf{x}) - \frac{3}{5}f_{NL}^{local}\left(\zeta_G(\mathbf{x})^2 - \left\langle\zeta_G(\mathbf{x})^2\right\rangle\right) \tag{2.45}$$

where $\zeta_G(\mathbf{x})$ is a Gaussian field. The parameter f_{NL}^{local} therefore has a clear interpretation as parametrising the deviation from Gaussianity. From this form, one obtains a bispectrum of the form (2.53). This is the version of f_{NL} that is most usually referred to.

In [92], f_{NL} is defined individually for each template, in such a way that for some template X, in the equilateral limit $k_1 = k_2 = k_3$

$$B_\Phi^X(k,k,k) = \frac{6A_\Phi^2 f_{NL}^X}{k^6}, \tag{2.46}$$

for example equation (5) in [92]. When considering a deviation from scale

invariance, the form

$$B_\Phi^X(k,k,k) = \frac{6A_\Phi^2 f_{NL}^X}{k^{8-2n_s}}, \tag{2.47}$$

is used, as in equation (6) in [92]. The parameter A_Φ comes from

$$P_\Phi(k) = \frac{A_\Phi}{k^{4-n_s}} \approx \left(\frac{3}{5}\right)^2 \frac{H^2}{4\varepsilon c_s k^{4-n_s}} \tag{2.48}$$

where we used $\Phi = \frac{3}{5}\zeta$ and the slow-roll result for $P(k)$. In [18, 66, 67] f_{NL} is similarly defined in terms of the shape function evaluated on equilateral triangles, as the templates considered had no scale dependence. One particular example is the DBI template, which we will present explicitly in (2.58). For now, we briefly state the definition of $f_{NL}^{\rm DBI}$, from [92]

$$B_\Phi^{\rm DBI}(k,k,k) = \frac{6A_\Phi^2 f_{NL}^{\rm DBI}}{k^6}. \tag{2.49}$$

Comparing to (2.58), we find that

$$f_{NL}^{\rm DBI} = -\frac{35}{108}\left(c_s^{-2} - 1\right). \tag{2.50}$$

In [5, 93] we see f_{NL} defined as the reduced bispectrum

$$f_{NL}(k_1,k_2,k_3) = \frac{5}{6}\frac{B(k_1,k_2,k_3)}{P(k_1)P(k_2)+P(k_1)P(k_3)+P(k_2)P(k_3)}, \tag{2.51}$$

which depends on k_1, k_2 and k_3. In the local case we see that this reduces to the usual definition of f_{NL}^{local} across the entire tetrapyd.

In the template-free pipeline that we present in this work, we will take f_{NL} as a fitting coefficient of our expansion of the numerically calculated primordial shape. This will be described in Chapter 5, where we will use an f_{NL} which is scenario specific, and assessing the consistency of a scenario with data will be equivalent to assessing the consistency of $f_{NL} = 1$.

2.5.1 Templates

The momentum dependence of the various possible bispectrum templates can be complex and it is therefore useful to define a standard notation in which to express this dependence. For example in [94, 95] the elementary symmetric

polynomials are used to compactly represent the results in a way that is manifestly symmetric[2]. Here we will use the notation employed in [85]:

$$\begin{aligned} K_p &= \sum_{i=1,2,3} k_i^p, \\ K_{pq} &= \frac{1}{\Delta_{pq}} \sum_{i \neq j} k_i^p k_j^q, \\ K_{prs} &= \frac{1}{\Delta_{prs}} \sum_{i \neq j \neq l} k_i^p k_j^r k_l^s, \end{aligned} \tag{2.52}$$

where Δ_{pq} is 2 if $p = q$, 1 otherwise and Δ_{prs} is 6 if $p = r = s$, 2 if $p = r \neq s$ (and permutations), and 1 if p, r, s are all distinct.

The local template is the shape which results from assuming the perturbation is quadratic in a Gaussian field (2.45)

$$S^{local}(k_1, k_2, k_3) = \frac{6}{5} A_\zeta^2 f_{NL}^{local} \left(\frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_3 k_1} + \frac{k_3^2}{k_1 k_2} \right) \tag{2.53}$$

$$= \frac{6}{5} A_\zeta^2 f_{NL}^{local} \frac{K_3}{6\, K_{111}}. \tag{2.54}$$

This template peaks on squeezed triangles, i.e. when $k_1 \approx k_2 \gg k_3$, and is used to test for multi-field effects [77]. It can also be modified to include the expected scaling

$$S^{local}(k_1, k_2, k_3) = \frac{6}{5} f_{NL}^{local} \Big(P_\zeta(k_2) P_\zeta(k_3) + P_\zeta(k_3) P_\zeta(k_1) + P_\zeta(k_1) P_\zeta(k_2) \Big) (k_1 k_2 k_3)^2. \tag{2.55}$$

Non-Gaussianity arising from inflation with a non-canonical kinetic term, and in the effective field theory of inflation [96, 97], can typically be described by the equilateral template

$$S^{equil}(k_1, k_2, k_3) = \frac{18}{5} A_\zeta^2 f_{NL}^{equil} \frac{(k_2 + k_3 - k_1)(k_3 + k_1 - k_2)(k_1 + k_2 - k_3)}{k_1 k_2 k_3} \tag{2.56}$$

This template peaks on equilateral triangles, i.e. when $k_1 \sim k_2 \sim k_3$.

[2]This representation is useful here in presenting analytic templates, however in performing our calculations we will require a one-dimensional basis, not a basis in three variables. There are also numerical advantages to the Legendre polynomials, as discussed in [94].

2.5.2 Basic shapes

We will now briefly review the shape functions that result from standard canonical inflation, and from non-canonical DBI inflation. With a canonical kinetic term the slow-roll result for the shape is

$$S^{Malda}(k_1,k_2,k_3) = -\frac{1}{32}\frac{H^4}{12\varepsilon^2}\left((3\varepsilon-2\eta)\frac{K_3}{K_{111}}+\varepsilon\left(K_{12}+8\frac{K_{22}}{K}\right)\right) \tag{2.57}$$

with $\eta = 2\varepsilon$ for the quadratic potential (2.37). At the primordial level, this is well approximated by the separable local template (2.53). However, the amplitude of this shape is expected to be tiny, since $S^{Malda} \sim \mathcal{O}(\varepsilon)\mathcal{P}_\zeta^2$ and the dominant contributions (the squeezed configurations) are expected to have no observable effect [98].

For the featureless DBI scenario, the shape function is [18]:

$$S^{\rm DBI}(k_1,k_2,k_3) = \\ -\frac{35}{108}\left(\frac{1}{c_s^2}-1\right)6\left(\frac{H^2}{4\varepsilon c_s}\right)^2\left(\frac{3}{5}\right)\left(-\frac{3}{7}\right)\frac{K_5+2K_{14}-3K_{23}+2K_{113}-8K_{122}}{K_{111}K^2} \tag{2.58}$$

to leading order in slow-roll. Since the amplitude of this shape is predicted to be far larger than (2.57) ($S^{\rm DBI} \sim \mathcal{O}(c_s^{-2})\mathcal{P}_\zeta^2$) a constraint on the magnitude of this template can be translated into one on the effective sound speed. The *Planck* analysis found a lower limit $c_s^{\rm DBI} \geq 0.087$ at 95% significance [77]. The shape (2.58) can be approximated by the separable equilateral template (2.56).

2.5.3 Scaling

These templates can be modified to be more physically realistic by including scaling similar to the scalar spectral index n_s [77]. For example, we can add some scale dependence to the DBI template (2.58) in a reasonable first approximation by including a prefactor, as was done in [92]. We define the product scaling template

$$S^{\rm DBI}_{prod}(k_1,k_2,k_3) = \left(\frac{k_1k_2k_3}{k_\star^3}\right)^{\frac{n_{NG}}{3}} S^{\rm DBI}(k_1,k_2,k_3) \tag{2.59}$$

and the sum scaling template

$$S^{\rm DBI}_{sum}(k_1,k_2,k_3) = \left(\frac{k_1+k_2+k_3}{3k_\star}\right)^{n_{NG}} S^{\rm DBI}(k_1,k_2,k_3) \tag{2.60}$$

with $n_{NG} = 2(-2\varepsilon - \varepsilon_s - \eta) - 2\varepsilon_s = 2(n_s - 1) - 2\varepsilon_s$. This improves the fit by matching the expected scaling of the shape function along the equilateral limit, which results from slow-roll corrections.

2.5.4 Shapes from features during inflation

For our more stringent validation tests we work with feature model scenarios based on the above base models. It has long been known that sharp, localised features in the inflationary potential can generate large non-Gaussianities [64], possibly with observable effects. To explore non-Gaussianity coming from such sharp features we include a kink [99]

$$V(\phi) = V_{\phi^2}(\phi)\left(1 - c\tanh\left(\frac{\phi_f - \phi}{d}\right)\right). \tag{2.61}$$

To explore non-Gaussianity from deeper in the horizon we imprint extended resonant features on the basic potential

$$V(\phi) = V_{\phi^2}(\phi)\left(1 + bf\sin\left(\frac{\phi}{f}\right)\right). \tag{2.62}$$

For more details on these models, see [65].

We now turn to feature templates that result from the preceding potential features. The result of adding a feature of the form (2.61) is to imprint oscillatory features on the bispectrum of the form

$$S^{\cos}(k_1,k_2,k_3) \approx \cos(w(k_1+k_2+k_3)) \tag{2.63}$$

though more realistically there is some phase, some shape dependence, and a modulating envelope, as detailed in [38]. The result of adding a resonant feature of the form (2.62) is to generate logarithmic oscillatory features in the shape function of the form

$$S^{\ln-\cos}(k_1,k_2,k_3) \approx \cos(w\ln(k_1+k_2+k_3)). \tag{2.64}$$

With a non-canonical kinetic term, this can also cause oscillations in the folded

limit (out-of-phase with the equilateral oscillations) as well as a modulating shape, as detailed in [73].

2.6 Step 1: the interaction Hamiltonian

2.6.1 Set-up

We begin with the action

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{2}R + P(X,\phi)\right]. \tag{2.65}$$

Note that unlike (1.30) we have explicitly included the Ricci scalar R. One proceeds by expanding this action in the perturbations (typically using the ADM formalism) to obtain the constraint equations. One can then solve the constraints (perturbatively, aided by choice of gauge) and substitute back into the action, leaving only the dynamical variables. At the level of the background one then obtains the Friedmann equations; at second order one obtains the quadratic action, which evolves the free "interaction picture" fields, as we will discuss in Section 2.7.1; and at third order, one obtains the cubic interaction Lagrangian, the source of deviations from Gaussianity.

To write the Hamiltonian one needs the momentum π, conjugate to ζ. For a simple Lagrangian such as $\mathcal{L}_0 = \frac{1}{2}\dot{\zeta}^2$, we find $\pi \equiv \frac{\partial \mathcal{L}_0}{\partial \dot{\zeta}} = \dot{\zeta}$. However if the Lagrangian contains cubic terms containing $\dot{\zeta}$, this relation will change and may not be easily invertible. However one can show that it will still be true that $\mathcal{H}_{int} = -\mathcal{L}_{int} + \mathcal{O}(4)$, so calculating the cubic interaction Hamiltonian to third order in the perturbations from the cubic interaction Lagrangian is trivial[3].

[3]Ignoring terms with no time derivatives, take $\mathcal{L} = \frac{1}{2}\dot{\zeta}^2 + g\dot{\zeta}^3$ as an example. Then, $\pi = \frac{\partial \mathcal{L}}{\partial \dot{\zeta}} = \dot{\zeta} + 3g\dot{\zeta}^2$. Calculating the Hamiltonian density

$$\mathcal{H} = \pi\dot{\zeta} - \mathcal{L} \tag{2.66}$$

$$= \pi(\pi - 3g\dot{\zeta}^2) - \frac{1}{2}(\pi - 3g\dot{\zeta}^2)^2 - g\dot{\zeta}^3 \tag{2.67}$$

$$\approx \frac{1}{2}\pi^2 - 3g\dot{\zeta}^3 + 3g\dot{\zeta}^3 - g\dot{\zeta}^3 \tag{2.68}$$

$$\approx \frac{1}{2}\pi^2 - g\dot{\zeta}^3 \tag{2.69}$$

where we have neglected terms higher order in the perturbations.

We will work with a Hamiltonian that has been split into three parts,

$$H = H_b + H_0 + H_{int}, \tag{2.70}$$

where H_b evolves the homogeneous background, H_0 is quadratic in ζ and evolves the free fields, and the interaction Hamiltonian H_{int} contains the terms cubic and above. In practice we will only use the part cubic in ζ and $\dot{\zeta}$, which can be labeled $H^{(3)}_{int}$. We will use this to calculate the higher order correlations coming from the interactions. We will occasionally drop the superscript and refer to the cubic interaction Hamiltonian as simply the interaction Hamiltonian H_{int}.

2.7 Step 2: the primordial bispectrum

2.7.1 The in-in formalism

The standard formalism for calculating higher-order correlators for models of inflation is the in-in formalism [24, 100]. For a careful discussion, see especially Appendix A2 of [100]—here, we will simply sketch the derivation. See also [101], [102–105] for discussions and examples. The setting for this calculation is the interaction picture. In the Schrödinger picture the time dependence is carried by the states, whereas in the Heisenberg picture the time dependence is carried by the operators. For notational convenience we will write the operator $\hat{H}$ as simply H. We will also write $\tilde{H} \equiv H_0 + H_{int}$ for the part of the Hamiltonian which evolves the perturbations, which is order quadratic and higher, since the linear part evolves the background fields. We can write the solutions to the Schrödinger and Heisenberg equations of motion with reference to some initial time t_i, where we take the pictures to coincide:

$$|\psi, t\rangle_S = e^{-i\int_{t_i}^{t}\tilde{H}dt}\,|\psi, t_i\rangle_S \tag{2.71}$$

$$\mathcal{O}_H(t) = e^{i\int_{t_i}^{t}\tilde{H}dt}\mathcal{O}_S e^{-i\int_{t_i}^{t}\tilde{H}dt} \tag{2.72}$$

with $\mathcal{O}_S(t) = \mathcal{O}_S(t_i)$ and $|\psi, t\rangle_H = |\psi, t_i\rangle_S$. We will be interested in the case where $\tau_i \to -\infty$. In contrast to these two pictures, the interaction picture splits the time dependence between the operators and states—the operators evolve according to the free Hamiltonian (the quadratic part) while the states see the

interaction

$$|\psi, t\rangle_I = e^{i\int_{t_i}^{t}\tilde{H}_0 dt} e^{-i\int_{t_i}^{t}\tilde{H}dt} |\psi, t_i\rangle_S \tag{2.73}$$

$$\mathcal{O}_I(t) = e^{i\int_{t_i}^{t}\tilde{H}_0 dt}\mathcal{O}_S e^{-i\int_{t_i}^{t}\tilde{H}_0 dt}. \tag{2.74}$$

We wish to calculate the equal time expectation value of some operator $\langle\mathcal{O}(t)\rangle$. To calculate this directly is difficult for a theory with interactions. In the interaction picture however, the operators evolve only according to the free Hamiltonian, so their time dependence can be more easily obtained. To take advantage of this, we would like to rewrite our desired expectation value as

$$\langle\mathcal{O}(t)\rangle = \left\langle F^{-1}(t,t_0)\mathcal{O}^I(t)F(t,t_0)\right\rangle \tag{2.75}$$

to relate the correlators of the interacting theory to the correlators of the free theory. The operator F is built out of two unitary time evolution operators

$$F(t,t_0) = U_0^{-1}(t,t_0)U(t,t_0) \tag{2.76}$$

that obey

$$U_0(t_0,t_0) = 1 = U(t_0,t_0). \tag{2.77}$$

As we will see, we will be able to roughly interpret $F(t,t_0)$ as evolving the interaction picture fields back to the initial point t_0 (where the Heisenberg and interaction pictures coincide) and then evolving the operators forward in the Heisenberg picture, giving the time evolution of the full expectation value.

The operator $U(t,t_0)$ evolves the Heisenberg picture operators in time using $\tilde{H}$

$$\delta\phi(t) = U^{-1}(t,t_0)\delta\phi(t_0)U(t,t_0) \tag{2.78}$$

$$\delta\pi(t) = U^{-1}(t,t_0)\delta\pi(t_0)U(t,t_0). \tag{2.79}$$

We can write the Heisenberg equation of motion, which has explicit time dependence in $\tilde{H}$ due to coefficients that depend on the background evolution,

$$\frac{d}{dt}\delta\phi(\mathbf{x},t) = i\left[\tilde{H}\left[\delta\phi(t),\delta\pi(t);t\right],\delta\phi\right]. \tag{2.80}$$

From this we can calculate

$$i\left[\tilde{H}\left[\delta\phi(t),\delta\pi(t);t\right],\delta\phi\right] = -U^{-1}\dot{U}U^{-1}\delta\phi(t_0)U + U^{-1}\delta\phi(t_0)\dot{U} \tag{2.81}$$

$$= -U^{-1}\dot{U}\delta\phi(t) + \delta\phi(t)U^{-1}\dot{U} \tag{2.82}$$

$$= \left[-U^{-1}\dot{U},\delta\phi(t)\right] \tag{2.83}$$

where $\dot{U} = \frac{dU}{dt}$, and we have temporarily omitted the arguments of $U(t,t_0)$. We then have

$$\frac{d}{dt}U(t,t_0) = -iU(t,t_0)\tilde{H}\left[\delta\phi(t),\delta\pi(t);t\right]U^{-1}(t,t_0)U(t,t_0) \tag{2.84}$$

$$\implies \frac{d}{dt}U(t,t_0) = -i\tilde{H}\left[\delta\phi(t_0),\delta\pi(t_0);t\right]U(t,t_0) \tag{2.85}$$

where particular attention should be paid to the evaluation time of the operators from which the $\tilde{H}$ is built.

In contrast, $U_0(t,t_0)$ evolves the interaction picture operators in time using only the part of the Hamiltonian quadratic in the perturbations

$$\delta\phi^I(t) = U_0^{-1}(t,t_0)\delta\phi^I(t_0)U_0(t,t_0) \tag{2.86}$$

$$\delta\pi^I(t) = U_0^{-1}(t,t_0)\delta\pi^I(t_0)U_0(t,t_0) \tag{2.87}$$

which by a similar calculation (recalling that $\delta\phi^I$ evolves according to the Heisenberg equation of motion with H_0) gives us

$$\frac{d}{dt}U_0(t,t_0) = -iH_0\left[\delta\phi^I(t_0),\delta\pi^I(t_0);t\right]U_0(t,t_0). \tag{2.88}$$

From (2.76), (2.85) and (2.88) we can find the equation of motion for $F(t,t_0)$

$$\frac{d}{dt}F(t,t_0) = -iH_{int}\left[\delta\phi^I(t),\delta\pi^I(t);t\right]F(t,t_0) \tag{2.89}$$

where H_{int} is cubic and higher order in the perturbations, but is now built from the interaction picture fields evaluated at t. This can be solved using the standard time ordering operator T

$$F(t,t_0) = T\exp\left(-i\int_{t_0}^{t}H_{int}(t)dt\right). \tag{2.90}$$

We have not yet accounted for the evolution of the vacuum state. We wish

the interacting vacuum $|\Omega\rangle$ to coincide with the free vacuum $|0\rangle$ in the far past. We achieve this by taking the integration contour in (2.90) to be moved slightly onto the complex plane, introducing a factor which kills off the interactions in the distant past [24]. That is, instead of taking $\tau_0 = -\infty$ we take $\tau_0 = -\infty(1 + i\epsilon)$. This is known as the $i\epsilon$ prescription. To calculate F^{-1}, we must take care to use the anti-time ordering operator, and take the lower bound of the integration to be $-\infty(1 - i\epsilon)$.

We would like to calculate the bispectrum of ζ at tree level. Expanding (2.90) in (2.75), and recalling that the expectation value of a product of three interaction picture fields vanishes

$$\langle \zeta_{\mathbf{k_1}}(\tau)\zeta_{\mathbf{k_2}}(\tau)\zeta_{\mathbf{k_3}}(\tau)\rangle = -i\int_{-\infty(1-i\varepsilon)}^{\tau} d\tau' a(\tau')\,\langle 0|\zeta_{\mathbf{k_1}}(\tau)\zeta_{\mathbf{k_2}}(\tau)\zeta_{\mathbf{k_3}}(\tau)H_{int}(\tau')|0\rangle + c.c \tag{2.91}$$

where all the operators on the right-hand side are in the interaction picture and H_{int} is the interaction Hamiltonian written in terms of ζ and its (interaction picture) conjugate momentum $\dot{\zeta}$. We will now take H_{int} to contain only terms cubic in the perturbations. From this calculation we obtain the dimensionless shape function $S(k_1, k_2, k_2)$, defined in (2.33), which is then used as input into (2.110). As an example, if one takes $H_{int} \propto \dot{\zeta}^3$, this set-up can produce the standard EFT shape

$$S(k_1, k_2, k_3) = \frac{k_1 k_2 k_3}{(k_1 + k_2 + k_3)^3}. \tag{2.92}$$

The central point, as noticed in [23], is that the integrand of (2.91) is intrinsically separable in its dependence on k_1, k_2 and k_3, and that the time integral can be done in such a way as to preserve this separability. This intrinsic separability has clearly been lost in the example in (2.92), but can be regained (to arbitrary numerical precision) by approximating it with a sum of separable terms. Our general aim will be to directly calculate this sum for a broad range of inflation models.

There is some freedom in how to represent the interaction Hamiltonian, as the equation of motion of the free fields can be used, along with integration by parts [106]. This can be used, as pointed out in [23], to avoid numerically difficult cancellations. Some presentations of this calculation use a field redefinition to eliminate terms proportional to the equation of motion from the Lagrangian. As pointed out in [5], this is unnecessary as these terms will never contribute to

the bispectrum result. In fact, in some scenarios (such as resonant models) it introduces a numerically difficult late time cancellation between a term in the interaction Hamiltonian and the correction to the correlator that adjusts for the field redefinition.

We now detail an explicit example of an in-in calculation. We will take

$$H_{int} = \int d^3x \; g(t) a^3 \dot{\zeta}^3 \tag{2.93}$$

where $g(t)$ is some function of time. This example is particularly simple as it contains no spatial derivatives. We will include those when we outline our formalism fully in Section 3.1. We wish to calculate

$$\langle \zeta_{\mathbf{k_1}}(t)\zeta_{\mathbf{k_2}}(t)\zeta_{\mathbf{k_3}}(t)\rangle = \mathrm{Re}\left(\left\langle -2i\zeta^I_{\mathbf{k_1}}(t)\zeta^I_{\mathbf{k_2}}(t)\zeta^I_{\mathbf{k_3}}(t)\int_{-\infty(1+i\varepsilon)}^{t} dt' H_{int}(t')\right\rangle\right). \tag{2.94}$$

We note that the operators inside the expectation value on the right hand side are time-ordered, as $t' \leq t$. All the operators on the right hand side are in the interaction picture, though for notational convenience we will drop the superscript I. To proceed with this calculation we will use Wick's theorem

$$\langle 0\,|\zeta_{\mathbf{k_1}}\ldots\zeta_{\mathbf{k_n}}|\,0\rangle = \left\langle 0\,\middle|:\zeta_{\mathbf{k_1}}\ldots\zeta_{\mathbf{k_n}}: + : \sum \text{ pairwise contractions}: \middle|\,0\right\rangle \tag{2.95}$$

where : : denotes normal ordering. Normal ordering places operators so that those which annihilate $|0\rangle$ are to the right and those which annihilate $\langle 0|$ are put to the left. Thus, if a normal ordered term contains uncontracted operators, the term vanishes inside an expectation value.

By "$\sum$ pairwise contractions" we mean the sum over every possible permutation of contracting one pair, and of contracting two pairs, and so on. The contraction of two operators $\zeta_{\mathbf{k_1}}$ and $\zeta_{\mathbf{k_2}}$ is defined as the difference between their time ordered product and their normal ordered product. Since the operators are built up of creation and annihilation operators, this will simply vanish, or be proportional to the commutator of $\hat{a}$ and $\hat{a}^\dagger$. In either case, it can be pulled out of the expectation value. Therefore, the only surviving contribution will the term with every $\zeta_{\mathbf{k}}$ contracted in a pair.

To evaluate the pairwise contractions we expand ζ similarly to (2.12),

$$\langle \zeta_{\mathbf{k_1}}(t_1)\zeta_{\mathbf{k_2}}(t_2)\rangle = \left\langle 0 \left| \left(\zeta_{k_1}(t_1)\hat{a}_{\mathbf{k_1}} + \zeta^*_{k_1}(t_1)\hat{a}^\dagger_{-\mathbf{k_1}}\right)\left(\zeta_{k_2}(t_2)\hat{a}_{\mathbf{k_2}} + \zeta^*_{k_2}(t_2)\hat{a}^\dagger_{-\mathbf{k_2}}\right)\right| 0\right\rangle \tag{2.96}$$

$$= \left\langle 0 \left| \zeta_{k_1}(t_1)\hat{a}_{\mathbf{k_1}}\zeta^*_{k_2}(t_2)\hat{a}^\dagger_{-\mathbf{k_2}}\right| 0\right\rangle \tag{2.97}$$

$$= \zeta_{k_1}(t_1)\zeta^*_{k_2}(t_2)\left\langle 0 \left| \left[\hat{a}_{\mathbf{k_1}}, \hat{a}^\dagger_{-\mathbf{k_2}}\right]\right| 0\right\rangle \tag{2.98}$$

$$= \zeta_{k_1}(t_1)\zeta^*_{k_2}(t_2)(2\pi)^3\delta^{(3)}(\mathbf{k_1}+\mathbf{k_2}) \tag{2.99}$$

where we used (2.14).

Inserting (2.93) into (2.94) we get

$$\langle \zeta_{\mathbf{k_1}}(t)\ \zeta_{\mathbf{k_2}}(t)\zeta_{\mathbf{k_3}}(t)\rangle \tag{2.100}$$

$$= \mathrm{Re}\left(\left\langle -2i\zeta_{\mathbf{k_1}}(t)\zeta_{\mathbf{k_2}}(t)\zeta_{\mathbf{k_3}}(t)\int_{-\infty(1+i\varepsilon)}^{t} dt' \int d^3x\ \left(g(t)a^3\dot{\zeta}^3(\mathbf{x},t')\right)\right\rangle\right). \tag{2.101}$$

Continuing, we replace ζ with its Fourier transform

$$\left\langle -2i\zeta_{\mathbf{k_1}}(t)\zeta_{\mathbf{k_2}}(t)\zeta_{\mathbf{k_3}}(t)\int_{-\infty(1+i\varepsilon)}^{t} dt' \int d^3x\ \left(g(t)a^3\dot{\zeta}^3(\mathbf{x},t')\right)\right\rangle \tag{2.102}$$

$$= \Bigg\langle -2i\zeta_{\mathbf{k_1}}(t)\zeta_{\mathbf{k_2}}(t)\zeta_{\mathbf{k_3}}(t)\int_{-\infty(1+i\varepsilon)}^{t} dt'\left(g(t)a^3\right) \cdot \int \frac{d^3p}{(2\pi)^3}\frac{d^3q}{(2\pi)^3}\frac{d^3r}{(2\pi)^3}\dot{\zeta}_{\mathbf{p}}(t')\dot{\zeta}_{\mathbf{q}}(t')\dot{\zeta}_{\mathbf{r}}(t')\int d^3x e^{i\mathbf{x}\cdot(\mathbf{p}+\mathbf{q}+\mathbf{r})}\Bigg\rangle \tag{2.103}$$

$$= \Bigg\langle -2i\zeta_{\mathbf{k_1}}(t)\zeta_{\mathbf{k_2}}(t)\zeta_{\mathbf{k_3}}(t)\int_{-\infty(1+i\varepsilon)}^{t} dt'\left(g(t)a^3\right) \cdot \int \frac{d^3p}{(2\pi)^3}\frac{d^3q}{(2\pi)^3}\frac{d^3r}{(2\pi)^3}\dot{\zeta}_{\mathbf{p}}(t')\dot{\zeta}_{\mathbf{q}}(t')\dot{\zeta}_{\mathbf{r}}(t')(2\pi)^3\delta^{(3)}(\mathbf{p}+\mathbf{q}+\mathbf{r})\Bigg\rangle. \tag{2.104}$$

We now perform the contractions. If we contract one of the $\zeta_{\mathbf{k_i}}$ with another, then the delta function will force one of $\mathbf{p}$, $\mathbf{q}$ or $\mathbf{r}$ to be zero. Since we want $\zeta_{\mathbf{k}}$ to be a perturbation and not a contribution to the background, this cannot contribute.

The results of the contractions will therefore have the form

$$\left\langle \zeta_{\mathbf{k_1}}(t)\dot{\zeta}_{\mathbf{p}}(t')\right\rangle = \zeta_{k_1}(t)\dot{\zeta}^*_p(t')(2\pi)^3\delta^{(3)}(\mathbf{k_1}+\mathbf{p}). \tag{2.105}$$

and permutations. Using the delta functions to perform the momentum integrals, switching to conformal time, and recalling (2.32), we obtain

$$\begin{aligned} B(k_1,k_1,k_3) = \mathrm{Re}\Bigg(&-2i\zeta_{k_1}(0)\zeta_{k_2}(0)\zeta_{k_3}(0) \\ &\cdot\int_{-\infty(1+i\varepsilon)}^{0} d\tau'\Big(g(\tau')a^4(\tau')\Big)\dot{\zeta}^*_{k_1}(\tau')\dot{\zeta}^*_{k_2}(\tau')\dot{\zeta}^*_{k_3}(\tau') + \text{perms}\Bigg). \end{aligned} \tag{2.106}$$

If one wanted to numerically calculate the bispectrum resulting from (2.93), one could use this form. Until this point we have assumed nothing about the solutions of the equations of motion, and we have not made a slow-roll approximation (although we have of course only included the tree-level effects). After one had chosen some prescription for numerically implementing the $i\varepsilon$ prescription, one would need to evolve a set of ζ_k according to the equations of motion and then perform the above time integral.

This is the point where our work will diverge from the standard calculation—this will be detailed in Section 3.1. For now, we merely note that the integrand of (2.106) is explicitly separable in k_1, k_2 and k_3.

We will now continue the calculation within the slow-roll approximation, using (2.18). We approximate $g(\tau')$ and the slow-roll parameters as constant, and use $a \approx -1/(H\tau)$. We see that we will need to perform the following integral by parts

$$\int_{-\infty(1+i\varepsilon)}^{0} d\tau\ \tau^2 e^{-ic_sK\tau} = \frac{-2i}{c_s^3K^3} \tag{2.107}$$

where we recall $K = k_1+k_2+k_3$. Using this, we finally obtain

$$S(k_1,k_2,k_3) = -3\frac{gH^5}{8\varepsilon^3}\frac{k_1k_2k_3}{(k_1+k_2+k_3)^3} \tag{2.108}$$

where the factor of 6 comes from the different possible contractions, all of which give the same result in this example. We can make contact with f_{NL}^{EFT2} of [16, 60] by setting $g(t) = -\frac{\varepsilon}{Hc_s^2}(1-c_s^2)\left(1+\frac{2\tilde{c}_3}{3c_s^2}\right)$ and recalling $\Phi = \frac{3}{5}\zeta$.

2.7.2 The squeezed limit consistency condition

The squeezed limit of single-field bispectra will not cause observable deviations from a Gaussian universe, due to a cancellation when switching to physical coordinates [98]. Here, we will only consider primordial phenomenology in comoving coordinates, so despite this cancellation, the squeezed limit is still a useful validation test of our results, using the standard single-field squeezed limit consistency condition [55, 107]. With $\mathbf{k_S} \equiv (\mathbf{k_2} - \mathbf{k_3})/2$:

$$S(k_1, k_2, k_3) = -\left[(n_s - 1)|_{k_S} + \mathcal{O}\left(\frac{k_1^2}{k_S^2}\right)\right] P_\zeta(k_1) P_\zeta(k_S), \quad k_1 \ll k_S \tag{2.109}$$

where $S(k_1, k_2, k_3)$ is again our dimensionless shape function. That the error in the consistency relation decreases at least quadratically in the long mode was shown in [107]. We will use (2.109) in Chapter 4 as a check on our results.

2.8 Step 3: the CMB bispectrum

In this section we will present a basic example of how the primordial bispectrum $B(k_1, k_2, k_3)$ is connected to the CMB bispectrum, which will be discussed in more detail in the next section. For the CMB, we calculate the reduced CMB bispectrum using [85]

$$b_{l_1 l_2 l_3} = \left(\frac{2}{\pi}\right)^3 \int_0^\infty dr\ r^2 \int d^3k\ (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3) \prod_{i=1}^{3} \left[j_{l_i}(k_i r) \Delta_{l_i}(k_i)\right], \tag{2.110}$$

analogously to (2.28). We see that the expression for the shape function $(k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3)$ appears directly.

We will repeat the simple example outlined in [85]. If we take $S(k_1, k_2, k_3) = 1$ for all configurations, then the above four-dimensional integral simplifies greatly. If we also restrict to large angular scales then we can use the same large-scale Sachs-Wolfe approximation[4] that was used in Section 2.2

$$\Delta_l(k) = \frac{1}{3} j_l(\chi_* k) \tag{2.111}$$

[4]The factor of 3 in this equation compared to the factor of 5 in (2.27) is due to $\zeta = 5\Phi/3$ on large scales during matter domination.

where χ_* is the comoving distance to last scattering. The integral then becomes

$$b_{l_1 l_2 l_3} = \left(\frac{2}{3\pi}\right)^3 \int_0^\infty dr r^2 \int d^3k \prod_{i=1}^{3} \left[j_{l_i}(k_i r) j_l(\chi_* k_i)\right]. \tag{2.112}$$

Using the following result

$$\int_0^\infty dk\ j_l(k) j_l(xk) = \frac{\pi}{2}\frac{x^{-(l+1)}}{2l+1} \quad \text{for } x > 1 \tag{2.113}$$

$$= \frac{\pi}{2}\frac{x^l}{2l+1} \quad \text{for } x < 1 \tag{2.114}$$

we arrive at an expression for the CMB angular bispectrum

$$b_{l_1 l_2 l_3} = \frac{f_{NL}}{27}\frac{1}{(2l_1+1)(2l_2+1)(2l_3+1)}\left[\int_0^1 dx x^{l_1+l_2+l_3+2} + \int_1^\infty dx x^{-l_1-l_2-l_3-1}\right] \tag{2.115}$$

$$= \frac{f_{NL}}{27}\frac{1}{(2l_1+1)(2l_2+1)(2l_3+1)}\left[\frac{1}{l_1+l_2+l_3+3} + \frac{1}{l_1+l_2+l_3}\right]. \tag{2.116}$$

Note also that the acoustic oscillations and other relevant physics have not been taken into account here—these would imprint oscillations at higher l. Nevertheless, it is instructive to write down the scaling behaviour of this simple example, and to see how separability is lost between (2.112) and (2.116) even for this simple case.

2.9 Step 4: CMB bispectrum estimation

2.9.1 Linear regression

The CMB bispectrum derived from a given inflationary scenario is denoted

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3}\rangle = \mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} b_{l_1 l_2 l_3} \tag{2.117}$$

where isotropy demands that $b_{l_1 l_2 l_3}$ does not depend on the m_i, and the quantity $\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3}$ is the Gaunt integral

$$\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} \equiv \int d\Omega Y_{l_1 m_1}(\hat{\mathbf{n}}) Y_{l_2 m_2}(\hat{\mathbf{n}}) Y_{l_3 m_3}(\hat{\mathbf{n}}). \tag{2.118}$$

The bispectrum measured from the CMB is denoted

$$\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} b^{obs}_{l_1 l_2 l_3} = a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}. \tag{2.119}$$

The standard method for estimating the CMB bispectrum is to calculate the least squares fit between the theory bispectrum and the observed bispectrum. That is, we find λ such that the following expression is minimised:

$$L = \sum_{l_i, m_i} \left(\lambda \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} - \frac{a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right)^2. \tag{2.120}$$

The solution to this is simply

$$\lambda = \frac{\frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \cdot \frac{a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}}}{\left| \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right|^2} \tag{2.121}$$

where $\cdot$ denotes summation over l_i and m_i. Defining the normalisation [108, 109]

$$N_B = \sum_{l_i, m_i} \frac{\left(\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} b_{l_1 l_2 l_3} \right)^2}{C_{l_1} C_{l_2} C_{l_3}} \tag{2.122}$$

we can rewrite this in the usual way (with λ identified as the result of the estimator, ϵ)

$$\epsilon = \frac{1}{N_B} \sum_{l_i, m_i} \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle \, a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}}{C_{l_1} C_{l_2} C_{l_3}}. \tag{2.123}$$

This estimator is optimal [110], so its variance saturates the Cramer-Rao bound and can be calculated theoretically.

This procedure must be modified to account for experimental noise, beam effects and the presence of a mask (to exclude to regions of the sky saturated by our galaxy and other foreground contaminants). The covariance matrix is approximated as diagonal, and the following quantities are defined

$$\tilde{C}_l \equiv b_l^2 C_l + N_l, \tag{2.124}$$

$$\tilde{b}_{l_1 l_2 l_3} \equiv b_{l_1} b_{l_2} b_{l_3} b_{l_1 l_2 l_3}, \tag{2.125}$$

to incorporate the experimental beam b_l and the noise power spectrum N_l from

the experiment. The estimator can then be written as [108, 109, 111]

$$\epsilon = \frac{1}{N_B^2} \sum_{l_i m_i} \frac{\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} \bar{b}_{l_1 l_2 l_3}}{\tilde{C}_{l_1} \tilde{C}_{l_2} \tilde{C}_{l_3}} \left(a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} - 3 C^{sim}_{l_1 m_1, l_2 m_2} a_{l_3 m_3} \right) \tag{2.126}$$

which retains optimality [111]. Here, the diagonal approximation is insufficient for the linear term (which is especially important for local type shapes), so the full covariance matrix $C^{sim}_{l_1 m_1, l_2 m_2}$ is approximated using an ensemble average over a suite of Gaussian simulations. Excluding this term would result in a large false detection of non-Gaussianity, due to the effects of masking the part of the sky obscured by the galaxy. While the estimator is still optimal, Monte Carlo simulations are also used to estimate the variance, to ensure systematics are taken into account correctly. Many CMB skies are generated from Gaussian initial conditions, and then the estimator is applied to each. This gives a distribution of ϵ, from which we can calculate 1σ and 2σ regions for ϵ. Then the result of the estimator (when applied to the real CMB sky) can be compared to this distribution to determine the significance of the result.

2.9.2 Complexity of bispectrum estimation

To calculate the bispectrum of temperature fluctuations in the CMB, one uses transfer functions (which we saw previously in (2.28)) to evolve and project the primordial bispectrum onto our sky. In principle, this is the same process as power spectrum estimation. However, for the bispectrum the computational challenge is far greater, requiring both compute-intensive and large in-memory components. As a result of this complexity, this step is computationally impractical for generic primordial bispectra. Progress can be made by finding an approximation to the primordial shape that is separable, and using this simplification to make the calculation tractable through the *KSW* estimator [59, 76, 112]. For example, one may find that a particular inflation scenario generates a primordial bispectrum with a high correlation with some standard shape, then look at how well that standard shape is constrained by the CMB. The Modal decomposition method of [84–86] leveraged these simplifications in a more structured way for generic bispectra, broadening the range of constrained models.

The measure of non-Gaussianity in the CMB that is most usually quoted is f_{NL}, referring to f_{NL}^{local}. This number describes how well a particular template, the local template, describes the correlations in the CMB; this template is used as a proxy for the class of inflation models that produce similar bispectra. Similar

quantities for the equilateral and orthogonal templates are also commonly quoted. In addition to broadening the range of constrained models through increases in efficiency, the Modal decomposition method of [84–86] allows to go beyond this paradigm, efficiently constraining inflationary templates in the CMB using all of the shape information; essentially constraining an f_{NL} specific to a given template. This bypasses the approximation step at the level of the templates, of finding a separable approximation to the primordial template. The results of these methods are constraints on the parameters of certain inflation models through the approximate phenomenological templates. These constraints can be found in [60, 77].

The numbers f_{NL}^F are useful summary parameters. From the data-side, they represent the result of a complex and intensive process of estimating the amplitude of the template F, given some data. From the theory-side, one can use them to take an inflation scenario and compare it to that data, if one can find a standard template with a high correlation with the shape resulting from that scenario. However, despite its usefulness, this paradigm does have drawbacks. It acts as an information bottleneck, losing some constraining power when one approximates the real shape function by some standard template. In particular, if one is interested in a feature model, it may be be difficult to see how constraints on existing features can be applied.

Recall the definition of separability (1.56). The link between the separability of the primordial bispectrum and the reduced CMB bispectrum can be seen from (2.110). If the primordial bispectrum is separable then the overall dimension of the calculation can be reduced from seven to five in (2.110), since the spherical Bessel functions j_{l_i} and the transfer functions Δ_{l_i} already appear in a separable way. This property can also be used to efficiently generate non-Gaussian initial conditions for simulations [83].

KSW-type estimators take as their starting point the realisation that many templates can be rewritten as a simple finite sum of separable functions. For example, the simple local (2.53), equilateral (2.56) and orthogonal [92] templates can be built up from a small set of simple monomials, namely k^{-1}, 1, k and k^2. This method is able to constrain shapes with that contain very high frequency linear oscillations[5], using a Fourier basis instead of monomials. It is limited in that it cannot function for shapes whose shape dependence is not sum-separable, however.

[5]Since $e^{iw(k_1+k_2+k_3)} = e^{iwk_1}e^{iwk_2}e^{iwk_3}$.

Since separability is so vital, the usual strategy is to approximate non-separable templates by separable ones. For example, the (2.58) is closely approximated by the equilateral template (2.56). Much success has been had in constraining non-Gaussianity in the CMB using separable approximations to these approximate templates. Other methods target oscillations [113, 114], by expanding the shape function in $k_1 + k_2 + k_3$, thus limiting their ability to capture shapes whose phase varies across the tetrapyd.

The Modal method [115] is more versatile, leveraging the benefits of separability in a broader set of models through expansion. For example, this method can deal with primordial templates with envelopes. The Modal code proceeds using two sets of basis functions, one in k space at the end of inflation, and another in l space for the CMB. For a given primordial template, the best-fit linear combination of the primordial basis is found, generating a set of coefficients. The template need not be a sum of separable functions, but should be well-approximated by this decomposition in the primordial basis. The coefficients of this decomposition are then projected onto the CMB basis, to calculate the constraint.

2.10 Previous work on in-in separability.

In [23] it was pointed out that one can compute the inflationary bispectrum using the tree-level in-in formalism in such a way as to preserve its intrinsic separability. In addition to making this point, [23] lays out some of the basic structure of an implementation of that computation, and validates the method on simple, featureless scenarios. This work built on the philosophy of [84–86] in which a formalism was developed to leverage the tractability of separable CMB bispectrum estimation for generic primordial bispectra, by expanding them in a separable basis. The idea of [23] is an extension of that philosophy to the primordial level, and our work is in expanding that methodology so that it can be applied to interesting examples. In [84–86] an orthogonal basis on the tetrapyd was used, removing the need to fit non-physical configurations. One of the main differences between that work and this is that we cannot use this basis here without sacrificing the in-in separability we are trying to preserve.

In this work we explore the details of this calculation in much greater detail than was considered in [23]. We restructure the methods, improving on the work of [23] in terms of flexibility of basis choice and efficiency of the calculation.

We also detail a particular set of basis functions that improves upon those described in [23] in its rate of convergence, its transparency, and its flexibility. We do this without sacrificing orthogonality. This is detailed in Chapter 3. Our improvements over the methods sketched in [23] allow us to validate on non-trivial bispectra for the first time, including sharp deviations from slow-roll, which we present in Section 4.7. We quote our results in terms of a measure that is easier to interpret than the correlation defined in [23], and that includes the magnitude as well as the shape information on the full tetrapyd. This is discussed in Section 3.2.

Some work has also been done on basis sets for three-dimensional bispectra [116–118], in the context of the tetrapyd after inflation.

2.11 Configuration-by-configuration codes

Previous work on the numerical calculations of inflationary non-Gaussianity include the BINGO code [119], Chen et al [64, 65], the work of Horner et al [120–122] and the Transport Method [93, 123–125]. All but the last directly apply the tree-level in-in formalism k-configuration by k-configuration for a given model; they integrate a product of three mode functions and a background-dependent term from the interaction Hamiltonian, of form similar to (3.3). The eventual result is a grid of points representing the primordial bispectrum.

The most advanced publicly released code for the calculation of inflationary perturbations is based on the Transport Method. Like the other mentioned work it calculates the bispectrum k-configuration by k-configuration. However the method is different in its details. Instead of performing integrals, a set of coupled ODEs is set up and solved. The power spectra and bispectra themselves are evolved, their time derivatives calculated by differentiating the in-in formalism[6]. The publicly released code is very sophisticated, able to deal with multiple fields in curved field spaces, recently being used to explore the bispectra resulting from sidetracked inflation [75].

[6]One could imagine applying the same philosophy to our method. Certainly, at first sight this seems more natural, that if the core quantities in our method are the coefficients in some basis expansion, why not evolve them directly? Why take the apparently circuitous route (that we will see in Chapter 3 and Chapter 4) of evolving the $\zeta_k(\tau)$, and decomposing them at every timestep? The answer is that the "equations of motion" for the coefficients of the expansion obtained by substituting a mode expansion of $\zeta_k(\tau)$ into (2.7) are coupled in an infinite hierarchy, making this a difficult direction to pursue. This is, however, a basis dependent statement and the possibility remains that a basis may be found to ameliorate this problem.

2.11.1 Usage in recent works

In [57, 75] and especially in [126], a large ensemble of inflationary trajectories is considered using the transport method [124]. For these trajectories, the bispectrum is calculated for a certain number of k-configurations. In these works, f_{NL}^{equil} and f_{NL}^{flat} refer to the definition given in (2.51) evaluated on equilateral and flat configurations respectively. It was found that squeezed limit configurations were much more expensive to calculate using that method.

While this analysis is very sophisticated, all configuration-by-configuration methods face the same problems: firstly, that calculating enough points in the bispectrum to ensure that the whole picture has been captured is expensive, especially for non-trivial features. Even once that has been achieved, what is obtained is a grid of points. Since this is not separable, it must be processed further to be usefully compared to observation.

2.12 CMB-BEst

The method implemented in CMB-BEst [2] is roughly a generalisation of the *KSW* method. Contrasting the Modal method, CMB-BEst requires only one basis, a primordial one, in k space. The size of the basis is referred to as $p_{\max}$. This basis is then projected forwards onto the CMB. This is a massively resource-intensive calculation with the naïve implementation scaling as $p_{\max}^6$. It also requires terabytes of RAM for reasonable method parameters. This calculation was tackled in [2], which implemented optimisations which allowed this calculation to be run for $p_{\max} = 30$ using reasonable supercomputer resources. The motivation for implementing this algorithmically and numerically difficult calculation is that it need only be done once per basis. Once it is done, then any primordial bispectrum represented in that basis can be constrained in the CMB immediately.

CMB-BEst is a generalisation of the *KSW* method in the sense that for a specific choice of basis, it simplifies to the *KSW* method. However, by choosing a more descriptive basis set (for example, one of the basis sets we have developed in Chapter 3) a much broader range of models can be constrained. This was one of our main motivations in carefully exploring and comparing the basis sets in Chapter 3, as more descriptive basis sets translate directly into more constraints on inflationary models.

In the methods implemented in CMB-BEst, the covariance matrix in the

linear term of (2.126) is not assumed to be diagonal, as this was found to lead to inaccurate results. It is calculated by taking an ensemble average of 140 Gaussian simulations. f_{NL} estimates are also made for all of these simulations, with the variance of this distribution providing a measure to assess the significance of any f_{NL} detection in the true CMB.

The convergence of the final estimate is always the primary arbiter for setting method parameters, even at the primordial level. It was shown in [2] that the ratio of $k_{\rm max}$ to $k_{\rm min}$ was important in determining the success of the method. If $k_{\rm max}/k_{\rm min}$ was too small then important information was lost, and validation estimates of f_{NL} for the local shape did not agree with previous methods. In that work it was determined that $k_{\rm max}/k_{\rm min} = 1000$ was sufficient, hence that is the value we will use in testing our basis sets.

Chapter 3

Decomposing Primordial Shapes

As we have seen, at tree level the bispectrum is an integral over a separable integrand (see (2.91) and the example presented in (2.106)). This separability is usually lost in the final result. In this chapter we outline a formalism that preserves the separability, by expanding the integrand in a separable basis. We also explore possible choices for this separable basis, testing them on various shape functions from the literature and highlighting the choices with sufficiently fast and broad convergence to be useful in realistic inflationary parameter scans.

3.1 Setting up the formalism

Given its separable form, the tree-level in-in formalism is amenable to more efficient calculation using separable modes, as first mentioned in [23]. That work extended the separable methodology previously implemented for the CMB bispectrum [84–86]. Our goal in this work is the efficient calculation of more general bispectra which may have significant (possibly oscillatory) features, requiring searches across inflationary parameters. To achieve this, we represent the shape function (2.33) using a set of basis functions as

$$S(k_1, k_2, k_3) = \sum_n \alpha_n \, Q_n(k_1, k_2, k_3) \,, \tag{3.1}$$

where the basis functions $Q_n(k_1, k_2, k_3)$ are explicitly separable functions of their arguments. We shall consider constructing the separable basis functions $Q_n(k_1, k_2, k_3)$ out of triplet products of normalized one-dimensional modes $q_p(k)$

as

$$Q_n(k_1, k_2, k_3) \equiv q_p(k_1)\, q_r(k_2)\, q_s(k_3)\,. \tag{3.2}$$

Here, n labels the integer triplet $n \leftrightarrow \{prs\}$ in some appropriate manner. This modal expansion is terminated at some $p_{\rm max}$ for which $\max(p, r, s) < p_{\rm max}$. The specific bispectrum information of each scenario is encoded in α_n.

Translating this result into a constraint from the CMB will require a large once-off computational cost—however, this cost will be paid once per set of basis functions Q_n, not per scenario. The details of this once-per-basis calculation will be presented in [2]. As such, while the general computational steps we describe will be independent of the basis, it is vital we explore possible sets of basis functions $Q_n(k_1, k_2, k_3)$ and their effects on convergence; that will be the main subject of this chapter. In this initial section we will set the notation we will use to recast the standard numerical in-in calculation into a calculation of α_n in (3.1). We will sketch the steps involved, including accounting for the effect of spatial derivatives in the interaction Hamiltonian on our final result. The rest of this chapter is then devoted to developing and comparing possible basis sets. In Chapter 4 we will then use the formalism and the basis sets of this chapter to make precise the numerical considerations of the calculation, especially our methods of dealing with the high-frequency oscillations at early times.

The values of the coefficients α_n will depend on the choice of basis, but the methods we will describe to calculate them will be practical for any basis. Our aim will be to separate out the dependence on k and τ_s, without losing any of the information contained in the tree-level in-in formalism, except in the sense that is controlled by $p_{\rm max}$, the size of the basis. We will set up an efficient numerical implementation of the calculation, a necessary consideration to allow this method to be useful in exploring parameter spaces in primordial phenomenology. Throughout we will see that we are able to preserve the separability of the dependence on k_1, k_2 and k_3.

The tree-level in-in formalism for the bispectrum (2.91) is inherently separable given the form of the cubic interaction Hamiltonian $H_{\rm int}$. This can be clearly seen in the integrand of the example in (2.106). Consider indexing with $i = 1, 2, 3...$ the interaction vertices in $H_{\rm int}$, so then after the contractions have been performed, the bispectrum (2.91) can be expressed as a sum over separable contributions of the form:

$$S(k_1,k_2,k_3) = \sum_i I^{(i)}(k_1,k_2,k_3)$$
$$= \mathrm{Re}\sum_i \Bigg[v^{(i)}(k_1,k_2,k_3) \int_{-\infty(1-i\varepsilon)}^{0} d\tau\, w^{(i)}(\tau) F^{(i)}(\tau,k_1)\, G^{(i)}(\tau,k_2)\, J^{(i)}(\tau,k_3)$$
$$+ \text{cyclic perms} \Bigg] \tag{3.3}$$

where $w^{(i)}(\tau)$ is a function of the scale factor and the other background parameters (1.38) for the i-th interaction vertex, while the terms $F^{(i)}, G^{(i)}, J^{(i)}$ are given by the Fourier mode functions $k^2\zeta_k(0)\zeta_k^*(\tau)$ or their time derivatives $k^2\zeta_k(0)\dot{\zeta}_k^*(\tau)$. Spatial derivative terms also separate because of the triangle condition (2.34). For example, a term such as $\partial_i\zeta\,\partial_i\zeta$ becomes $(-\mathbf{k_2}\cdot\mathbf{k_3})\zeta_{k_2}(\tau)\,\zeta_{k_3}(\tau)$. Since $\mathbf{k_2}\cdot\mathbf{k_3} = (k_1^2 - k_2^2 - k_3^2)/2$, this yields a sum of separable terms. These time-independent contributions are contained in $v^{(i)}(k_1,k_2,k_3)$, as they do not force us to compute extra time integrals. Note that $v^{(i)}(k_1,k_2,k_3)$ need not be symmetric in its arguments.

We can connect this form to the example given in (2.106). In that example, we see there is no sum over i as we took only one term of H_{int}. We also see that $v^{(i)}(k_1,k_2,k_3) = 1$ as there are no spatial derivatives. Collecting the terms with no k dependence, we see that $w^{(i)}(\tau) = -2ig(\tau)a^4(\tau)$. The remaining terms (for this example) then give

$$F^{(i)}(\tau,k) = G^{(i)}(\tau,k) = J^{(i)}(\tau,k) = k^2\zeta_k(0)\dot{\zeta}_k^*(\tau) \tag{3.4}$$

where the k^2 comes from the definition of the shape function (2.33), and we obtain $\zeta_k(\tau)$ by numerically solving the equation of motion (2.7).

Using the approximate mode functions (2.18), another explicit example is the second interaction term in (4.8), i.e. $H_{\rm int}^{(1)} = \zeta'^2\zeta$. It takes the form

$$F^{(1)}(\tau,k) = G^{(1)}(\tau,k) \;=\; k^2c_s^2\tau\frac{H^2}{4\varepsilon c_s k}e^{-ik\tau_s}\,, \tag{3.5}$$
$$J^{(1)}(\tau,k) \;=\; (1+ik\tau_s)\frac{H^2}{4\varepsilon c_s k}e^{-ik\tau_s}\,. \tag{3.6}$$

In this slow-roll approximation, such terms in (3.3) are straightforward to integrate analytically (using the $i\varepsilon$ prescription), provided the time-dependence of the slow-roll parameters and the sound speed is neglected [24]. However, for high

precision bispectrum predictions we must incorporate the full time-dependence, while solving (2.7) to find accurate mode functions $\zeta_{\mathbf{k}}(\tau)$ numerically. Obtaining the full 3D bispectrum at high resolution using this direct method is computationally demanding however, because it requires repetitive integration of (3.3) at each specific configuration of the wavenumbers (k_1, k_2, k_3), a problem which is drastically compounded by bispectrum parameter searches e.g. for oscillatory models.

The terms contained in $v^{(i)}(k_1, k_2, k_3)$ depend on the structure of the spatial derivatives in the interaction Hamiltonian, but not the specific scenario. These terms are separable; we will discuss their precise form in Section 4.5. We include their contribution to the final result after the time integrals have been computed. In contrast, the factors which depend only on time ($w_i(\tau)$) depend on the scenario but do not need to be decomposed in k. The remaining factors have both k and time dependence; they must be decomposed in k at every timestep. As we have mentioned, these terms look like $F^{(i)}(k, \tau) = k^2 \zeta_{\mathbf{k}}(0)\zeta^*_{\mathbf{k}}(\tau)$, with a possible time derivative. The factor of k^2 could be absorbed into $v^{(i)}(k_1, k_2, k_3)$, but we have the freedom to keep it here to aid convergence.

If the expressions being expanded have some known pathology in their k-dependence, we can then see two ways of dealing with this. The basis can be augmented to efficiently capture the relevant behaviour (as we will explore in the rest of this chapter) or the behaviour can be absorbed into the analytic prefactor, $v^{(i)}(k_1, k_2, k_3)$.[1] We use the former, as the numerics of the latter are less transparent and less physically motivated.

The internal basis used for the decomposition at each timestep need not match that which is used for the final result. Indeed, in dealing with the spatial derivatives we will find it useful to change to a different basis than the one used to perform the time integrals of the decompositions—we will discuss this further later in this section.

We will now link (3.3) to (3.1). Consider representing the primordial shape function $S(k_1, k_2, k_3)$ in (3.3) as a mode expansion for each interaction term $I^{(i)}(k_1, k_2, k_3)$ as

$$S(k_1, k_2, k_3) = \sum_i I^{(i)}(k_1, k_2, k_3) = \sum_i \sum_n \alpha_n^{(i)} Q_n(k_1, k_2, k_3)\,, \qquad (3.7)$$

where $Q_n(k_1, k_2, k_3)$ is separable, built out of some orthonormal set $q_p(k)$ as

[1] At early times the modes are highly oscillatory in both k and τ_s, to which we will certainly give special attention.

in (3.2). Armed with this set of modes, we can expand any of the interaction terms $F^{(i)}(\tau,k)$, $G^{(i)}(\tau,k)$, $J^{(i)}(\tau,k)$ in (3.3) as:

$$F^{(i)}(\tau,k) = \sum_p f_p^{(i)}(\tau)\, q_p(k)\,, \tag{3.8}$$

$$\text{where} \quad f_p^{(i)}(\tau) = \int_{k_{\min}}^{k_{\max}} dk\, F^{(i)}(\tau,k)\, q_p(k)\,. \tag{3.9}$$

Substituting these expansions into (3.3), we obtain the following decomposition for the i-th vertex contribution,

$$\begin{aligned} I^{(i)}(k_1,k_2,k_3) &= \mathrm{Re}\left[v^{(i)}(k_1,k_2,k_3) \int_{-\infty(1-i\varepsilon)}^{0} d\tau\, w^{(i)}(\tau) \right. \\ &\left. \cdot \sum_p f_p^{(i)}(\tau) q_p(k_1) \sum_r g_r^{(i)}(\tau) q_r(k_2) \sum_s h_s^{(i)}(\tau) q_s(k_3) + \text{cyclic perms} \right] \qquad (3.10) \\ &= v^{(i)}(k_1,k_2,k_3) \sum_{prs} \left(\mathrm{Re} \int_{-\infty}^{0} d\tau\, w^{(i)}(\tau) \right. \\ &\left. \cdot\, f_p^{(i)}(\tau)\, g_r^{(i)}(\tau)\, h_s^{(i)}(\tau) \right) q_p(k_1)\, q_r(k_2)\, q_s(k_3) + \text{cyclic perms} \qquad (3.11) \end{aligned}$$

where the dependence on k_i has now been brought outside the time integral. For the sake of compactness of notation we will now use P to stand for the triplet p,r,s and $\tilde{P}$ to stand for the triplet $\tilde{p},\tilde{r},\tilde{s}$. Note that we have dropped the $i\varepsilon$ prescription—this is due to the fact that the integrals are now convergent, as pointed out in [23], and as we discuss in Section 4.4. Writing $q_P(k_1,k_2,k_3) = q_p(k_1)\, q_r(k_2)\, q_s(k_3)$, we continue,

$$\begin{aligned} I^{(i)}(k_1,k_2,k_3) &= v^{(i)}(k_1,k_2,k_3) \sum_P \tilde{\alpha}_P^{(i)}\, q_P(k_1,k_2,k_3) + \text{cyclic perms} \\ &= \sum_P \tilde{\alpha}_P^{(i)} \sum_{\tilde{P}} V_{P\tilde{P}}^{(i)}\, q_{\tilde{P}}(k_1,k_2,k_3) + \text{cyclic perms} \\ &= \sum_{\tilde{P}} \alpha_{\tilde{P}}^{(i)}\, q_{\tilde{P}}(k_1,k_2,k_3) + \text{cyclic perms}\,, \end{aligned} \tag{3.12}$$

where we have written

$$\tilde{\alpha}_P^{(i)} = \tilde{\alpha}_{prs}^{(i)} = \mathrm{Re} \int_{-\infty}^{0} d\tau\, w^{(i)}(\tau)\, f_p^{(i)}(\tau)\, g_r^{(i)}(\tau)\, h_s^{(i)}(\tau)\,, \tag{3.13}$$

and included the time-independent k-prefactors from the interaction Hamiltonian

by writing

$$v^{(i)}(k_1, k_2, k_3) q_P(k_1, k_2, k_3) = \sum_{\tilde{P}} V^{(i)}_{P\tilde{P}} q_{\tilde{P}}(k_1, k_2, k_3)\,, \tag{3.14}$$

and

$$\alpha^{(i)}_P = \sum_{\tilde{P}} \tilde{\alpha}^{(i)}_{\tilde{P}} V^{(i)}_{\tilde{P}P} + \text{cyclic perms.} \tag{3.15}$$

The numerical calculation of $V^{(i)}_{P\tilde{P}}$ (as defined by (3.14)) is highly efficient (despite naïvely scaling as p^6_{max}) because $v^{(i)}(k_1, k_2, k_3)$ is a sum of separable terms. The details of these terms depend only on the spatial derivatives in the interaction Hamiltonian, not the scenario being considered, so the matrix can be precomputed and stored. Note that this is not the only way one can organise this calculation to explicitly preserve the separability. One could also include the contributions coming from the spatial derivatives first, decomposing (as in (3.8)) not only terms like $k^2 \zeta_{\mathbf{k}}(0) \zeta^*_{\mathbf{k}}(\tau)$, but also terms that include each power of k_1, k_2 or k_3 that appears in $v(k_1, k_2, k_3)$. The index i in the sum in (3.7) would then run over not only each vertex in the interaction Hamiltonian, but also each separable term within those vertices. We do not choose this path as, for the sake of efficiency, we wish to minimise the number of time integrals of the form (3.13) we need to calculate.

Note the basis sets on the left and right hand side of (3.14) need not match. In fact, if those two basis sets do match, then generically information will be lost—for example, if the basis set on the left is the Legendre polynomials up to order n, then terms in $v^{(i)}(k_1, k_2, k_3)$ with positive powers will introduce higher order dependencies on k, and negative powers will introduce $1/k$ behaviour. In practice, to prevent this loss of information, we take the basis set on the right hand side of (3.14) to be an expanded version of that on the left. For example, if the left hand basis was the Legendre polynomials up to order n, the right hand basis would include polynomials up to order $n + 2$, and also include a term to capture $1/k$ behaviour. This will be discussed in detail later in this chapter.

We can see from (3.13) that the number of time integrals needed is controlled by $N_V \times {p^3_{\text{max}}}^2$, where N_V is the number of interaction vertices and p_{max} is the size of the basis. Since the calculational cost of doing the internal decompositions

[2] In fact the number is not quite p^3_{max}. Since we have extracted the spatial derivatives, the only remaining possible source of asymmetric k-dependence comes from ζ^3, $\zeta^2\zeta'$, $\zeta'^2\zeta$ or ζ'^3 so the time integral in (3.13) will always be (at least) symmetric in p and r.

depends only linearly on the size of internal basis, improvements there are dwarfed by improvements gained from reducing the number of terms needed in the final basis.

Having calculated the contribution of each H_{int} vertex separately, indexed as above by (i), the overall shape function (3.7) is then simply

$$S(k_1, k_2, k_3) = \sum_n \left(\sum_i \alpha_n^{(i)} \right) Q_n(k_1, k_2, k_3) = \sum_n \alpha_n \, Q_n(k_1, k_2, k_3) \,. \qquad (3.16)$$

Depending on the scenario, some vertex contributions will converge faster than others or be completely negligible; for efficiency the maximum modal resolution defined by $p_{\rm max}$ can be allowed to be different for each vertex.

It is worth emphasising that the raison d'etre for this approach is that all time integrals (3.13) are now independent of the k-configuration[3]. In a configuration-by-configuration method one improves the precision by decreasing the spacing which defines the density of the grid of points within the tetrapyd. Instead, in the modal approach, we increase precision by adding more modes to the shape function expansion (3.16) until the result converges at high precision. At first sight, this appears to increase the dimensionality of the calculation. Directly integrating the in-in formalism requires one time integration for each k-configuration, i.e. N_k^3 integrals, ignoring symmetry. The method detailed here requires decomposing the modes, then a time integral for every coefficient, i.e. $p_{\rm max}^3$ integrals (again ignoring symmetry) in addition to the decomposition. However for every model we have explored from the literature, our expansion in $p_{\rm max}$ converges far faster than in the number of k-modes that would be required to have confidence in a sampled bispectrum. This is clear in smooth bispectra such as (2.57) and (2.58), but is also true of bispectra with complicated features, as quantified by the figures later in this chapter.

To be efficiently connected to a late-time observable a sampled bispectrum would have to be fit by a smooth template, a complication that is automatically taken care of in this formalism. We note that the primordial basis can be chosen independently of the final bispectrum basis employed for observational tests—a change of basis $Q_{pqr} \to \tilde{Q}_{pqr}$ can be trivially and cheaply achieved through a linear transformation Γ with the new expansion coefficients given by $\tilde{\alpha}_{abc} = \Gamma_{ap}\Gamma_{bq}\Gamma_{cr}\alpha_{pqr}$.

[3]They are not independent of $k_{\rm min}$ and $k_{\rm max}$ which define the domain of interest, which is analogous to the coefficients of a Taylor expansion depending on its expansion point.

Discussion of convergence in this chapter is considered only at the primordial level, with no concept of the signal to noise of an actual experiment. There could be a basis that converges faster in some observationally weighted sense, efficiently describing the primordial modes which will matter most at late times. We leave discussion of this point to a later work, as converting between the two, after the in-in computation is completed, is trivial.

Having now set our notation and outlined the calculation, in the following sections we discuss possible basis sets, before presenting the details on the numerical implementation of these methods in Chapter 4.

3.2 Testing basis sets

We define the inner product of two bispectrum shape functions as

$$S_1 \cdot S_2 = \langle S_1 \, , S_2 \rangle = \int_{T_k} d^3k \; S_1(k_1, k_2, k_3) \; S_2(k_1, k_2, k_3) \, , \tag{3.17}$$

where T_k refers to the tetrapyd, the region of the cube $[k_{\rm min}, k_{\rm max}]^3$ that obeys the triangle inequality. Following [127] we define the two correlators:

$$\mathcal{S}(S_1, S_2) = \frac{S_1 \cdot S_2}{\sqrt{(S_1 \cdot S_1)(S_2 \cdot S_2)}} \, , \qquad \mathcal{A}(S_1, S_2) = \sqrt{\frac{S_1 \cdot S_1}{S_2 \cdot S_2}} \, . \tag{3.18}$$

Here, we refer to $\mathcal{S}(S_1, S_2)$ as the shape correlator between the two bispectra; $\mathcal{A}(S_1, S_2)$ is the amplitude correlator. In principle, we could add some observationally motivated weighting to the above measure, as considered in [84–86], but in this work we restrict ourselves to accurately calculating the full primordial bispectra, weighting each configuration equally.

Writing $|S|^2 = S \cdot S$, we can then re-express a measure of the relative error between one bispectrum template and another:

$$\begin{aligned} \mathcal{E}(S_1, S_2) &= \sqrt{\frac{|S_1 - S_2|^2}{|S_2|^2}} = \sqrt{\frac{|S_1|^2 - 2S_1 \cdot S_2 + |S_2|^2}{|S_2|^2}} \\ &= \sqrt{\mathcal{A}(S_1, S_2)^2 - 2\mathcal{A}(S_1, S_2)\mathcal{S}(S_1, S_2) + 1}. \end{aligned} \tag{3.19}$$

This error measure takes into account differences in overall magnitude as well as shape. If we are only interested in comparing the differences coming from the

shape, we can scale the bispectra so that $\mathcal{A}(S_1, S_2) = 1$ and so

$$\mathcal{E}^{(\mathcal{A}=1)}(S_1, S_2) = \sqrt{2(1 - \mathcal{S}(S_1, S_2))}. \tag{3.20}$$

With this measure of relative difference, a shape correlation of 0.9 corresponds to a relative difference of 45%, a shape correlation of 0.99 corresponds to a relative difference of 14%, a shape correlation of 0.999 corresponds to a relative difference of 4%. Thus (for our purposes in calculating feature bispectra) this more exacting measure $\mathcal{E}$ from [127] is a far better representation of actual convergence between two shape functions than the correlation used in [23], as it is easier to interpret and more stringent.

The full relative difference (3.19) also includes the amplitude in its measure, which will be important in obtaining an precise link between fundamental inflationary parameters and the resulting primordial bispectrum. We will use this measure to test the accuracy and efficiency of our basis expansion in reconstructing the standard templates, and later to quantify the convergence of our validation examples in Section 4.7. In that section we also plot residuals on slices through the tetrapyd, relative to the representative value

$$S^{rep} = \sqrt{\frac{S \cdot S}{\int_{T_k} d^3k}}. \tag{3.21}$$

One degree of freedom that we have not exploited is weighting the decomposition to optimise the convergence of the final observable. This could be advantageous as it is possible that the CMB could be more sensitive to certain k-configurations than others, and so we would like those configurations to converge most efficiently. In [85] a good approximation was found for the appropriate weighting on the tetrapyd. In this work however we focus on calculating the primordial shape function accurately and efficiently with a flat weighting, and leave exploiting this freedom to a future work.

An important part of this work is testing the expected convergence of our various basis sets on templates of primordial shapes, before testing them in the setting of the in-in formalism. This testing is important as when testing in the context of the full in-in formalism it can be difficult to distinguish between genuine lack of convergence, and numerical errors coming from other sources. By testing on templates we can estimate the optimal possible convergence for a given shape, once all physical effects that contribute in the tree-level in-in calculation

are taken into account. Since the feasibility of our method depends on being able to efficiently capture interesting and realistic shapes, this will determine which basis sets are worth implementing and testing in the in-in setting.

3.3 Building basis sets

3.3.1 Basis set building blocks

Our general strategy will be to start with the Legendre polynomials or the Fourier basis functions as a foundation for our basis set. We will augment this set with additional basis functions motivated by the expected behaviour of primordial shapes in general. In this way we can quickly capture the known behaviours, but still have the flexibility to converge to a wide range of shape functions. It also has the advantage that the Legendre polynomials and the Fourier basis functions are automatically orthogonal, so we will only have to use the modified Gram–Schmidt process (which brings in numerical difficulties) a minimal number of times.

The Legendre polynomials are a basis set with broad descriptive power. The Fourier basis functions also have broad descriptive power, but are limited by converging poorly to non-periodic functions. This problem can be ameliorated through augmenting the basis with extra basis functions.

We augment our basis sets with extra (orthogonalised) basis functions. We build these using the more numerically stable modified Gram-Schmidt process, orthogonalising the new function with respect to all the basis elements already in the basis.

3.3.2 Basis choices

We now begin our discussion of specific possible sets of separable basis functions $Q_n(k_1, k_2, k_3)$ for use in the expansion (3.1). Whether the goal is to explore primordial phenomenology or for direct comparison with observations, the convergence of our basis set will determine the efficiency and practicality of our methods. Given the basis-agnostic formalism we outlined in 3.1 and the basis-agnostic methods we will outline in Chapter 4 we are free to choose our set of basis functions in (**??**) to optimise for efficient convergence, ensuring our results are useful for comparison with observations. There are a wide variety of options available, such as polynomial bases or Fourier series, that can be chosen for the

$q_p(k)$. While not strictly necessary for the method, it is more convenient if the resulting 3D basis functions $Q_n(k_1, k_2, k_3)$ are orthogonal on the cubic region of selected wavenumbers, making it much more straightforward to obtain controlled convergence. Overall, then, rapid convergence is the key criterion in choosing the basis functions $q_p(k)$ in (3.2), thus determining the nature of the numerical errors in the calculated bispectrum. However, since we are going beyond the featureless examples of [23] this matter deserves considerable care and close attention. Ideally we would have a three-dimensional basis that can efficiently capture a wide variety of shapes on the tetrapyd, with relatively few modes. In this work we aim for basis functions that work well in a wide variety of scenarios, so we endeavour to use as little specific information as possible (e.g. guessing the frequency of bispectrum oscillations from the power spectrum of a given scenario), though we will allow ourselves to use a representative value of the scalar spectral index, n_s^*. It is worth emphasising that a major advantage of the flexibility of the basis in the methods detailed in the following sections is the ease with which the basis can be modified to yield drastic increases in the rate of convergence at the primordial level, for the purposes of exploring primordial phenomenology.

In this section we will use some standard templates to investigate different possible sets of basis functions. An important issue is that when leveraging the separability of the in-in formalism, we are essentially forced to expand the shape function on the entire cube $[k_{\rm min}, k_{\rm max}]^3$. This is because the only decomposition we actually perform is a one-dimensional integral over $[k_{\rm min}, k_{\rm max}]$ (as we saw in (3.9)). With a uniform weighting, this integral does not know anything about the distinction between the tetrapyd and the cube. This is important as it means the non-physical configurations outside the tetrapyd will affect the convergence of our result on the tetrapyd, the region where we require efficient convergence. To mimic this in testing our sets of basis functions, each shape will be decomposed on the entire cube, but the quoted measures of convergence will be between the shape and its reconstruction on the tetrapyd only (unless stated otherwise).

For a shape like (2.57) the non-physical off-tetrapyd configurations will not have a large effect, as the bispectrum on the faces of the cube is comparable to the bispectrum in the squeezed limit of the tetrapyd. On the other hand, for a shape of the equilateral type such as (2.58), this effect can be disastrous if not handled properly. This can be easily seen from (2.56), in the limit of small k_3. The triangle condition in that limit enforces $(k_2 - k_1)^2 \leq k_3^2$. This implies that $0 \leq k_3^2 - (k_2 - k_1)^2 \leq k_3^2$, forcing the shape to go to zero in that limit despite the

k_3 in the denominator. On the non-physical part of the face, $k_2 - k_1$ is not small, and so the shape is boosted by $1/k_3$ relative to the equilateral configurations. These regions then dominate any attempted basis expansion. To overcome this problem, as we shall discuss, we will extend our basis to explicitly include this $1/k$ behaviour[4].

One useful starting choice for modal bispectrum expansions has been the shifted, normalised Legendre polynomials $P_r(x)$:

$$q_r(k) = \left(\frac{2r+1}{k_{\rm max} - k_{\rm min}}\right)^{1/2} P_r(\bar{k})\,, \tag{3.22}$$

with a rescaling of the argument $\bar{k}$ to ensure the wavenumber k falls within the chosen (observable) domain $k_{\rm min} < k < k_{\rm max}$, that is,

$$\bar{k} = \frac{2k - k_{\rm max} - k_{\rm min}}{k_{\rm max} - k_{\rm min}}\,. \tag{3.23}$$

There is freedom to vary this mapping, which we shall exploit in Section 3.5. We shall label as $\mathcal{P}_0$ the basis function set of pure Legendre polynomials in (3.22), with $r = 0, 1, \ldots, p_{\rm max}-1$. These were considered also in [23], however, while they prove to be particularly functional building blocks for other modal applications, in the context of the in-in formalism they converge so slowly even for simple shapes as to be inadequate when taken on their own. This poor rate of convergence with $\mathcal{P}_0$ for two local- and equilateral-type shapes is shown in Figure 3.1. This is due to the $1/k$ behaviour inherent in these shapes, which is compounded in the equilateral models by pathologies exterior to the tetrapyd, as we have discussed.

The two basis function sets actually used in [23] to calculate primordial bispectra were as follows. The first was the Legendre polynomials taken with a

[4]There are results in the literature that describe generic $K = k_1+k_2+k_3$ poles in correlators—see for example [128]. A simple example can be understood by recalling that in standard calculations using the in-in formalism, the $i\varepsilon$ prescription is used to damp out contributions in the infinite past. This does not work for $K = 0$. While the resulting divergence (in K) is clearly outside the physical region of the tetrapyd, we will see its effects in the physical configurations. Given that this three-dimensional behaviour is generic, one might worry that we should take more care in building it into our one-dimensional basis. However, the excellent convergence in Section 4.7 shows that $\mathcal{P}_1$ and $\mathcal{P}_{01}^{n_s}$ can capture this behaviour well, and that this worry is unwarranted. In fact, since this behaviour comes from the oscillations at early times, observing this behaviour is a useful check on our results.

log-mapping between k and the polynomial argument as

$$q_r(k) = \left(\frac{2r+1}{\ln k_{\rm max} - \ln k_{\rm min}}\right)^{1/2} P_r\left(\overline{\ln k}\right), \tag{3.24}$$

$$\overline{\ln k} = \frac{2\ln k - \ln k_{\rm max} - \ln k_{\rm min}}{\ln k_{\rm max} - \ln k_{\rm min}}. \tag{3.25}$$

The second basis was implicitly mentioned in a reference to the possibility of multiplying the functions to be decomposed by k, and dividing that factor out when evaluating the result. In our language, this is equivalent to working with an unnormalised basis set of the Legendre polynomials divided by k:

$$q_r(k) = \left(\frac{2r+1}{k_{\rm max} - k_{\rm min}}\right)^{1/2} \frac{P_r(\bar{k})}{k}, \tag{3.26}$$

where the rescaled $\bar{k}$ is defined in (3.23). This can also be thought of as expanding the bispectrum $k_1k_2k_3\,S(k_1,k_2,k_3)$ in $\mathcal{P}_0$, instead of the shape function $S(k_1,k_2,k_3)$ itself. The consequence is that neither (3.24) nor (3.26) are orthogonal with respect to the flat weighting of the inner product (3.17). However, as shown in [23], these two basis sets (3.24) and (3.26) are able to approximate the three canonical bispectrum shapes. Nevertheless, our aim is to go beyond the featureless examples investigated in [23], so we require a basis that can capture many different forms of bispectrum features. To this end, we prefer not to weight the large or small wavelengths in our fit, as is done in (3.24) and (3.26). The deciding factor for which weighting is optimal to include in the primordial inner product is information about which configurations are most important for observables, that is, the expected signal-to-noise. We will not discuss this matter in detail here, except to note that motivated by the form of (2.110), we will take as our aim the accurate calculation of the primordial shape function with a flat weighting. Based on this motivation, we will not pursue (3.24) and (3.26) any further.

One could certainly also consider sets of basis functions more tailored to a particular example, or indeed even use power spectrum information to, on the fly, generate a basis tailored to a rough form of the expected bispectrum features. We save this possibility for future work. While it may be useful in the context of purely primordial phenomenology, connecting to CMB-BEst requires one broadly useful basis.

Our general strategy will be to augment these basic building blocks with

a small number of extra basis elements, while retaining orthogonality, using the standard modified Gram-Schmidt process. We will now introduce notation that we will use to describe the construction of our basis sets. We begin with some initial set of orthonormal basis functions, q_r, with $r = 0, \ldots, N_{initial} - 1$ (in practice these will usually be the Legendre polynomials or the Fourier basis functions). If we then wish to use some function f to this initial set, which we will denote here by B_0, then we define

$$\tilde{f}(k) \;=\; \mathrm{Orth}\left[f(k),\; B_0\right] \;\equiv\; f(k) - \sum_{q \in B_0} \frac{\langle f, q\rangle}{\langle q, q\rangle} q(k) \tag{3.27}$$

and add $\tilde{f}$ to our basis set. We note that the inner product here $\langle f, g\rangle$ is the 1D integral of the product $f(k)g(k)$ from $k_{\min}$ to $k_{\max}$. The resulting basis is orthogonal, provided sufficient care is taken to avoid numerical errors. When we augment an initial basis by multiple extra functions, each function is added in to the set one at a time, and orthogonalised with respect to all the functions already in the set.

Motivated by the form of the basic templates, we can introduce a basis function to capture the $1/k$ behaviour. We define $\mathcal{P}_1$ such that the first $p_{\max} - 1$ basis functions match $\mathcal{P}_0^{p_{\max}-1}$, and the final basis function is

$$q_{p_{\max}-1}(k) = \mathrm{Orth}\left[1/k,\; \mathcal{P}_0^{p_{\max}-1}\right]. \tag{3.28}$$

As we see in Figure 3.1, the convergence properties for the augmented basis $\mathcal{P}_1$ are dramatically improved.

In addition to our Legendre basis functions, the pure $\mathcal{P}_0$ basis and the augmented $\mathcal{P}_1$ basis, we will also introduce a Fourier series basis denoted by $\mathcal{F}_0$ and defined by

$$q_0(k) = 1\,, \quad q_{2r-1}(k) = \sin(\pi r \bar{k})\,, \quad q_{2r}(k) = \cos(\pi r \bar{k})\,, \quad 1 \le r \le (p_{\max}-3)/2 \tag{3.29}$$

augmented by k and then by k^2 for a total basis size of $p_{\max}$. The basic Fourier series have to be augmented by the linear k and quadratic k^2 terms in order to satisfactorily approximate equilateral shapes (reflecting in part the preference for periodic functions). As with $\mathcal{P}_1$ defined in (3.28), we will similarly create an augmented Fourier basis $\mathcal{F}_1$ by adding the inverse $1/k$ term to the $\mathcal{F}_0$ basis, i.e. using $q_{p_{\max}}(k) = \mathrm{Orth}\left[1/k,\; \mathcal{F}_0\right]$ with (3.27). When we refer to convergence, we

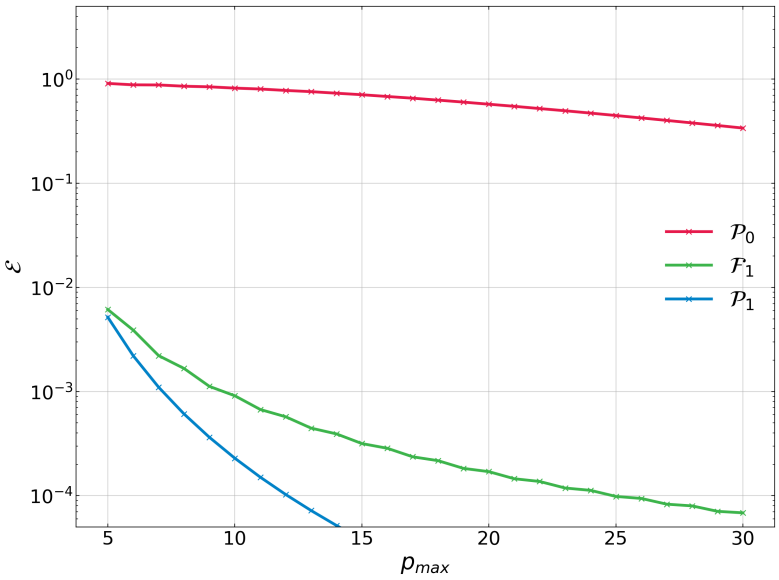
mean in increasing number of Legendre polynomials (or sines and cosines) within the initial set. The total size of the set will always be referred to as $p_{\max}$.

Notation	**Building Blocks**	**Augmented by**	**Definition**
$\mathcal{P}_0$	Legendre polynomials		(3.22)
$\mathcal{F}_0$	Fourier Series	k, k^2	(3.29)
$\mathcal{P}_1$	Legendre polynomials	k^{-1}	
$\mathcal{F}_1$	Fourier Series	k, k^2, k^{-1}	
$\mathcal{P}_1^{n_s}$	Legendre polynomials	$k^{-1+(n_s^*-1)}$	(3.30)
$\mathcal{P}_{01}^{n_s}$	Legendre polynomials	$k^{n_s^*-1}$, $k^{-1+(n_s^*-1)}$	(3.31)
scaling	Legendre polynomials	k^{-1}, $\ln(k)k^{-1}$	
resonant	$\frac{P_n(\bar{k})}{\sqrt{k}}$, with $\bar{k} = \frac{2\ln(k)-\ln(k_{\min}k_{\max})}{\ln(k_{\max})-\ln(k_{\min})}$		

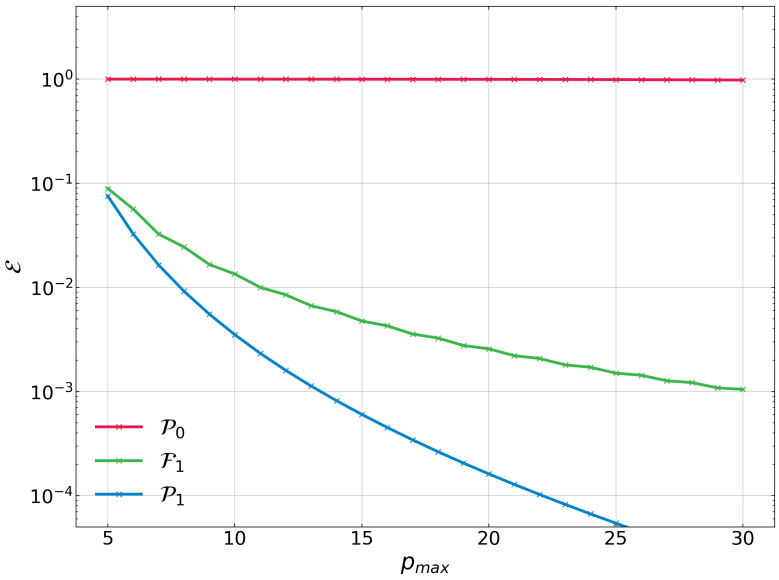
Table 3.1: Basis summary—the augmentation of the basis is done using (3.27). The size of each basis is referred to as $p_{\max}$. Some of these basis sets are plotted in figures 3.7 and 3.8.

In order to compare the efficacy of the different basis function sets, we have investigated their convergence on Maldacena's shape (2.57) and the DBI shape (2.58). To mimic the in-in calculation, we expand the shape on the cube, but test the result on the tetrapyd using (3.19). The results are shown in Figure 3.1. We find that the Legendre polynomials basis set $\mathcal{P}_0$ converges so slowly as to be unusable (with the Fourier modes $\mathcal{F}_0$ worse and not plotted). However, the augmented Legendre basis $\mathcal{P}_1$ (including $1/k$) leads to rapid convergence with an improvement of four orders of magnitude at $p_{\max} = 15$. The augmented Fourier basis $\mathcal{F}_1$ also converges quickly relative to $\mathcal{P}_0$, but is outdone by $\mathcal{P}_1$. Though we do not show the convergence on the cube, we find that for Maldacena's template this is of the same order of magnitude as the error on the tetrapyd. For the DBI shape, however, the fit on the tetrapyd lags significantly behind, due to the effect of the large non-physical configurations. This explains the order of magnitude difference between the convergence at each $p_{\max}$ for the two shapes in Figure 3.1.

Next, we investigate oscillatory model templates. The simple feature model (2.63) and the resonance model (2.64) have scale dependence, but no shape dependence (in that they only depend on the perimeter of the triangle, $K = k_1 + k_2 + k_3$). We test our sets of basis functions on these two shapes, and also when they are multiplied by (2.58) to obtain a feature template with both shape and scale dependence. As shown in Figure 3.2, $\mathcal{F}_0$ naturally outperform the basis sets built from Legendre modes for a pure oscillation. However when the equilateral-type



(a) Reconstructing the Maldacena Template



(b) Reconstructing the DBI Template

Figure 3.1: Convergence comparisons for the Legendre and Fourier basis functions for (a) the Maldacena template (2.57) and (b) the DBI template (2.58). The pure Legendre $\mathcal{P}_0$ basis requires many terms to fit the $1/k$ behaviour in both Maldacena's template (2.57) and the DBI template (2.58). In contrast, the $\mathcal{P}_1$ basis (with an orthogonalised $1/k$ term) mitigates this dramatically, with the error already reduced by a factor of 100 at $p_{\rm max} = 5$. The Fourier $\mathcal{F}_1$ basis performs well, but converges more slowly than the $\mathcal{P}_1$ basis. Note that the convergence errors for (2.58) are larger than (2.57) because of the larger contributions outside the tetrapyd dominating the fit. In this plot and the following, unless otherwise stated, $k_{\rm max} = 1000 k_{\rm min}$.

DBI template (2.58) is superimposed, even the augmented Fourier modes $\mathcal{F}_1$ converge poorly. Instead, the basis sets built from Legendre modes offer a better more robust option, with the *scaling* basis performing best. In Figure 3.3 we see that for a logarithmic oscillation, the *resonant* basis converges in the fewest modes.

Finally, we consider convergence in the light of the more subtle scale-dependence due to the spectral index n_s of the power spectrum. The simple canonical examples in Figure 3.1 had shape dependence and no scale dependence, but this would only be expected of scenarios unrealistically deep in the slow-roll limit. When we include this scale dependence, using (2.59) with n_s^*, it proves very useful to include these deviations from integer power laws in the basis functions. We will now consider two cases, first augmenting $\mathcal{P}_0$ by a scale-dependent $1/k$ term using the orthogonalisation procedure (3.27),

$$q_{p_{\max}-1}(k) = \text{Orth}\left[k^{-1+(n_s^*-1)},\ \mathcal{P}_0\right], \tag{3.30}$$

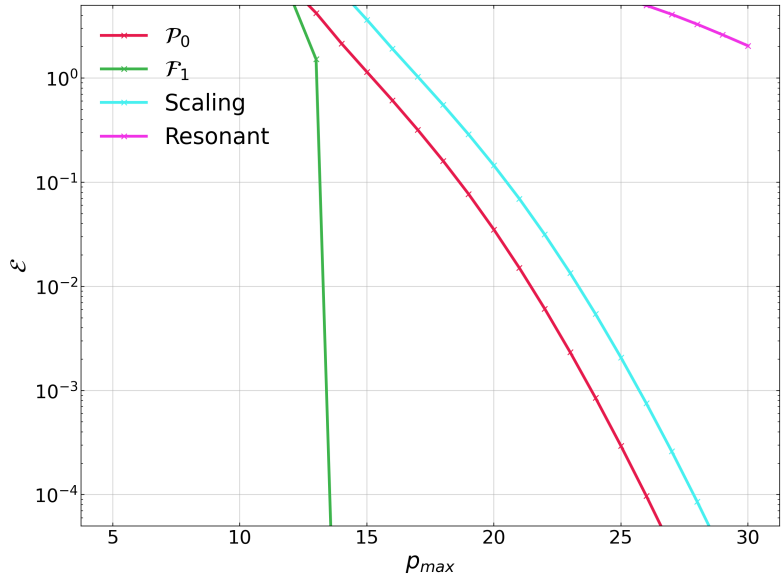
which we refer to as $\mathcal{P}_1^{n_s}$. Secondly, we instead augment $\mathcal{P}_0$ with an additional scale-dependent terms

$$k^{(n_s^*-1)}, \qquad k^{-1+(n_s^*-1)} \tag{3.31}$$

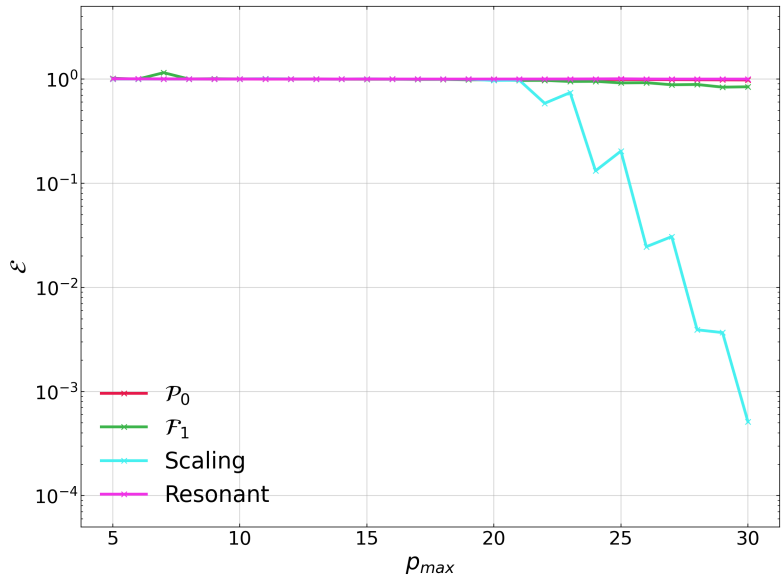
which we refer to as $\mathcal{P}_{01}^{n_s}$.

It is worth noting that while $\mathcal{P}_0$ with (e.g.) $p_{\max} = 20$ is a subset of $\mathcal{P}_0$ with any higher $p_{\max}$, $\mathcal{P}_1^{n_s}$ with (e.g.) $p_{\max} = 20$ is not a subset of $\mathcal{P}_1^{n_s}$ with any higher $p_{\max}$ as the augmented term will be orthogonalised with respect to a different set of functions.

As we see in Figure 3.4, for equilateral type shapes even a small overall scale dependence causes significant degradation in the convergence of the original augmented Legendre basis $\mathcal{P}_1$. However, incorporating the spectral index n_s into the basis functions $\mathcal{P}_1^{n_s}$ and $\mathcal{P}_{01}^{n_s}$ results again in rapid convergence to the scale-dependent DBI template, which can then be accurately approximated with a limited number of modes. We conclude using power spectrum information to augment the basis functions with terms incorporating the expected dependence on the spectral index enables the efficient approximation of high precision primordial bispectra. This is still not optimal, however, as the scaling of the bispectrum is not uniquely defined by the scaling of the power spectrum, so more flexibility would be desirable—we will explore this further in Section 3.4.

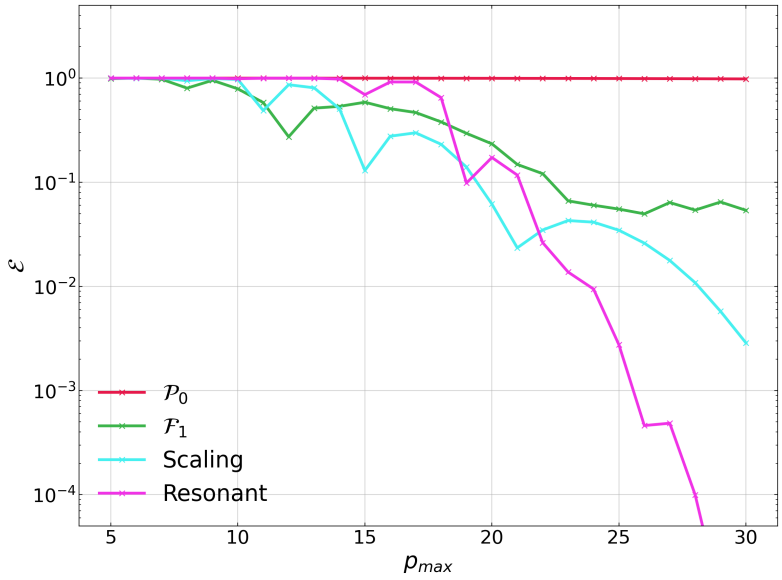


(a) $\cos(f(k_1 + k_2 + k_3))$

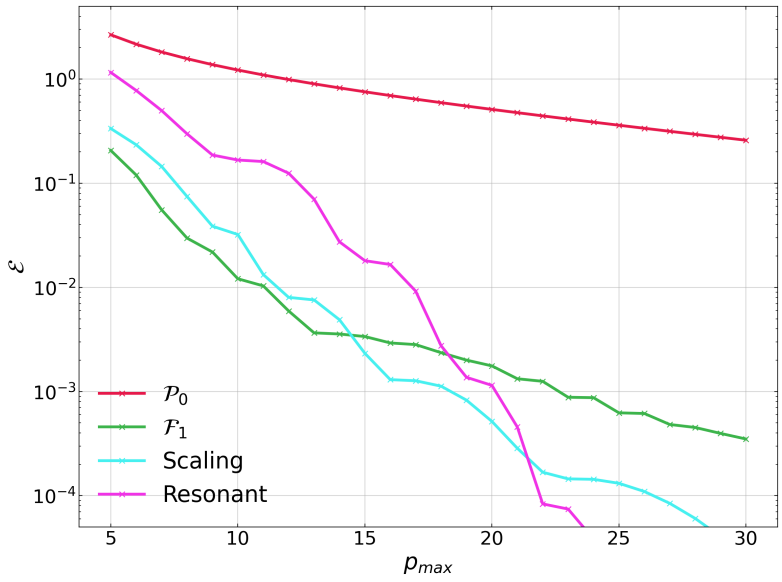


(b) $\cos(f(k_1 + k_2 + k_3))S^{DBI}$

Figure 3.2: Convergence comparison for oscillatory models. For linear oscillations we choose $f = 150.64$ (so we obtain 15 whole oscillations in the k-range). (a) As expected, the $\mathcal{F}_1$ basis fits an oscillation with no shape dependence (2.63) (that is periodic in the k-range) perfectly. For this special case, the $\mathcal{P}_0$, *scaling* and *resonant* sets of basis functions require more modes to accurately describe the shape. (b) However, moving to the more complex and realistic case of a feature with scale and shape dependence (in this case the product of (2.58) and (2.63)), we see that again the *scaling* basis converges with the fewest modes. Note that before the expansion has fully converged, the fit on the tetrapyd can actually degrade slightly when the basis set is extended. This is an artifact of fitting on the cube and restricting (3.19) to the physical configurations on the tetrapyd; when considered over the entire cube the fit improves monotonically. The *resonant* basis, naturally, does not converge well to a linear oscillation.

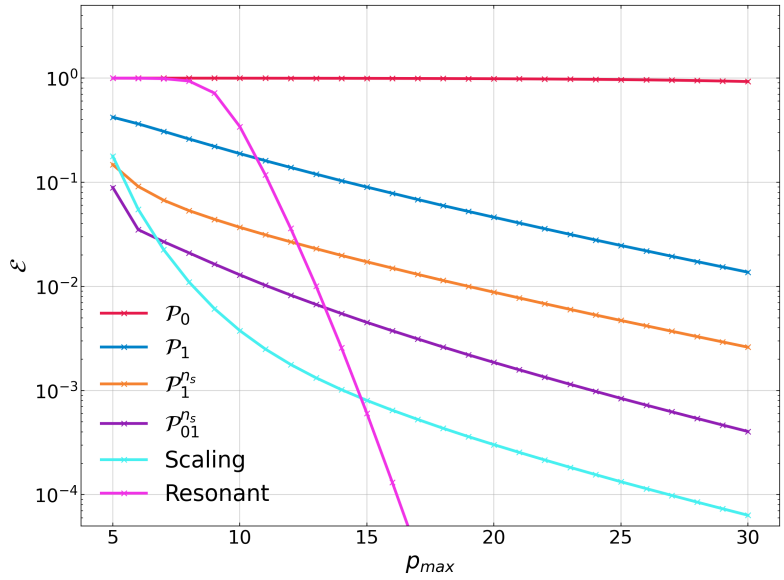


(a) $\cos(f\log(k_1+k_2+k_3))S^{DBI}$ on the tetrapyd

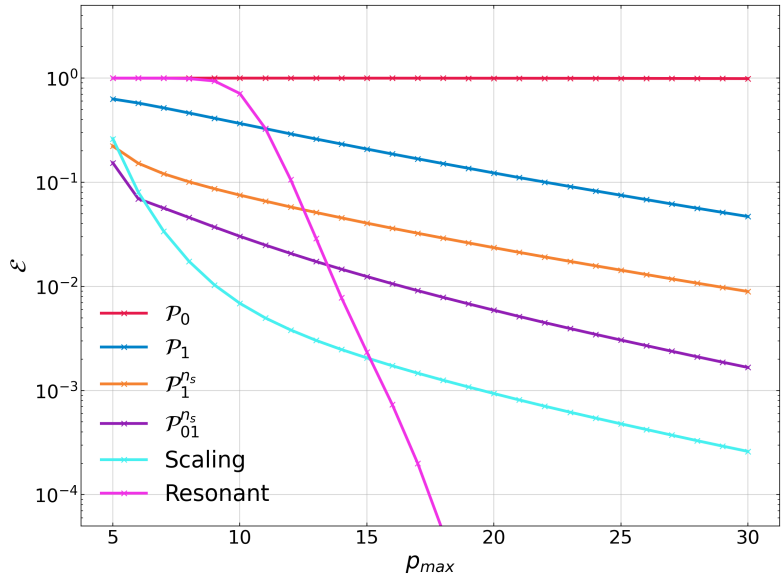


(b) $\cos(f\log(k_1+k_2+k_3))S^{DBI}$ on the cube

Figure 3.3: We plot the convergence of various basis sets for the more complex case of a resonant feature with scale and shape dependence, in this case the product of (2.58) and (2.64). We choose $f = 4.55$ (so we obtain 5 whole oscillations in the k-range). (a) On the tetrapyd, we see that the *scaling* basis struggles to converge for this fixed-frequency template, due to the combination of the high frequency logarithmic oscillation and the non-trivial shape dependence coming from $S^{DBI}(k_1, k_2, k_3)$. As expected for logarithmic oscillations, the *resonant* basis performs best for this template, confirming that the basis functions defined in (3.5) perform well even when there is non-trivial shape dependence. (b) On the cube, we see that the results converge orders of magnitude faster—see discussion in Section 3.3.2.

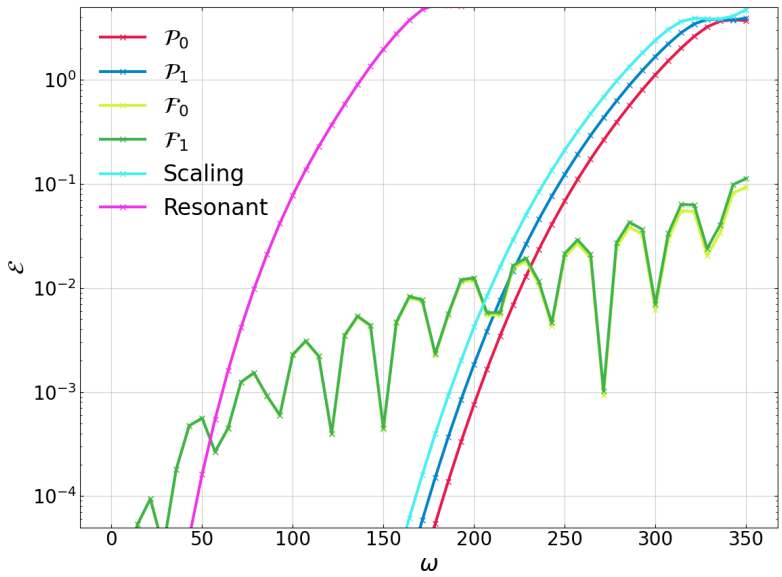


(a) $k_{\rm max}/k_{\rm min} = 550$

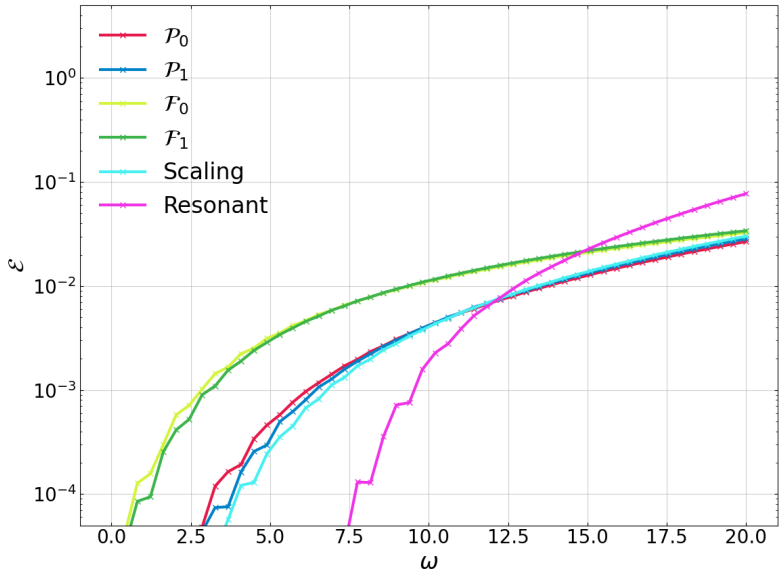


(b) $k_{\rm max}/k_{\rm min} = 1000$

Figure 3.4: For the scale-dependent DBI template (2.59), by including a minimal amount of power spectrum information using (3.30) and (3.31) ($\mathcal{P}_1^{n_s}$ with $n_s^* = 0.9649$), we can decrease the convergence error by nearly an order of magnitude (compared to $\mathcal{P}_1$), reaching sub-percent errors at $p_{\rm max} = 30$. The *scaling* basis performs far better again, even without the extra power spectrum information. By far the fastest convergence, however, results from the *resonant* basis. The convergence of this template also has a dependence on $k_{\rm max}/k_{\rm min}$. The convergence power of these basis sets will allow us to efficiently capture the scale dependence of the numerically calculated shape function.

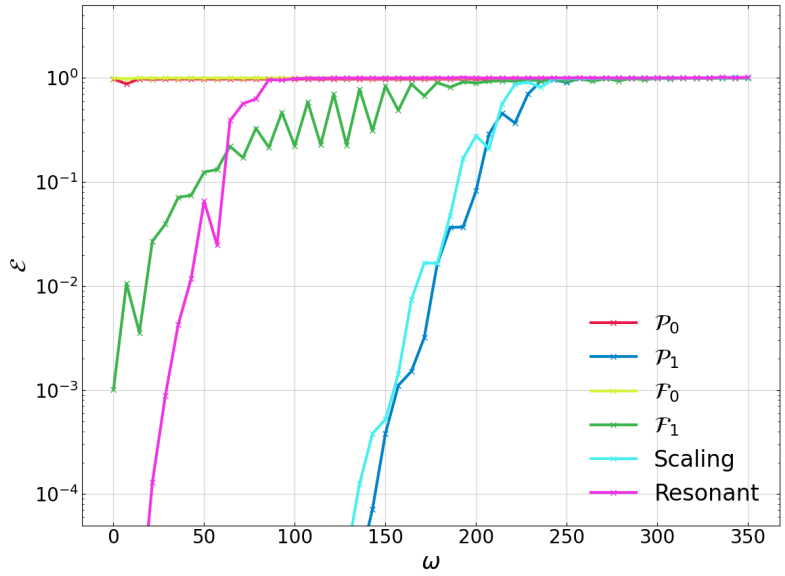


(a) $\cos(\omega(k_1 + k_2 + k_3))$

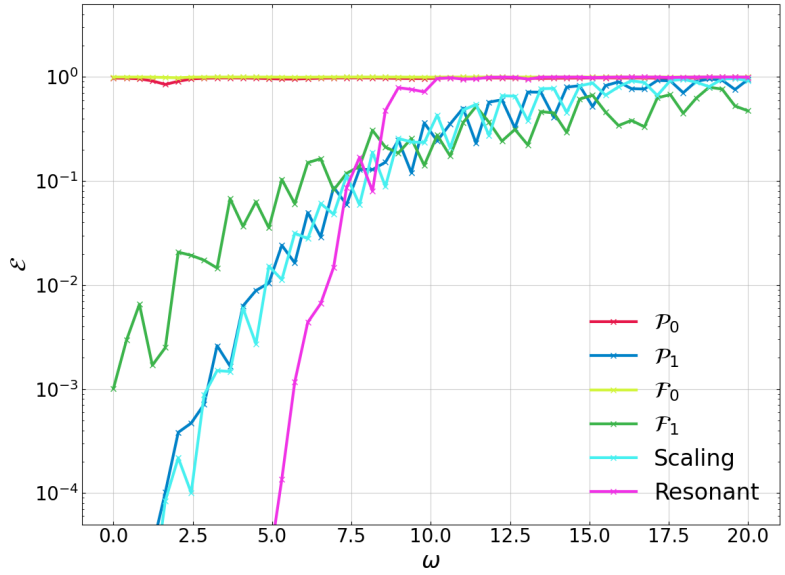


(b) $\cos(\omega \ln(k_1 + k_2 + k_3))$

Figure 3.5: Here we test our basis sets on oscillatory templates, which are a good proxy for feature models. These plots are useful to aid in deciding which basis to run CMB-BEST for, in determining which covers the widest range of features. The $\mathcal{P}_1^{n_s}$ basis is not included in every plot, but would always perform between the $\mathcal{P}_0$ and *scaling* basis. The basis sets are defined in Table 3.1, and some are plotted in figures 3.7 and 3.8. Basis sets $\mathcal{P}_1$ and *scaling* work well for the linear oscillations, converging up to around $\omega \approx 200$. For logarithmic oscillations the resonant basis works best, as expected. However, the improvement is less dramatic.



(a) $\cos(\omega(k_1 + k_2 + k_3))S^{DBI}$



(b) $\cos(\omega \ln(k_1 + k_2 + k_3))S^{DBI}$

Figure 3.6: We hold the basis size $p_{\max}$ fixed at 30 and increase the frequency of the oscillatory template. For linear oscillations with non-trivial shape dependence (a) we see that $\mathcal{P}_1^{n_s}$ and the *scaling* basis converge for a significantly larger frequency range than the *resonant* basis, as expected. We also see that $\mathcal{F}_0$ and $\mathcal{F}_1$ converge very poorly due to the DBI shape, only covering around a tenth of the frequency range that the *scaling* basis converges for, up to $\omega \approx 170$. For logarithmic oscillations the resonant basis converges best, capturing the complex oscillatory templates up to around $\omega \approx 7$. Beyond this point none of the basis sets converge so the relative performance is irrelevant, but we can see that the *resonant* basis actually has the largest error in this region—this is due to that basis being constructed out of orthogonal Legendre polynomials with a log-scaled argument.

3.4 The *scaling* basis

Despite having the same power spectrum scaling, different models can have different bispectrum scalings. Therefore while the power spectrum can give us a rough estimate of the scaling of the bispectrum, the $\mathcal{P}_1^{n_s}$ basis may not converge sufficiently quickly if (3.30) is not a sufficiently close match to the required scaling. We would instead like a basis that could fit a range of fractional powers of k. We achieve this using the *scaling* basis. The *scaling* basis is built using the Legendre polynomials, augmented (in the sense of (3.27)) with

$$k^{-1}, \qquad \ln(k)k^{-1} \tag{3.32}$$

This is motivated by

$$k^{\epsilon} = e^{\epsilon \ln\left(\frac{k}{k_{\rm min}}\right) + \epsilon \ln(k_{\rm min})} \tag{3.33}$$

$$\approx e^{\epsilon \ln(k_{\rm min})}\left(1 + \epsilon \ln\left(\frac{k}{k_{\rm min}}\right)\right) \tag{3.34}$$

which is valid when $\epsilon \ln\left(k_{\rm max}/k_{\rm min}\right)$ is sufficiently small. Since for the DBI case the scaling exponent ϵ is expected to be of the same order as $n_s - 1$ and ε_s, we see that this approximation is good for our case, where (as we will discuss in Section 3.7) we have $\ln\left(\frac{k_{\rm max}}{k_{\rm min}}\right) \approx 6.91$. By adding k^{-1} and $\ln(k)k^{-1}$ separately to our basis, their relative coefficient will be set to the value which provides the best fit to the final shape, providing the flexibility needed to fit shapes with different scalings with the same basis, so we have no need to know anything about the scaling a priori (except that it is small).

For the feature templates, for example Figure 3.5, we see that the *scaling* basis performs equivalently to $\mathcal{P}_1^{n_s}$ and $\mathcal{P}_0$. This is because these template have no scaling or complex shape dependence, and the fit to the oscillatory feature is due to the Legendre polynomials in the basis. In Figure 3.6 we see that the $1/k$ behaviour is necessary to achieve an acceptable fit, but since these templates do not need non-integer scaling, the *scaling* basis performs equivalently to $\mathcal{P}_1$, as expected.

As we can see in Figure 3.4 however, the *scaling* basis converges quickly to a DBI template with a realistic scaling. It outperforms the $\mathcal{P}_1^{n_s}$ basis, despite not assuming power spectrum information.

3.5 The *resonant* basis

We finally describe a basis specifically designed to capture logarithmic oscillations, a type of feature that is usually very difficult to accurately fit. This goes against our philosophy of desiring a basis that is not tied to any specific shape, which was motivated by the fact that CMB-BEST is expensive but need only be run once per basis. However, logarithmic oscillations are an important type of feature in the literature [60]. Running CMB-BEST twice, once for a basis that can cover a broad range of general features, and once for logarithmic oscillations in particular, may be a viable strategy to cover a very broad range of models.

The *resonant* basis is built using the Legendre polynomials, however the argument has been scaled logarithmically. It differs from the basis described in [23] in that it also includes a factor of $\frac{1}{\sqrt{k}}$ to retain orthogonality. We define the nth basis element as

$$\frac{P_n(\bar{k})}{\sqrt{k}}, \qquad \bar{k} = \frac{2\ln(k) - \ln(k_{\rm min}k_{\rm max})}{\ln(k_{\rm max}) - \ln(k_{\rm min})}. \tag{3.35}$$

We then see that

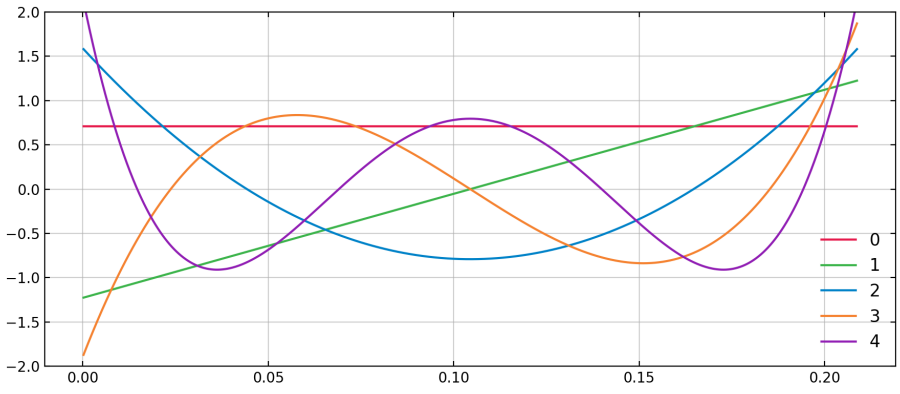
$$\int_{k_{\rm min}}^{k_{\rm max}} dk \frac{P_m(\bar{k})}{\sqrt{k}} \frac{P_n(\bar{k})}{\sqrt{k}} = \int_{\ln k_{\rm min}}^{\ln k_{\rm max}} d\ln k \; P_m(\bar{k}) P_n(\bar{k}) \propto \delta_{mn} \tag{3.36}$$

so the *resonant* basis is orthogonal due to the way we defined $\bar{k}$ for this basis set.

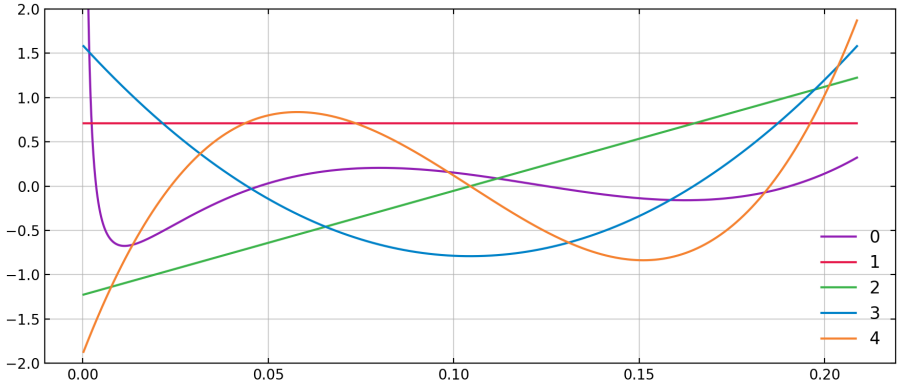
In Figure 3.5 we see that for linear oscillations, as expected, the *resonant* basis converges for a much smaller range for frequencies (for fixed basis set size $p_{\rm max} = 30$) than the other basis sets we test. For logarithmic oscillations however, we see that the range of accessible frequencies is nearly doubled when compared to the other basis sets. We also note that for logarithmic oscillations, wherever we have acceptable convergence the *resonant* basis converges best. We also see however that in the frequency range where none of our basis sets converge, the *resonant* basis has the largest error. This is because the *resonant* basis is designed to be a set of orthogonal logarithmic oscillations, and so will be orthogonal to frequencies outside of its range. On the other hand, while the *scaling* basis (for example) has a worse fit for lower frequencies, it has a better fit to higher frequencies as it is not orthogonal to them. We emphasise however that this only occurs in the region where none of our basis sets provide acceptable convergence for $p_{\rm max} = 30$.

We also note the surprising result in Figure 3.4 that the *resonant* basis

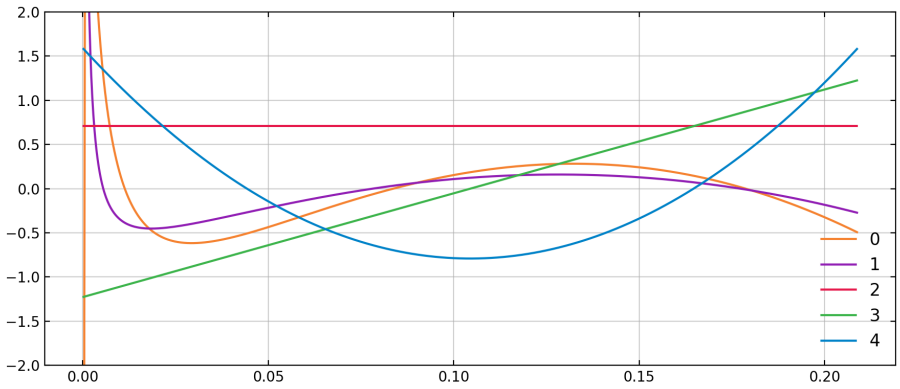
performs best at converging to the scale-dependent DBI template. However, the convergence of the *scaling* basis is still perfectly adequate in this case.



(a) $\mathcal{P}_0$

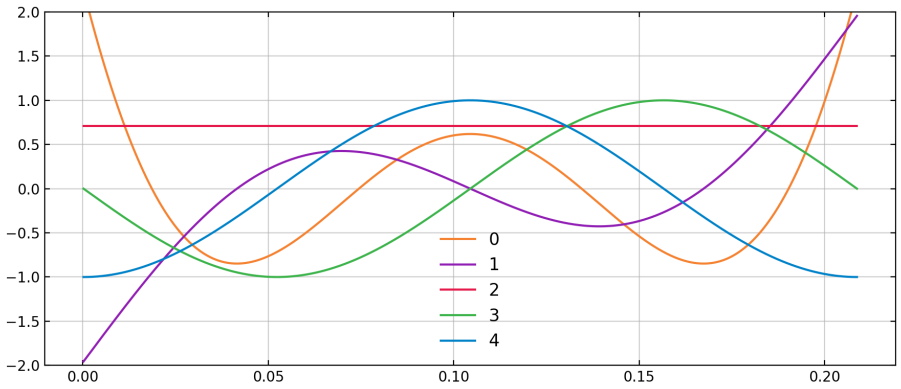


(b) $\mathcal{P}_1$

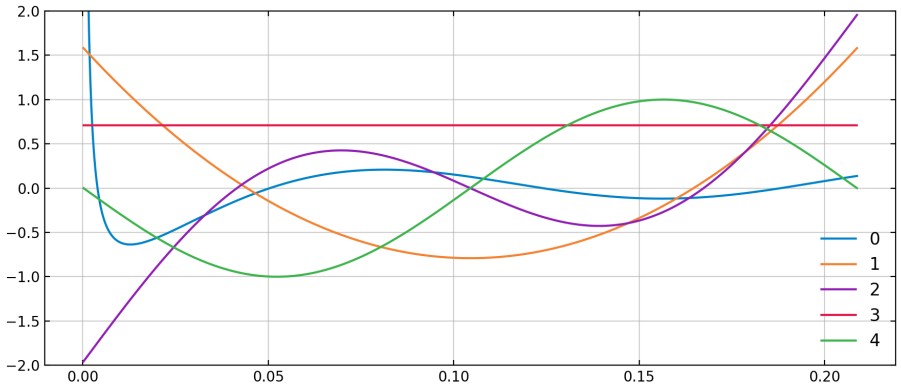


(c) *scaling* basis

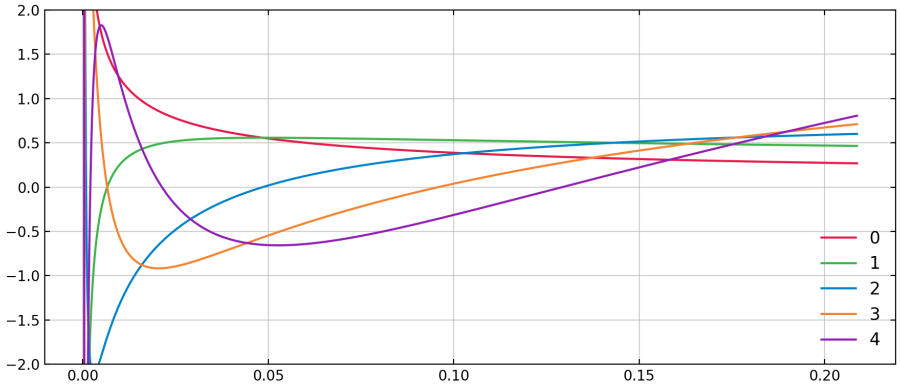
Figure 3.7: We plot the $\mathcal{P}_0$, $\mathcal{P}_1$ and *scaling* basis sets from Table 3.1, for $p_{\max} = 5$. Note that these sets have Legendre polynomial basis elements in common, but differ in which functions they are augmented by, which are added to the start of the indexing.



(a) $\mathcal{F}_0$



(b) $\mathcal{F}_1$



(c) *resonant* basis

Figure 3.8: We plot the $\mathcal{F}_0$, $\mathcal{F}_1$ and *resonant* basis sets from Table 3.1, for $p_{\max} = 5$. Note that $\mathcal{F}_0$ and $\mathcal{F}_1$ have Fourier basis elements in common, but differ in which functions they are augmented by. The *resonant* basis differs in all of its basis elements, lacking even a constant basis element due to the factor of $1/\sqrt{k}$ added to retain orthogonality.

3.6 Large non-physical contributions

When a shape is dominated by its non-physical configurations (those that do not obey the triangle inequality (2.35)) then we can see slow convergence on the tetrapyd, despite possibly fast convergence on the cube as a whole. We will refer to convergence on the tetrapyd as $\mathcal{E}^{tetra}$ and the convergence on the cube as $\mathcal{E}^{cube}$. For a pure oscillation we see no significant difference between the convergence on the cube and on the tetrapyd, as expected for a shape where the physical and non-physical configurations are of the same order of magnitude. For an oscillation with a DBI shape however, we see a large difference, as shown in Figure 3.3. On the cube we see fast, monotonic convergence—the *resonant* basis and *scaling* basis perform well, quickly bring $\mathcal{E}^{cube}$ below 0.1%. On the tetrapyd however, we see that $\mathcal{E}^{tetra}$ only drops below 0.1% for $p_{\max} < 30$ for the *resonant* basis.

We also see that the improvement in $\mathcal{E}^{tetra}$ is no longer monotonic, there are regions where increasing $p_{\max}$ results in a larger $\mathcal{E}^{tetra}$. To understand this it is important to remember that we are not simply adding modes as we increase $p_{\max}$—the augmented mode is changing each time, and thus while we expect $\mathcal{E}^{cube}$ to decrease monotonically, there is no such guarantee for point-wise convergence.

We also notice that $\mathcal{E}^{tetra}$ stays fixed at 1 for low values of $p_{\max}$. This is due to the basis fitting the large non-physical configurations, and thus in (3.19), $S_2 \gg S_1$ everywhere on the cube.

For the local shape the mean value of the shape function on the entire cube (thus accounting for volume effects) is approximately a factor of 40 larger than the mean when restricted to the tetrapyd. For the equilateral template, this factor becomes -300. This illustrates why the tetrapyd-vs-cube problem is so much more severe for equilateral-type shapes than for local shapes, as the non-physical configurations are an order of magnitude more dominant.

This phenomenon is a major obstacle to preserving the separability of the in-in formalism, due to the restriction that we must (effectively, if not in a literal sense) fit to the cube—we have seen that neglecting it and using a simple basis such as $\mathcal{P}_0$ would be disastrous. However, as we have seen, the more sophisticated basis sets described in this chapter are capable of overcoming this obstacle, without surrendering generality.

3.7 A tradeoff between $p_{\max}$ and $k_{\max}/k_{\min}$

While it may not be intuitively apparent, the value of $k_{\max}/k_{\min}$ affects convergence even in the absence of significant features. We can see this in Figure 3.4. In this figure we plot the convergence of various basis sets for the scale-dependent DBI template (2.59), for $k_{\max}/k_{\min} = 550$ and for $k_{\max}/k_{\min} = 1000$. We see that the convergence is slower for higher $k_{\max}/k_{\min}$. This means that to achieve the same convergence for this shape for a higher $k_{\max}/k_{\min}$, we must also go to a higher $p_{\max}$. For example, if one wanted to achieve an accuracy of 0.1% for this shape with the *scaling* basis, one would need $p_{\max} = 15$ for $k_{\max}/k_{\min} = 550$ and $p_{\max} = 20$ for $k_{\max}/k_{\min} = 1000$ We also see this effect for the scale-invariant DBI shape (2.58) but it significantly less dramatic.

We of course desire to use as much of the available data as we can. As such, it is desirable to have the largest possible k-range. However, as we can see, this must be weighted against the loss of descriptive power, as we will then have a smaller set of shapes which will converge sufficiently well to be constrained.

As described in [2], if too much of the squeezed limit is lost, constraining power on models such as those of the local type is lost (at the CMB level). It is shown in that work that $k_{\max}/k_{\min} = 1000$ is sufficient, and as we have seen in this chapter this value still retains convergence on a range of interesting models at the primordial level.

3.8 Conclusions

From the plots presented in this chapter it is clear that the problem of the large non-physical off-tetrapyd configurations is vitally important to the feasibility of the separable in-in calculation of inflationary bispectra. In this setting, where we are effectively forced to fit on the whole cube, we find that the basic Legendre polynomials or Fourier series basis sets alone are not sufficient to capture the shapes we hope to constrain. However, we found that by augmenting the Legendre polynomials with basis functions that could capture $1/k$ behaviour, the convergence increased dramatically and the range of shapes that could be constrained by this pipeline broadens significantly.

While $\mathcal{P}_1^{n_s}$ improves significantly over $\mathcal{P}_0$ for certain examples, as it is tied to a particular scaling it is not an ideal solution. This is due to the desire to create a full pipeline, which connects to real CMB data using CMB-BEst, and the constraint that the expense of that code (for reasonably large $p_{\max}$) limits

the full pipeline to one (or very few) basis sets. Thus we desire a single basis that can capture the broadest range of shapes, including different scalings. To improve upon $\mathcal{P}_1^{n_s}$, we introduced the *scaling* basis, which is augmented by terms which allow it more flexibility in capturing the scaling of the shape function, despite being built without any power spectrum information.

We also presented results from the *resonant* basis, a basis set designed to converge well for logarithmic oscillations, an important target due to the resonance mechanism. While this goes against the philosophy of developing one basis set that converges well across the broadest possible range of bispectrum shapes, it may be useful as an alternative way forward—performing two runs of CMB-BEst, i.e. with both the scaling basis (for linear oscillations and general features) and the resonant basis (for logarithmic oscillations).

Possible future improvements in the design of basis sets could be obtained using tensor decomposition methods, such as the PARAFAC method described in, for example, [129]. Given the output of Primodal, the coefficients α_{pqr}, one could imagine performing a tensor decomposition on α_{pqr}, approximating it as a sum of outer products of vectors. If this decomposition resulted in a compact description, a sum of only a few terms, then it is possible the methods of CMB-BEst as described in [2] could be efficiently applied to individual models, as this would effectively have a much smaller $p_{\max}$. This could allow a very broad, but still template-free, analysis of bispectrum constraints on inflationary models.

Now that we have demonstrated the feasibility of the overall method, in the next chapter we will present its detailed implementation in the context of the in-in formalism.

Chapter 4

Methods and Validation

In Chapter 3 we outlined a formalism for recasting the tree-level in-in calculation for the bispectrum into a form that preserves its separability. We then showed that despite the effects of the non-physical configurations we could still achieve sufficient convergence for a range of interesting models, by investigating how well various basis sets fit standard shape templates.

Given that understanding, in this chapter we will now turn our attention to the numerics of the calculation—specifically, the calculation of the integrals (3.9) and (3.13), and the numerical evolution of ζ_k according to (2.7). We will also validate the implementation of our methods (the PRIMODAL code) on phenomenologically interesting examples.

4.1 Numerics of mode evolution

4.1.1 The initial conditions

For the background quantities, we can evolve τ_s, ϕ and H numerically using (1.37), (1.42) and (1.41) respectively. In the examples we use, we set the initial conditions by prescribing ϕ_{start}, calculating an approximate value of $\phi'_{start,approx}$ using the slow-roll approximation, then using $\phi'_{start,approx}$ to obtain values for c_s^{start} and H_{start}. Taking those values for the sound speed and Hubble parameter as exact, along with the prescribed value for ϕ_{start}, we can calculate the corresponding exact value of ϕ'_{start} using the background Friedmann equation (1.7), obtaining consistent initial conditions to start our numerical evolution. This procedure was described in [71]. See [91] also for a brief discussion on consistent initial conditions.

The Bunch-Davies initial conditions for each mode are set sufficiently early that the modes have converged to the attractor solution before the start of the numerical in-in integration. Since it is necessary to start the numerical integration when all modes are deep in the horizon (as we will see in Figure 4.2) this means the modes must start their evolution when they are are oscillating with a high frequency. As they evolve and cross the horizon they will stop oscillating and eventually freeze-out. This freeze-out marks the end of the in-in integration, as once all the modes have frozen there will be no more contribution to (3.13). This happens as the integrand of (3.13) either becomes purely imaginary or vanishes (although this is dependent on the form of H_{int} used, as we will discuss in Section 4.6). We must use different numerical set-ups in early times versus late times to ensure the efficient calculation of an accurate result.

Similarly to [23], at early times we extract the factor of $e^{-ik\tau_s}$ from the mode functions[1] and numerically evolve $\zeta_k e^{ik\tau_s}$. Unless interrupted by a feature, this quantity decays exponentially at early times. At late times we switch to evolving ζ_k directly. For featureless slow-roll inflation the timing of the switch is simple; so long as it is around horizon crossing, or a couple of e-folds after, the precise location will not affect the result. This becomes trickier when we are dealing with a model with a step feature, for example. Here, we found that navigating the feature in the first set of variables causes difficulty for the stepper. Switching to ζ_k before the onset of the feature gave robust results without needing to loosen the tolerance.

4.2 Integration weights

There are two central types of numerical integral that we have to perform in this method, (3.13) and (3.9). In some regimes these integrals are highly oscillatory, in others they are not. In our method we evolve a fixed number of k-modes through the inflation scenario. At each timestep we must perform two decompositions (3.9), one for both ζ_k and ζ_k', by integrating over these given sample k-points. These decompositions at each timestep involve performing many integrals over the same sample points, one for each of the basis functions.

The integrals over time also require many integrals over the sample points—one for each of the shape coefficients, per vertex, over the decomposition coefficients calculated at each timestep. Thus, it is important that we have a general

[1] In fact [23] extracts a factor of $e^{-ikc_s(\tau)\tau}$, losing efficiency due to slow-roll corrections.

method to efficiently integrate many different functions over the same sample points. In this section we detail how to numerically evaluate these integrals accurately and efficiently.

We do this by calculating integration weights for each of the sample points. In the regimes where the functions are not expected to be oscillatory, we achieve this using the basic method of performing a second order Legendre decomposition on each pair of segments, and integrating the Legendre polynomials to obtain the weights. We break the integral up into segments $[x_0, x_1, x_2, \ldots, x_{N-1}]$. To calculate $\int_{x_{i-1}}^{x_{i+1}} f(x)dx$ we take each triplet of sample points $[x_{i-1}, x_i, x_{i+1}]$, and map it to $[\bar{x}_{i-1}, \bar{x}_i, \bar{x}_{i+1}]$, where

$$\bar{x}_{i-1} = -1, \tag{4.1}$$

$$\bar{x}_i = \frac{2x_i - (x_{i-1} + x_{i+1})}{x_{i+1} - x_{i-1}}, \tag{4.2}$$

$$\bar{x}_{i+1} = 1. \tag{4.3}$$

We then calculate

$$A_{pq} = P_p(\bar{x}_q) \tag{4.4}$$

for the pth Legendre polynomial, $p \in [0, 1, 2]$ and $q \in [i-1, i, i+1]$. Inverting $A_{pq}/(x_{i+1} - x_{i-1})$ gives us the coefficients of the fit, and the definite integral is simply the 0 coefficient, i.e. $(A^{-1})_{0q}$ is the contribution that these segments give to the weight for the sample point x_q. When all the weights w_i have been calculated, the resulting approximation for the integral is simply

$$\int_{x_0}^{x_{N-1}} f(x)dx \approx \sum_{i=0}^{i=N-1} f(x_i)w_i, \tag{4.5}$$

i.e. a simple dot product between the sampled function and the weights, which need only be calculated once before being used for an arbitrary number of functions f. Thus this method very efficient, without needing to sacrifice accuracy, and allows us to keep the sample points x_i flexible instead of mandated.

For oscillatory regimes, we can improve our convergence by including the known oscillatory frequency in our calculation of the weights. For example we calculate $\int_{x_{i-1}}^{x_{i+1}} \cos(wx)f(x)dx$ by performing a Legendre expansion of f, which is relatively slowly varying. By precomputing $\int_{x_{i-1}}^{x_{i+1}} \cos(wx)P_n(\bar{x})dx$ for each segment, we can then obtain a set of weights which provide a good approximation

to the integral, with the cost of integrating an extra function simply the cost of performing an extra dot product—again, very efficient without a sacrifice in accuracy.

To obtain a k-sample we must evolve a Fourier mode from Bunch-Davies initial conditions deep in the horizon until it becomes constant after horizon crossing. We denote by N_k the number of Fourier modes we evolve. Different choices of distributing the k-samples are possible; for example, one could distribute them with an even spacing, log-spacing or cluster them more densely near $k_{\min}$ and $k_{\max}$. We have found the best convergence results from distributing the k-samples according to the prescription of Gauss-Legendre quadrature.

One complication that arises in the calculation of the integration weights for the decompositions is that the oscillation frequency (in k) at each timestep depends on time. At early times τ_s is large, so in k this is highly oscillatory. This creates two problems. Firstly, this seems to require many samples in k to accurately calculate each $f_p^{(i)}(\tau)$, adding more modes that must be evolved in time. To bypass this, we extract the oscillatory part at early times. One may think that this would require us to recompute the integration weights for each timestep, reducing the efficiency of the method (or forcing us to use a less sophisticated integration method but to evolve more ζ_k samples). In fact, due to the simple $e^{ik\tau_s}$ behaviour the integration weights can be calculated in closed form for each τ_s. Even if this were not true however, since this early-time oscillatory behaviour is scenario-independent, one could simply numerically precalculate the weights for a given set of timesteps, thus maintaining accuracy without adding significant overhead to a parameter scan. See [130] for a review.

4.3 Decompositions

As mentioned in the previous section, at early times ζ_k is highly oscillatory as a function of k, taking the form $z_k e^{-ik\tau_s}$ for some much smoother z_k. Directly decomposing this would require evolving more ζ_k samples than is practical. Instead, as we have described, we extract the factor of $e^{-ik\tau_s}$ and build it into our integration weights.

One could expect that this forces us to calculate its basis expansion up to very high order if we want to accurately converge to $F^{(i)}(\tau, k)$ in (3.8). In fact, obtaining a convergent final bispectrum result does not require calculating the full convergent sum for $F^{(i)}(\tau, k)$, as the higher order parts will cancel in the

time integrals for any sufficiently smooth $S(k_1, k_2, k_3)$. That is, assuming the expansion in (3.16) converges within our $p_{\max}$, we need only compute the first $p_{\max}$ of the coefficients[2] in the expansion (3.8) of the $F^{(i)}(\tau, k)$, not until the actual sum (3.8) converges, since for high enough orders the integrals in (3.13) will integrate to zero. This will no longer be true, of course, in the presence of features—in that case the expansion must be calculated to high order.

Clearly, once τ_s becomes small enough these considerations will no longer be necessary and we can simply decompose the mode function directly. We do this around the horizon crossing of the geometric mean of $k_{\min}$ and $k_{\max}$. If there is an extreme feature which causes a large deviation from the usual slow-roll form this switch will need to be made sooner. Also, this method would need to be adapted for non-Bunch-Davies initial conditions, e.g. by using methods detailed in [131–133]. Since anything related to the basis but independent of the scenario can be precomputed, certain parts of this calculation do not hurt the efficiency of this method in the context of a parameter scan.

4.4 Starting the integration with a pinch

While our method has extra suppression at early times compared to the usual in-in calculation, the integral (3.13) still converges slowly in its limit at early times. As long as we are earlier in time than any relevant features or resonances, the integrand will oscillate far faster than its amplitude varies. Such a regime will not significantly contribute to the final result, but if treated without care will cause errors that will swamp the true result, possibly by orders of magnitude. To overcome this inefficiency we will discuss in detail how to start this time integration.

Since calculating each extra point in the time integrand requires a decomposition (3.9), it is worthwhile to consider how to perform the time integral with the minimal number of sample points. From the form of the Bunch-Davies mode functions, we expect the dominant frequency (in τ_s) to be $3k_{\max}$. Assuming we are earlier than any features that might change this, we can use this knowledge to sample the integrand at a far lower rate, building the oscillation into our quadrature weights, as previously described. A second important consideration comes from how early we begin sampling the integrand. We can of course only obtain a point in the time integrand after our mode functions have burned in

[2]This is slightly complicated by the later inclusion of the spatial derivatives in (3.10) and (3.14), which mixes the modes—see later discussion in that section.

from their set initial conditions to their true attractor trajectory. This means that sampling earlier in the time integrand requires us to set the initial conditions for the mode functions deeper in the horizon, a regime in which they are expensive to evolve.

The integrals of the form (3.13) that we must calculate have $\tau = -\infty$ as their lower limit. The highly oscillatory nature of the mode functions (in k) in these early times ($|(k_{\max} - k_{\min})\tau_s| \gg 1$) suppresses the coefficients of our basis expansion by a factor of $1/\tau_s$. As noted in [23], this means that we do not need to explicitly use the $i\varepsilon$ prescription to force the integrals to converge. In the case of using the Legendre polynomials as our basis, we can see this somewhat more precisely by considering the expansion [134]:

$$e^{-i\bar{k}(k_{\max}-k_{\min})\tau/2} = \sum_{n=0}^{\infty}(2n+1)i^n P_n(\bar{k}) j_n(-(k_{\max} - k_{\min})\tau/2) \tag{4.6}$$

for $\bar{k}$ in $[-1, 1]$. When $(k_{\max} - k_{\min})\tau/2$ is large, the spherical Bessel functions oscillate with an amplitude $\propto \frac{1}{\tau}$. Thus, the initial conditions (2.17) expanded in Legendre polynomials (and similar) give us an extra suppression of $1/\tau^3$ in (3.13).

While our method has extra suppression compared to configuration-by-configuration methods (and thus does not need the $i\varepsilon$ prescription to converge) it still converges rather slowly, as we push the lower limit to earlier times. This expensive sampling can be wasteful of resources, especially in a feature scenario where we know this region will not contribute to the final result. Care is required however, as starting the integration in the wrong way can easily lead to errors which can completely swamp the result, since higher order modes are more sensitive to early times. The authors of [64] used an artificial damping term to smoothly "turn on" their integrand. The point at which this is done can then be pushed earlier to check for convergence. However they found that the details of the damping needed to be carefully set to avoid underestimating the result. In [65] they replaced this method by a "boundary regulator"; they split the integral into early and late parts and used integration by parts to efficiently evaluate the early time contribution. As our integrand already has extra suppression compared to the configuration-by-configuration integrands considered in [64, 65], we can safely use the simpler first method.

We understand this situation by taking advantage of the asymptotic behaviour of highly oscillatory integrals (for a review see [135]). The leading order contribution depends on the value of the non-oscillatory part only at the endpoints,

and the next-to-leading order contribution depends on the derivative only at the endpoints. Thus we can approximate the integral $\int_{-\infty}^{T} f(\tau_s)e^{iw\tau_s}d\tau_s$ by replacing the non-oscillatory part $f(\tau_s)$ with a function with matching value and derivative at $\tau_s = T$, but which converges far faster. We use

$$f(\tau_s)e^{-\beta^2(\tau_s-T)^2}, \tag{4.7}$$

for $\tau_s < T$. In this way, for sufficiently large negative T, we obtain the accuracy of the first two terms of the asymptotic expansion ($O(\beta^2/w^2)$, $w = 3k_{\max}$) without needing to explicitly calculate the derivative at T. We also do not need any phase information, as one would need to accurately impose a sharp cut on the integrand.

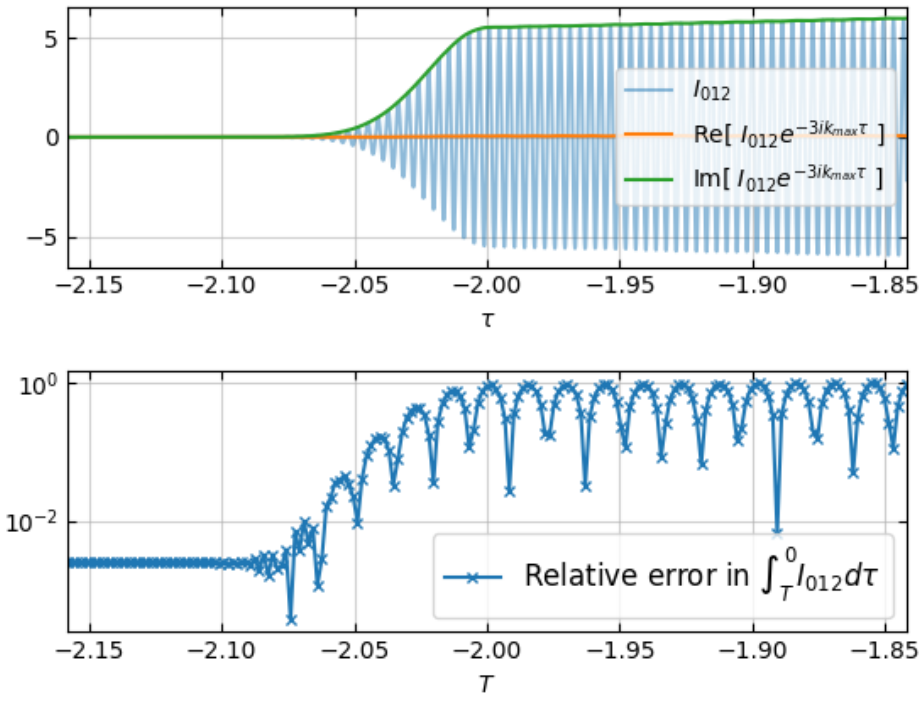
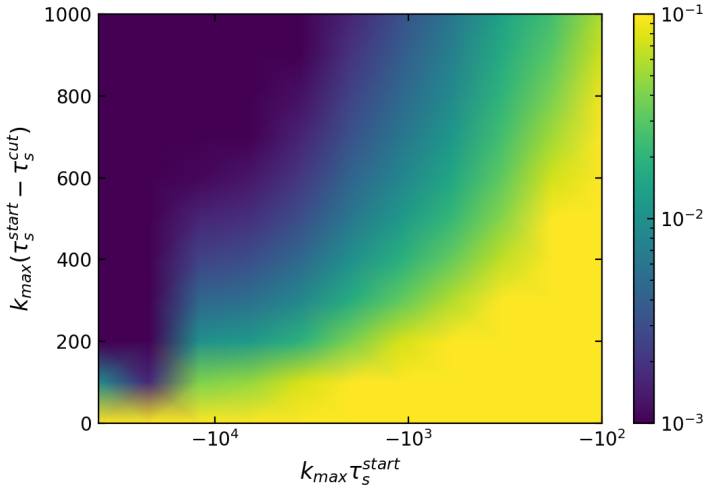
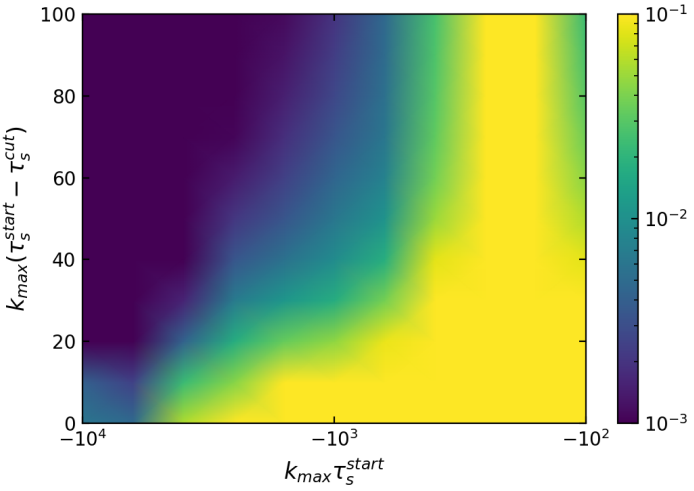


Figure 4.1: A toy example demonstrating the considerations involved in performing the time integrals (3.13). By carefully starting the time integrations, using the form (4.7), we can avoid errors that would otherwise swamp our result. The coefficient being calculated is the α_{012} coefficient of the $\mathcal{P}_0$ expansion of (2.92).

Thus the damping $e^{-\beta^2(\tau_s-T)^2}$ for $\tau_s < T$ smoothly sets the integrand to zero before a certain initial time, T. As long as T is sufficiently early and β is not too large their precise values have no significant effect on the final result. For



(a) $\mathcal{E}$ for DBI model.



(b) $\mathcal{E}$ for resonant model.

Figure 4.2: Here we show the convergence in τ_s^{start} and τ_s^{cut} for a DBI scenario and a resonant scenario (with a canonical kinetic term, on a quadratic potential). Note the different scales on the top and bottom plots. We quantify the convergence by plotting the error compared to the fully converged result, measured by $\mathcal{E}$. In both cases, the error is unacceptably high for a sharp cut, $\tau_s^{start} - \tau_s^{cut} = 0$, across the entire width of the scan. However, for even moderately positive values of $\tau_s^{start} - \tau_s^{cut}$ we see that the error can be reduced by orders of magnitude, for the same computational cost (i.e. the same τ_s^{cut}). In both cases we also see that for values of τ_s^{start} which are too late in time, we do not recover the correct result in the range of the scan. This is expected, as in this case we are losing relevant physical information.

definiteness, we take $\beta/w = 1 \times 10^{-4}$, small enough that the integrand has many oscillations while it is "turning on", so matches the contribution of an infinite limit to high accuracy. We demonstrate this in Figure 4.1 for a toy $H_{int} = (-1/\tau)\dot{\zeta}^3$, which produces the shape (2.92).

We also demonstrate this for a realistic DBI model and resonance model in Figure 4.2. Here we reparametrise the damping in a form that with a more obvious physical interpretation. We use the form $e^{-100\left(\frac{\tau_s^{start}-\tau_s}{\tau_s^{start}-\tau_s^{cut}}\right)^2}$. The point $\tau_s = \tau_s^{start}$ is the earliest point we calculate the integrand for, i.e. before this point we set the integrand to exactly zero—we see that at this point the integrand is suppressed by a factor of e^{-100}. In Figure 4.2 we see that $\tau_s^{start} - \tau_s^{cut} = 0$ (i.e. the naïve method of starting the integrand with a sharp cut) gives unacceptably large errors across the range of the scan. For sufficiently large τ_s^{start} this is ameliorated by a positive $\tau_s^{start} - \tau_s^{cut}$. This validates the accuracy and usefulness of this method.

This effect is more relevant for the DBI scenario, as the size of the errors coming from early times can easily dominate the final result, whereas in the resonant example the large contributions from the resonance at later times are more robust. For the resonant scenario (2.62) we have $c_s = 1$, so we expect (from [41]) the point of resonance τ_s^k for each mode k to be approximately given by $\tau_s^k = \frac{-|\phi'|}{2fk}$. This means that for our example, we expect the earliest resonance time to be $\tau_s \approx -\frac{0.14}{2(0.02)(2.088\times10^{-4})} \approx -1.7 \times 10^4$. Indeed, in Figure 4.2 we see that we must have $k_{\max}\tau_s^{start} < -10^3$ to correctly capture the feature.

4.5 The interaction Hamiltonian

The methods detailed in the previous section depend on the separability of the third-order interaction Hamiltonian, H_{int}, and the possibility of including the spatial derivatives in a numerically accurate and efficient way. To make precise how our methods take into account the details of H_{int}, we will take $P(X, \phi)$ inflation as an example. The full cubic interaction Hamiltonian, not neglecting

boundary terms, can be calculated as [5, 66, 67]

$$H_{int}(t) = \int d^3x \Big\{ -\frac{a^3\varepsilon}{Hc_s^4}\left(1-c_s^2-2c_s^2\frac{\lambda}{\Sigma}\right)\dot{\zeta}^3 + \frac{a^3\varepsilon}{c_s^4}\left(3-3c_s^2-\varepsilon+\eta\right)\zeta\dot{\zeta}^2 - \frac{a\varepsilon}{c_s^2}\left(1-c_s^2+\varepsilon+\eta-2\varepsilon_s\right)\zeta(\partial\zeta)^2 - \frac{a^3\varepsilon^2}{2c_s^4}(\varepsilon-4)\dot{\zeta}\partial\zeta\partial(\partial^{-2}\dot{\zeta}) - \frac{a^3\varepsilon^3}{4c_s^4}\partial^2\zeta(\partial(\partial^{-2}\dot{\zeta}))^2\Big\} \tag{4.8}$$

with $\Sigma = \frac{H^2\varepsilon}{c_s^2}$ and $\lambda = X^2P_{,XX} + \frac{2}{3}X^3P_{,XXX}$. See [5] for further details.

This is commonly quoted with a term proportional to the equation of motion, but this will never contribute [5, 69, 136, 137]. We do not need to make a slow-roll approximation (the quantities defined in (1.38) are not required to be small, except in that we wish to have a successful inflation scenario), nor do we need to neglect any terms in the interaction Hamiltonian. We do no field redefinition, so do not need to add a correction to the final bispectrum. Following the calculation of [5] (see also [69, 136, 137]) we do not work with any boundary terms.

4.6 Stopping the integration

Numerically, it is preferable to write H_{int} in a form with no boundary terms, whether they come from undoing a field redefinition or from integration by parts. Since the boundary term contribution will depend on the choice of when to end the integration, its time dependence must cancel with a late-time time-dependent contribution of some vertex, requiring us to track the necessary quantities much longer than otherwise needed to obtain the desired precision.

Schematically, the correction from a field redefinition would look like

$$\langle\zeta_{\mathbf{k_1}}\zeta_{\mathbf{k_2}}\zeta_{\mathbf{k_3}}\rangle = \left\langle\tilde{\zeta}_{\mathbf{k_1}}\tilde{\zeta}_{\mathbf{k_2}}\tilde{\zeta}_{\mathbf{k_3}}\right\rangle + \lambda\left\langle\tilde{\zeta}_{\mathbf{k_1}}\tilde{\zeta}_{\mathbf{k_2}}\right\rangle\left\langle\tilde{\zeta}_{\mathbf{k_1}}\tilde{\zeta}_{\mathbf{k_3}}\right\rangle + cyclic \tag{4.9}$$

where λ is some function of the slow-roll parameters. The correction terms will have a time dependence from λ, so the $\left\langle\tilde{\zeta}_{\mathbf{k_1}}\tilde{\zeta}_{\mathbf{k_2}}\tilde{\zeta}_{\mathbf{k_3}}\right\rangle$ term must have some late time contribution to cancel it. To obtain an accurate result, care would need to be taken with this cancellation, an unnecessary (and possibly quite difficult) complication.

Even among the forms of H_{int} that have no boundary terms, we have some extra freedom that can be used to increase the numerical accuracy of our calcu-

lation. By integrating by parts and using the equation of motion, the interaction Hamiltonian can be rewritten without picking up boundary terms [106]. Using (3.7) from [106], with $f = -\varepsilon/(c_s^2 H)$, we obtain the following form:

$$\begin{aligned} H_{int}(t) = \int d^3x \Bigg\{ & -\frac{a^3\varepsilon}{Hc_s^4}\left(-c_s^2 - 2c_s^2\frac{\lambda}{\Sigma}\right)\dot{\zeta}^3 + \frac{a^3\varepsilon}{c_s^4}\left(-3c_s^2\right)\zeta\dot{\zeta}^2 \\ & - \frac{a\varepsilon}{c_s^2}\left(-c_s^2\right)\zeta(\partial\zeta)^2 - \frac{a\varepsilon}{Hc_s^2}\dot{\zeta}(\partial\zeta)^2 \\ & - \frac{a^3\varepsilon^2}{2c_s^4}(\varepsilon-4)\dot{\zeta}\partial\zeta\partial(\partial^{-2}\dot{\zeta}) - \frac{a^3\varepsilon^3}{4c_s^4}\partial^2\zeta(\partial(\partial^{-2}\dot{\zeta}))^2\Bigg\}. \end{aligned} \tag{4.10}$$

To leading order, this formulation is made up of terms that give equilateral shapes when the slow-roll parameters are roughly constant. It was pointed out in [23] that using (4.8) in a scenario that results in an equilateral shape would require sensitive cancellations in the squeezed limit. Likewise, using (4.10) for a local scenario would require sensitive cancellations in the equilateral limit. As such, we can choose whether to use (4.8) or (4.10) depending on our expected result—both are fully implemented in the PRIMODAL code.

Finally, we mention some points relating to including the spatial derivatives in the calculation. As mentioned in [23], the spatial derivatives can be manipulated into simple prefactors of k_i using the triangle condition ($\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3} = 0$), and so preserve the separability of the result. To absorb these prefactors in our calculation, we precompute $k^p q_a(k)$ as a linear combination of the $q_a(k)$ for the relevant values of p, from which $V^{(i)}_{P\tilde{P}}$ defined in (3.14) is built. For certain sets of basis functions this matrix can be calculated analytically, but it is simpler and more robust to numerically calculate the relevant integral directly. The processing cost this incurs is small, and must only be paid once per basis. We note especially that this means the matrix can be stored and efficiently used in many scenarios. To summarise, we calculate the bispectrum contribution from each vertex in H_{int} separately: we assemble the integrands, integrate them with respect to time, include the prefactors coming from the spatial derivatives, then sum the resulting sets of basis coefficients. Of course, these methods are not restricted to this example of the interaction Hamiltonian of $P(X,\phi)$ models.

4.7 Validation

4.7.1 Validation methods

In this section we validate our implementation of our methods on different types of non-Gaussianity, sourced in different ways. While our actual results take the form of a set of mode expansion coefficients α_n, to make contact with previous results in the literature all of our validation tests take place on the tetrapyd, the set of physical bispectrum configurations.

We test that our results have converged using (3.19), between $p_{\max} = 45$ and $p_{\max} = 15$ for the featureless cases, and between $p_{\max} = 65$ and $p_{\max} = 35$ for the cases with features. We will refer to this as our convergence test. To verify that our results have converged to the correct shape, we perform full tetrapyd checks against known analytic results (where those are available, and in their regimes of validity) using (3.19), and point tests against the PyTransport code for the scenarios with canonical kinetic terms. Since all our scenarios are single-field, the most general test we have is the single-field consistency relation, which states that for small k_L/k_S, the shape function $S(k_S, k_S, k_L)$ must obey (2.109). The consistency condition should hold most precisely at the configurations with smallest k_L/k_S, the most squeezed being the three corners, $(k_{\max}, k_{\max}, k_{\min})$ and permutations. We want our test to be on an extended region of the tetrapyd however, so we choose the line

$$\frac{k_L}{k_S} = \frac{2k_{\min}}{k_{\max}}, \tag{4.11}$$

which connects $(k_{\max}, k_{\max}, 2k_{\min})$ to $(k_{\max}/2, k_{\max}/2, k_{\min})$. While we took $\frac{k_{\min}}{k_{\max}} = \frac{1}{1000}$ in Chapter 3, we will take $\frac{k_{\min}}{k_{\max}} = \frac{1}{550}$ in this chapter—this still allows $\frac{2k_{\min}}{k_{\max}}$ to be sufficiently squeezed to be a stringent test.

First, we investigate convergence on simple featureless models, both local-type (2.37) and equilateral-type (2.39). We find that in the basis $\mathcal{P}_{01}^{n_s}$ our results converge quickly and robustly as we increase the number of modes, where we quantify the convergence using (3.19). We compare the converged results against analytic templates (2.57) and (2.58), using the full shape information (3.19), finding them to match to high accuracy. Secondly we validate our methods on an example of non-Gaussianity from a feature: linear oscillations from a sharp step in the potential (2.61). The result converges robustly across the parameter range we explore. Throughout that range, we test the converged result using the squeezed

limit consistency condition, and perform point tests against PyTransport, finding excellent agreement. For small step size we can further validate against the analytic template of [38], using the full shape information, finding agreement to the expected level given the finite width of the step. The final type of non-Gaussianity we use for validation on is the resonance type, logarithmic oscillations generated deep in the horizon (2.62). We test the converged result against the PyTransport code, by performing point tests on a slice. We also present a resonant DBI scenario, with out-of-phase oscillations in the flattened limit, as pointed out in [73], resulting from non-Bunch-Davies behaviour of the mode functions. We also test both resonant scenarios using the squeezed limit consistency condition.

We display the phenomenology of our various validation examples by plotting slices through the tetrapyd, as detailed in Figure 4.3. Along with the phenomenology plots we plot the residual (with respect to the totally converged result) on the same slice, relative to the magnitude of the shape (3.21). We emphasise that while these plots display slices through the tetrapyd, our actual result describes the shape function on the entire three-dimensional volume of the tetrapyd, and we measure our convergence over this whole space.

While one of the main advantages of this method is its direct link to the CMB, in this section we only concern ourselves with validating the code, not the observational viability of the scenarios considered. We focus on accurately and efficiently calculating the primordial tree-level comoving bispectrum, validating on models popular in the literature.

4.7.2 Quadratic slow-roll

The first model we will consider is slow-roll inflation on a quadratic potential (2.37). We consider two scenarios, both with $m = 6 \times 10^{-6}$. The first is deep in slow-roll, which we achieve by choosing $\phi_0 = 1000$; then, choosing ϕ_0' according to the slow-roll approximation, we get $\frac{1}{2}\phi'^2 = \varepsilon \approx 0.2 \times 10^{-5}$. We can then choose the initial value for H to satisfy the Friedmann equation to sufficient precision. The second scenario is chosen to have a value for $n_s^* - 1$ consistent with the *Planck* result, by choosing $\phi_0 = 16.5$, so that $\varepsilon \approx 0.8 \times 10^{-2}$. The shapes are shown in Figure 4.4.

We choose the first scenario to have such a small value of ε so that we can use Maldacena's shape (2.57) as a precision test. Indeed, we find that it has a scaled relative difference (3.20) of 2.7×10^{-5} with this shape, contrasting a

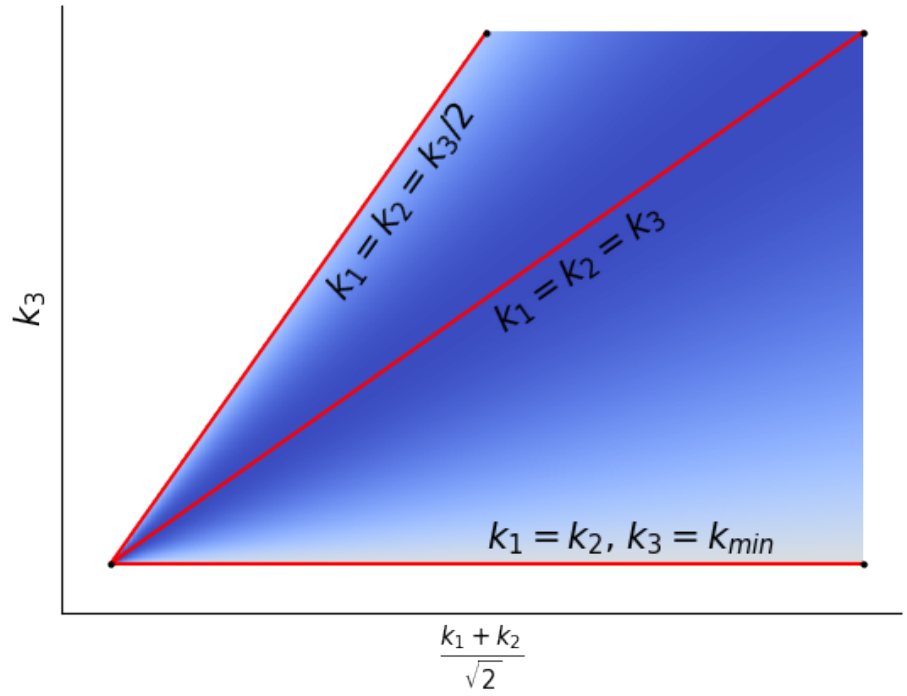


Figure 4.3: For ease of display, we will plot the two-dimensional $k_1 = k_2$ slice of the tetrapyd for each of our validation examples, as shown schematically here. Horizontal lines on this plot have constant k_3. The bottom edge is $k_3 = k_{\rm min}$, the top edge is $k_3 = k_{\rm max}$. The right edge is $k_1 = k_2 = k_{\rm max}$, the left edge is $k_1 = k_2 = k_3/2$, i.e. the limit imposed by the triangle condition. Plotted in red from top-left to bottom-right, are the flattened, equilateral and squeezed limits. This can be compared to Figure 2.1.

scaled relative difference of 0.077 with the local template (2.53). This confirms that our methods and our implementation in code can accurately pick up this basic type of featureless non-Gaussianity.

For the second scenario, we cannot validate on Maldacena's shape (2.57) or the local template (2.53), as for $\epsilon \approx 0.8 \times 10^{-2}$ we only expect these templates to match the true result to percent level accuracy. Indeed, we find that our result has a correlation of 0.998 with both (2.57) and (2.53), corresponding (in the sense of (3.20)) to a relative difference of 6%, as expected. Instead, we validate this model using the squeezed limit test described above, verifying our result to 0.05%.

This is a validation of the convergence of our basis, reaffirming the template decomposition results of Figure 3.1 in the setting of the in-in formalism. It is also

a stringent validation of our methods of including the higher-order coefficients, as insufficient care taken in the early-time sections of integrals (3.13), or in including the spatial derivatives from H_{int}, could have easily swamped the $p_{\max} = 45$ result.

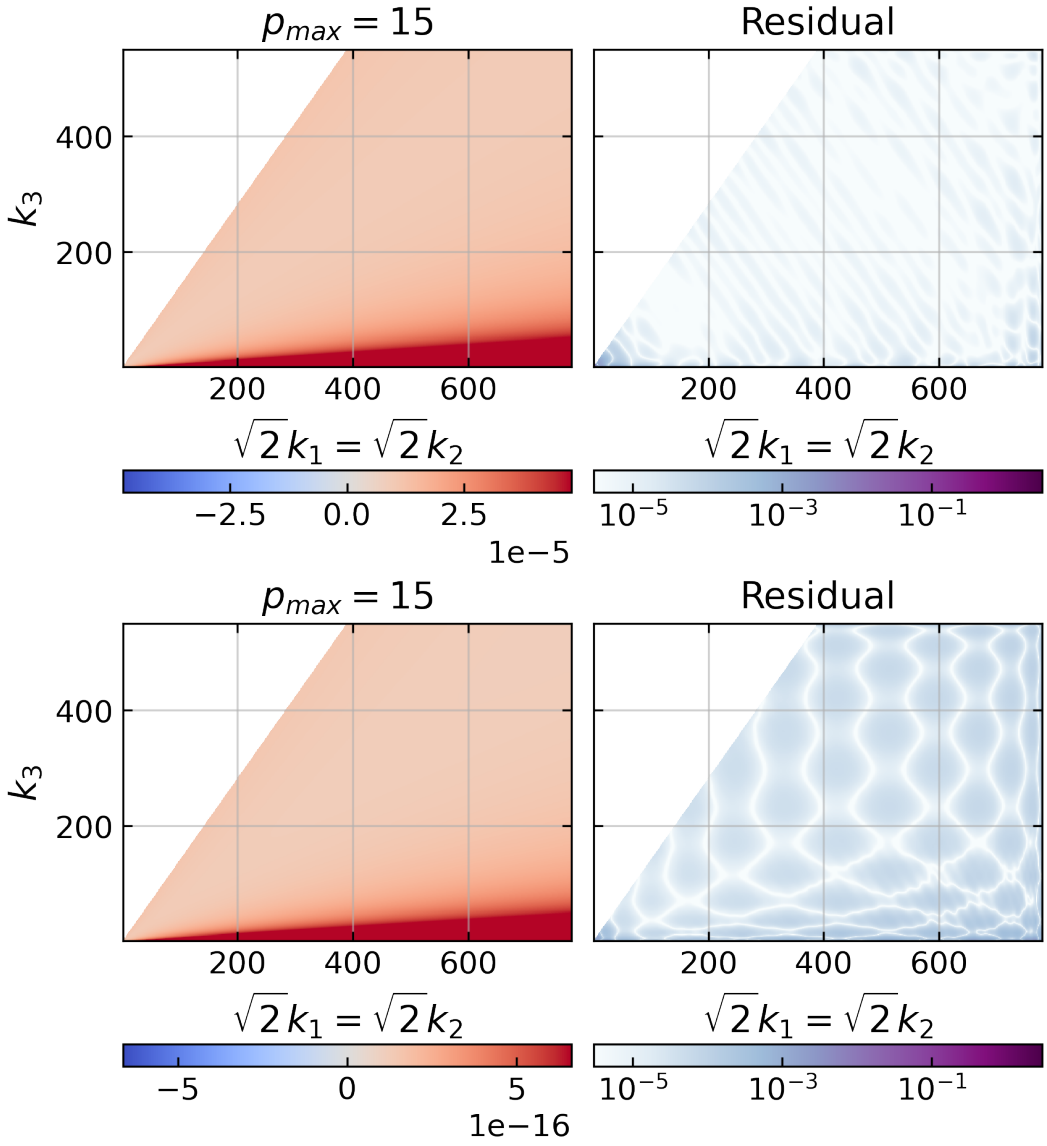


Figure 4.4: A canonical single-field model on a quadratic potential (2.61), slowly-rolling with $\varepsilon \approx 2 \times 10^{-6}$ in the top plot, and $\varepsilon \approx 0.8 \times 10^{-2}$ in the lower plot. This shape is dominated by its squeezed limit, and has a scale dependence determined by ε, very small in the top plot and "realistic" in the lower plot, relative to the *Planck* power spectrum. The first scenario converges well in the $\mathcal{P}_1$ basis, with a relative difference of 2.7×10^{-5} between $p_{\max} = 45$ and $p_{\max} = 15$. The second scenario converges well in the $\mathcal{P}_{01}^{n_s}$ basis (with $n_s^* - 1 = -0.0325$), with a relative difference of 7.9×10^{-5} between $p_{\max} = 45$ and $p_{\max} = 15$.

4.7.3 DBI inflation

Next, we show results for a similar pair of scenarios for DBI inflation. We choose $V_0 = 5.2 \times 10^{-12}$ with $m = \sqrt{0.29V_0/3}$ in (2.38) and (2.39). We choose $\phi_0 = 0.41$, and then the starting condition for H according to the slow-roll approximation, allowing us to choose ϕ_0' such that the Friedmann equation is satisfied to sufficient precision. The first scenario is deep in slow-roll, with $\lambda_{DBI} = 1.9 \times 10^{18}$, while the second scenario saturates the *Planck* limit on c_s, with $\lambda_{DBI} = 1.9 \times 10^{15}$. The resulting shapes are shown in Figure 4.5.

The scenario deep in slow-roll has a error of 0.082% relative to the DBI shape (2.58), and 13% relative to the equilateral template (2.56). The second scenario has a relative error of 2.9% with the scale-invariant DBI shape, and 14% with the equilateral template. Including some scale dependence in the template, using (2.59), we get a relative error of 0.27%. On the line defined by (4.11), both scenarios have a sub-percent difference from the consistency condition, with respect to the equilateral configurations, which decreases when configurations with a larger k_S/k_L are considered.

Including the minimal information of an individual, approximately representative value of $n_s^* - 1$ in $\mathcal{P}_{01}^{n_s}$ allows us to converge to these smooth shapes quickly and robustly, overcoming the tetrapyd-vs-cube difficulties described in 3. Our accurate match to these shapes validates our implementation in code, and the ability of the method (and our basis in particular) to capture very different types of bispectrum shapes, local and equilateral.

4.7.4 Step features

Moving on from simple featureless bispectra, we present the results of our validation tests on non-Gaussianity coming from a sharp feature in the potential. We use the same parameters for the quadratic potential as in the second scenario in fig 4.4. In (2.61) we fix $d = 1 \times 10^{-2}$ and $\phi_f = 15.55$ (as with the second canonical quadratic example, $\phi_0 = 16.5$). Figure 4.6 shows results for the shape function for two step sizes, $c = 5 \times 10^{-5}$ and $c = 5 \times 10^{-3}$. The resulting shape for small step sizes contains simple oscillations, linear in $k_1 + k_2 + k_3$, whose phase is almost constant across the tetrapyd. When the step size is small, as expected, our result matches the analytic result of [38], presented there in equations (48), (54), (55). We plot a comparison of the result of [38] and our result in Figure 4.8. For larger step size, we check the squeezed limit in Figure 4.7, where we also show point tests against the PyTransport code. Across this range of step sizes, for

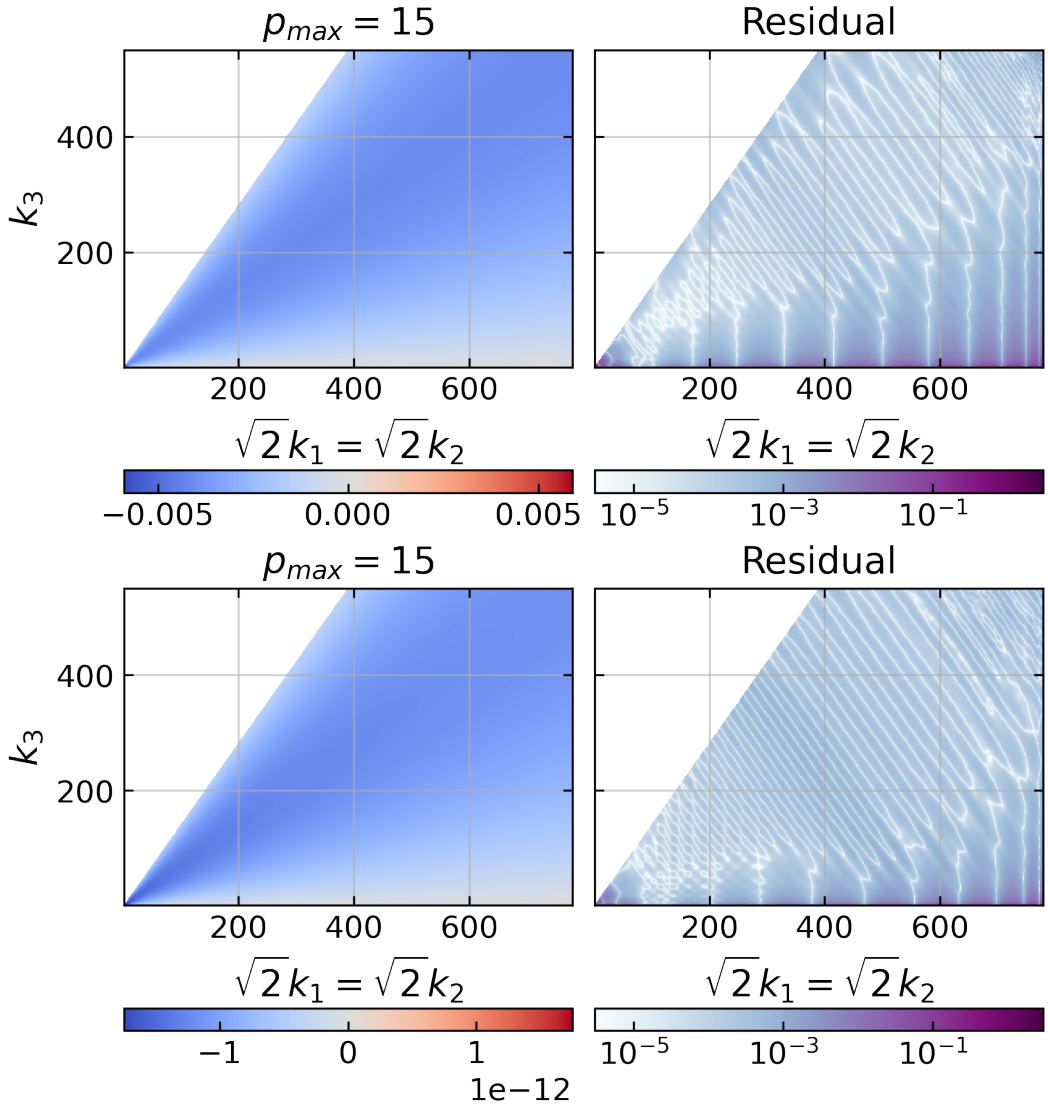


Figure 4.5: The upper plot shows the shape function for a DBI model deep in slow-roll. We set λ_{DBI} in (2.39) to 1.9×10^{18}, obtaining a scenario with $\varepsilon \approx 1.9 \times 10^{-6}$ and $c_s = 2.3 \times 10^{-3}$. This shape is dominated by its equilateral configurations, and has only a slight scale dependence. It converges well in the $\mathcal{P}_1$ basis, with a relative difference of 2.1×10^{-3} between $p_{\rm max} = 45$ and $p_{\rm max} = 15$. The lower plot shows a DBI model that saturates the *Planck* limit on c_s. We set λ_{DBI} in (2.39) to 1.9×10^{15}, obtaining a scenario with $\varepsilon \approx 8.0 \times 10^{-5}$ and $c_s = 8.0 \times 10^{-2}$. This shape is also dominated by its equilateral configurations, but has a scale dependence consistent with the measured power spectrum. It converges well in the $\mathcal{P}_{01}^{n_s}$ basis (with $n_s^* - 1 = -0.0325$), with a relative difference of 1.1×10^{-3} between $p_{\rm max} = 45$ and $p_{\rm max} = 15$.

the resulting shapes we obtain a full tetrapyd convergence test result (between $p_{\max} = 65$ and $p_{\max} = 35$) of between 0.17% and 0.15% and we verify the squeezed limit test to better than 0.5%.

These examples show the utility of our methods in calculating bispectra with non-trivial shape and scale dependence, going beyond the simple examples of [23]. They validate the calculation of the high order coefficients, and show that our code as implemented can handle sharp deviations from slow-roll, generating non-Gaussianity around horizon crossing.

4.7.5 Resonance features

Now we further validate our code against two resonance models. In contrast to the previous sharp kink, this feature is extended, requiring precision at earlier times. The first, shown in Figure 4.9, is a model with a canonical kinetic term, on a quadratic potential with a superimposed oscillation (2.62). We take $bf = 10^{-7}$, and $f = 10^{-2}$. The resulting bispectrum has oscillations logarithmic in $k_1 + k_2 + k_3$. In Figure 4.9 we see the excellent agreement between our result and the PyTransport result, once initial conditions in both codes are set early enough to achieve convergence. This validates the code on non-Gaussianity generated deeper in the horizon. Note the change of phase in the squeezed limit, though this is expected to be unobservable. We obtain a full tetrapyd convergence test result (between $p_{\max} = 65$ and $p_{\max} = 35$) of 0.93%, a squeezed limit test result of 1.1% (along the line defined by (4.11)), and a relative difference of 3.0% with respect to the PyTransport result, although this is only integrated over the two-dimensional slice presented in Figure 4.9.

The time taken for the PyTransport code (per configuration) varies by a factor of around forty between the equilateral limit and the squeezed limit, as we show in Figure 4.9. While the PyTransport code is extremely fast at calculating the shape function for a single k-configuration, to obtain this two-dimensional slice through the tetrapyd took around seven hours; to obtain the shape function on the full three-dimensional tetrapyd would take much longer. In contrast, our code took less than an hour on the same machine to calculate the full shape function, not limited to the shown slice. The overall speed increase is, therefore, a factor on the order of 10^2 to 10^3 for the full shape information, speaking only on the level of primordial phenomenology, in addition to the advantage that our result is in a form designed to be compared with observation. We expect that our implementation can be optimised beyond this.

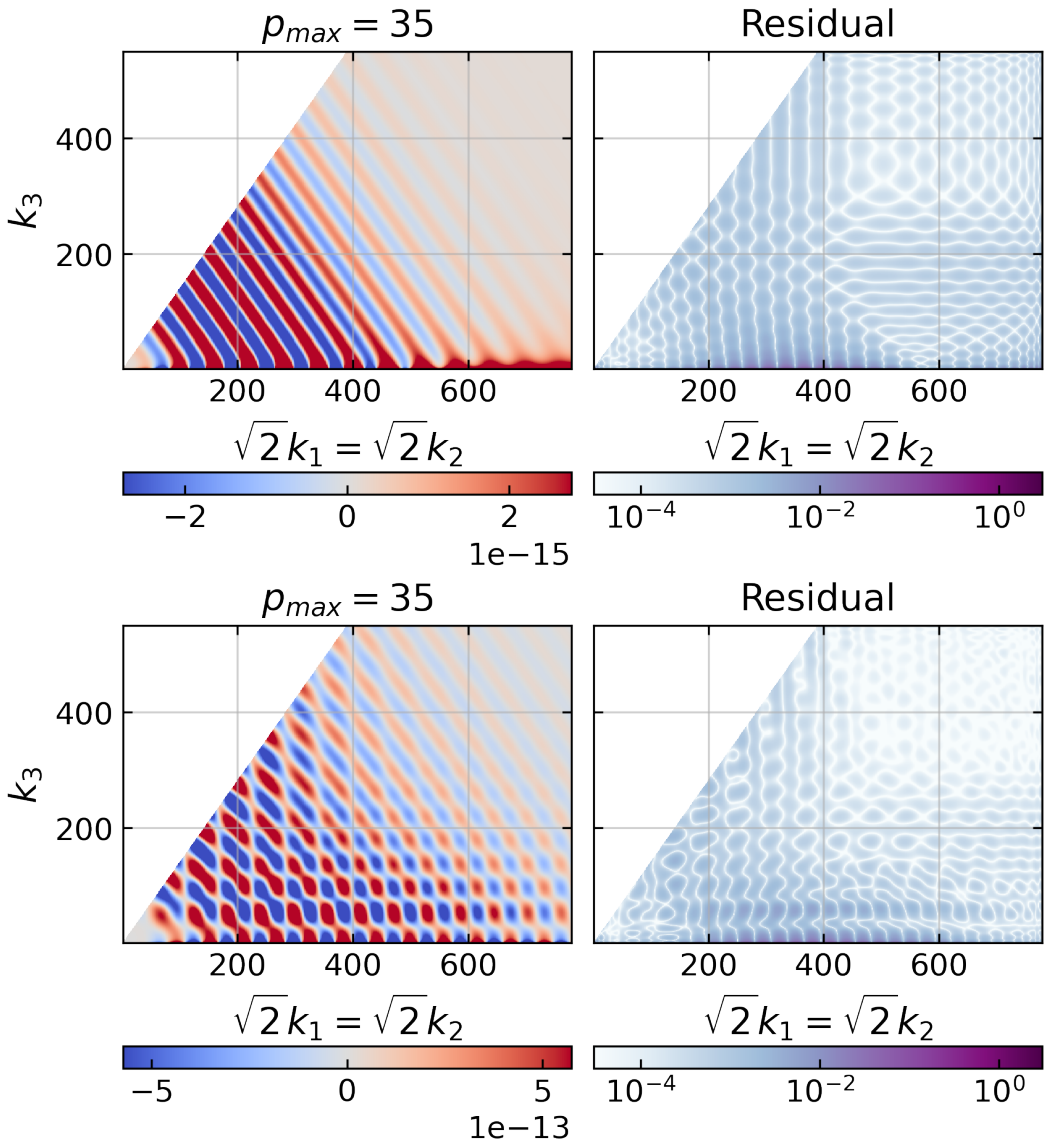


Figure 4.6: The tree-level shape function of a feature model (2.61), shown for step sizes of $c = 5 \times 10^{-5}$ (upper plot) and $c = 5 \times 10^{-3}$ (lower plot). The corresponding expansion parameter values of [38], $\mathcal{C} = 6c/(\varepsilon + 3c)$, are 0.035 and 1.3. For the smaller step size, the oscillations are almost entirely functions of $K = k_1 + k_2 + k_3$, except for a phase difference in the squeezed limit. The dependence is more complicated for $\mathcal{C} = 1.3$, however our result still converges well. In the $\mathcal{P}_{01}^{n_s}$ basis, with $n_s^* - 1 = -0.0325$, the results have a relative difference of 1.6×10^{-3} and 1.5×10^{-3}, respectively, between $p_{\max} = 65$ and $p_{\max} = 35$.

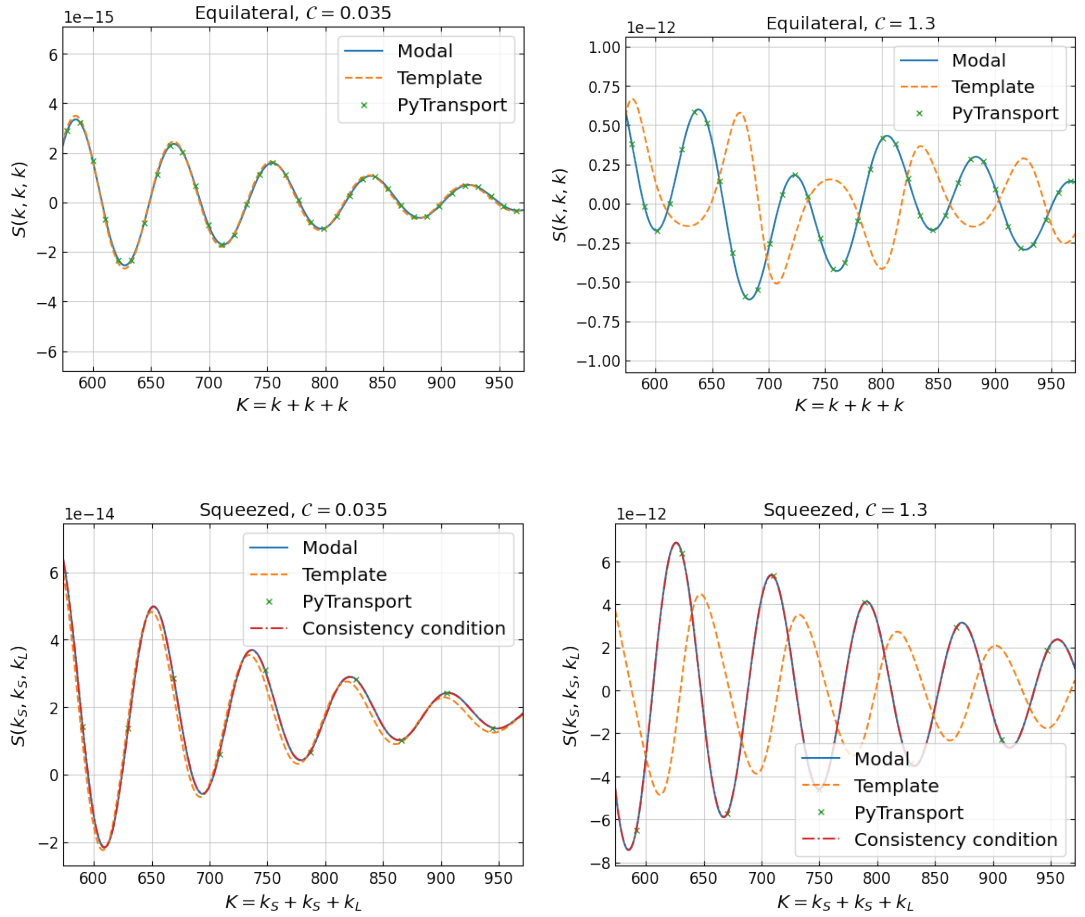


Figure 4.7: In the equilateral limit for the feature models (the top two figures) we validate our modal result against the PyTransport result. In the squeezed limit (the bottom two figures) we validate against PyTransport, and the consistency condition. In both limits, for both step sizes shown, we find excellent agreement. For the small step size (the two plots to the left), we additionally see a good match to the template of [38]. For the larger step size, the template amplitude is still accurate, but no longer captures the detailed shape information. This validates our code on non-Gaussianity generated by sharp features, and illustrates the general usefulness of our method. Our numerical results are accurate in a broader range than approximate templates, but are still smooth separable functions, unlike the results of previous numerical codes.

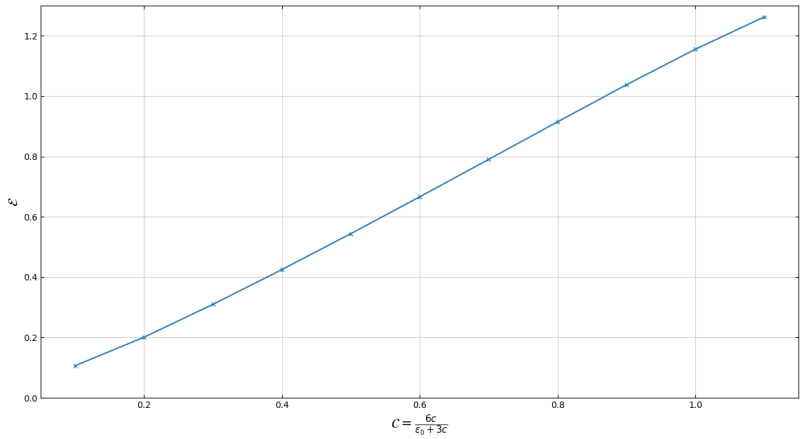


Figure 4.8: We sample more shapes with step sizes between the two feature models shown in Figure 4.6. We plot the relative difference, integrated over the full tetrapyd in the sense of (3.19), between the modal result and the analytic template of [38], as a function of the template parameter $\mathcal{C} = \frac{6c}{\varepsilon_0+3c}$ (where c is the step size and ε_0 is the value of the slow-roll parameter ε at ϕ_{step} when $c = 0$). We test our result by verifying the squeezed limit consistency condition to better than 1% throughout (not shown). The number of oscillations in the k-range is determined by the conformal time at which the kink in (2.61) occurs, which is kept constant across this scan. The width of the feature was also kept constant.

The second scenario we consider here also has an oscillation superimposed on its potential, but this time is a non-canonical model, the DBI model. The resulting bispectrum is shown in Figure 4.10. Note especially the out-of-phase oscillations in the flattened limit, which are potentially observable. For the purpose of displaying this phenomenology, we place a window on the oscillation in the potential, smoothing out the resulting oscillations in the shape at low $k_1 + k_2 + k_3$, to aid convergence. This validates our code on non-Gaussianity generated by deviations from Bunch-Davies behaviour [68, 73]. We obtain a convergence test result (between $p_{\max} = 65$ and $p_{\max} = 35$) of 0.15%, and a squeezed limit test result of 6.5%.

4.8 Conclusions

In this chapter we outlined methods that allow (3.13) and (3.9) to be computed precisely and efficiently for a wide range of basis sets. This allows us to (i) preserve the intrinsic separability of the tree-level in-in formalism and (ii) do so in a way that allows easy exploration of possible sets of basis functions, to find a set that converges quickly enough to be useful in comparison with observation.

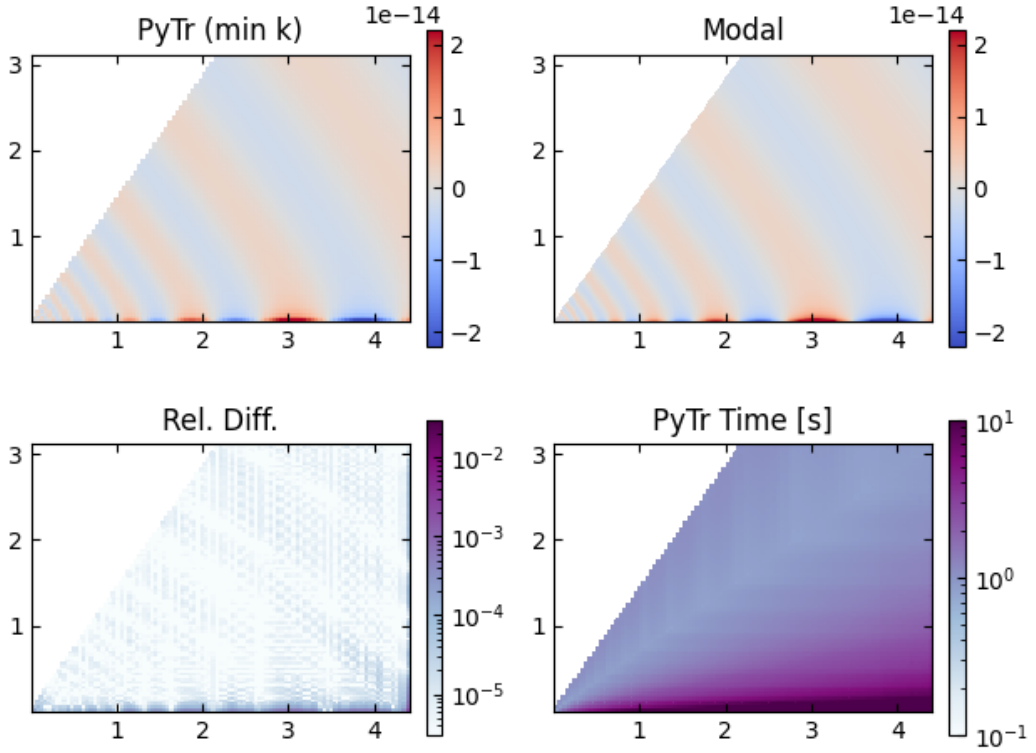


Figure 4.9: Resonance on a quadratic potential (2.62), testing our result using point tests against the PyTransport code. The logarithmic oscillations in the shape function are generated by periodic features deep in the horizon. The differences between our result and the PyTransport result are sufficiently small throughout that we can consider this a validation of our code on non-Gaussianity generated by periodic features deep in the horizon. In the $\mathcal{P}_{01}^{n_s}$ basis, with $n_s^* - 1 = -0.0325$, our result has a relative difference of 9.6×10^{-3} between $p_{\rm max} = 65$ and $p_{\rm max} = 35$.

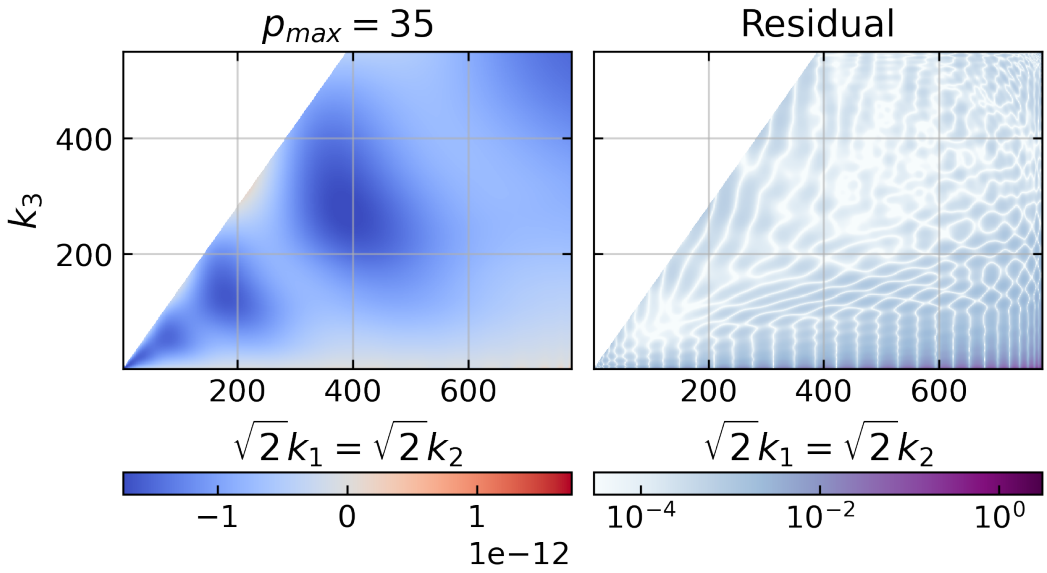


Figure 4.10: Non-Gaussianity generated by periodic features in a DBI model, including a phase difference in the flattened limit as described in [73]. For the purposes of demonstrating the phenomenology, we have placed an envelope on the oscillations in the potential to aid convergence. In the $\mathcal{P}_{01}^{n_s}$ basis, with $n_s^* - 1 = -0.0325$, the result has a relative difference of 1.9×10^{-3} between $p_{\max} = 65$ and $p_{\max} = 35$.

These methods were then validated on interesting and realistic example scenarios. We have confidence in this due to the varied tests we have performed—against templates, other numerical codes, and also against the squeezed limit consistency condition.

For the standard shapes, we have obtained the full shape information far faster than previous numerical methods, for the single-field models that we used as validation examples. We then demonstrated separable in-in methods for high orders and features for the first time, going beyond the work done in [23]. With our methods validated, in the next chapter we will present work performed in collaboration [2], validating the combined PRIMODAL and CMB-BEST pipeline by obtaining a constraint on the sound speed of a DBI inflation scenario from *Planck* data.

Chapter 5

An Inflationary Constraint from the CMB Bispectrum

5.1 DBI sound speed constraint

In this chapter, we use *Planck* data to place a constraint on DBI inflation. We then compare this result to a constraint on the same parameter obtained by the *Planck* collaboration [60], with the aim of validating the full PRIMODAL-CMB-BEST pipeline. The *Planck* result was obtained by constraining the amplitude of the DBI template (2.58), and using a slow-roll relation (2.50) to map this to a constraint on the sound speed. In our pipeline, in contrast, we scan over a fundamental parameter of the model, β_{IR}, with the other parameters of the model held fixed. This means the sound speed is not a constant, and is instead found by numerically evolving the background. We use c_s^* to refer to the sound speed at horizon crossing of the pivot scale. We obtain the full shape and amplitude information (within our k-range) for each scenario in our scan and the final constraint makes use of the entirety of that information, up to convergence in $p_{\max}$.

In the context of the scan (which we will describe in Section 5.2), we used the methods described in previous chapters (as implemented in the PRIMODAL code) to generate shape coefficients for each scenario in the scan. The CMB-BEST code, which implements the methods described in [2], then used the *Planck* temperature[1] data to place a constraint on β_{IR}. The bispectrum of each scenario is given an amplitude parameter, which we call f_{NL}, where the scenario prediction

[1]While at present the CMB-BEST code is implemented for temperature only, the methods are general and future implementations will include E-mode polarisation.

corresponds to $f_{NL} = 1$. Thus, if we can rule out $f_{NL} = 1$, we can judge that scenario to be inconsistent with the data. The CMB-BEST code determines the confidence interval of the result by applying the estimator to 140 full focal plane realistic Gaussian simulations [138]. The variance of the results from those Gaussian maps is then assumed to approximate the variance of the result when the estimator is applied to a CMB drawn from a universe with $f_{NL} = 1$.

At present CMB-BEST has been run for the $\mathcal{P}_1^{n_s}$ basis with $p_{\max} = 30$, for the *Planck* 2018 map and simulations, foreground cleaned using SMICA [60]. To make contact with it the PRIMODAL scan was run with the same basis[2]. As shown in Figure 5.1, we find

$$\beta_{IR} \leq 0.46 \qquad (95\%,\ T \text{ only}). \tag{5.1}$$

This translates into a constraint on the sound speed (at horizon crossing) of

$$c_s^* \geq 0.056 \qquad (95\%,\ T \text{ only}). \tag{5.2}$$

Equation (55) in [60] presents the constraint on the sound speed during DBI inflation derived from the CMB bispectrum using the *Modal* estimator

$$c_s^{\rm DBI} \geq 0.079 \qquad (95\%,\ T \text{ only},\ \textit{Modal}) \tag{5.3}$$

using the definition (2.50) and the template (2.58). While our goal was to use this previous constraint as a validation of our own pipeline, it is not unexpected that we did not reproduce it exactly. One possible reason for this is the fact we are using slightly different scales—since we are limited to $k_{\max}/k_{\min} = 1000$ ($k_{\min} = 2.09 \times 10^{-4}\text{Mpc}^{-1}$) for convergence reasons, we cannot use all the scales in the *Planck* data. Another possible reason is convergence in the number of maps[3], with *Modal* using approximately 300 maps while CMB-BEST uses 140. Numerical issues in either method could also contribute. See [2] for an in-depth discussion of these possibilities.

We can however understand the expected method variance of such estimators

[2]The CMB-BEST code has also been run for a *KSW* basis and a Fourier basis, but neither of these can describe features with a complicated shape dependence. Other basis sets described in Chapter 3 will be used in future runs of CMB-BEST.

[3]In addition to the variance of f_{NL}, the maps are used to approximate the full covariance matrix for the linear term in the estimator of [2]. Thus, the estimated value of f_{NL} will only converge once sufficiently many maps are used. However, this correction (which accounts for the effect of the mask) was found to typically be relevant for local-type shapes, and small for equilateral-type shapes, such as in DBI inflation.

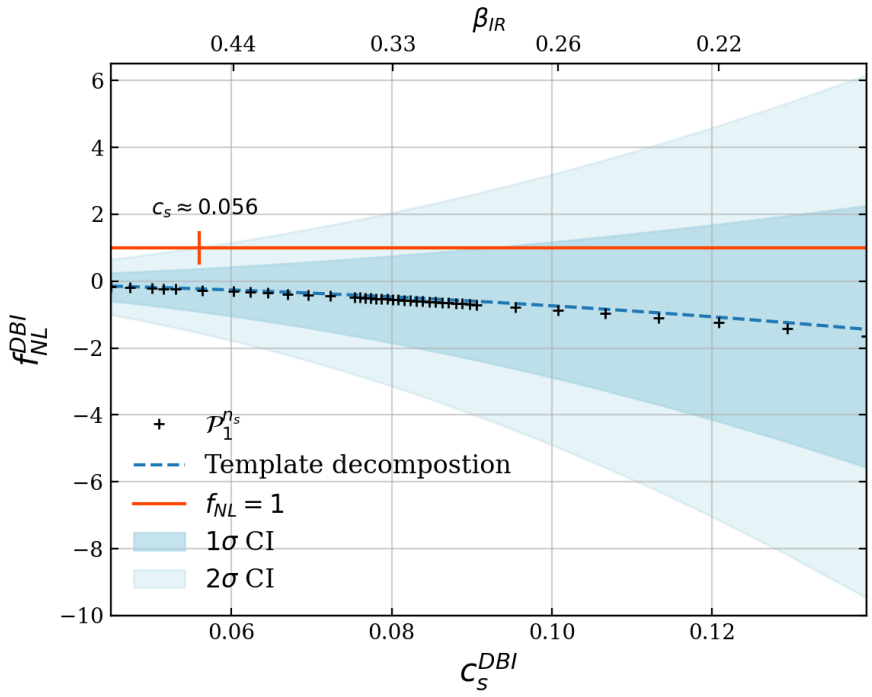


Figure 5.1: The constraint we obtain on DBI inflation. Here c_s^{DBI} is not an input parameter, unlike in the template case. Instead it is time dependent, and the value plotted as c_s^{DBI} here is the value at horizon crossing of the pivot scale, which can be used to label the scenarios we scanned over. We find that $\beta_{IR} < 0.46$ is outside of our 2σ confidence interval. This plot was obtained by scanning across values of β_{IR} and calculating the corresponding primordial bispectra using Primodal, then projecting those bispectra onto the CMB and comparing them to the *Planck* CMB temperature data using CMB-BEst. Since the amplitude is fixed by the scenario, we rule out a scenario by ruling out $f_{NL}^{\mathrm{DBI}} = 1$. Note that the β_{IR} is approximately inversely proportional to c_s^*. The confidence interval is the variance of the estimated f_{NL} obtained from 140 Gaussian maps.

by simply examining the scatter between the estimators used within the *Planck* 2018 analysis. Note that this is not statistical variance, but simply method variance, coming from the difficulty of CMB bispectrum estimation. We show this in Figure 5.2 for the equilateral template (2.56) which is closely related to the DBI template. We see that the result obtained by CMB-BEst is different, but still lies between the results of the *KSW* and *Modal* estimators.

5.2 Connecting to CMB-BEst

The potential we use is the IR one discussed in [88], as we outlined in Section 2.4. See also [71, 89] for further discussions. We use equation (2.39), which has four

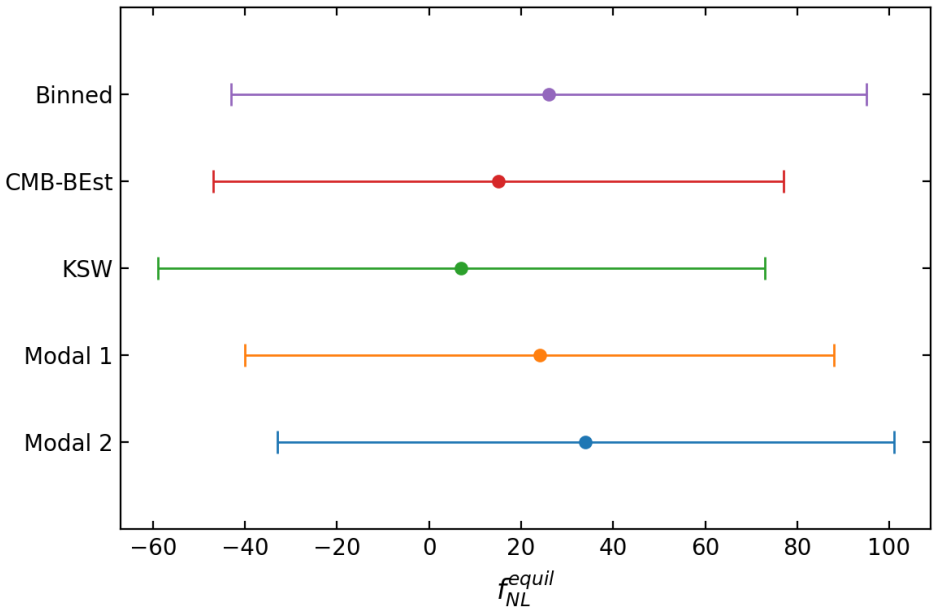


Figure 5.2: A comparison of the various estimation methods used in *Planck*, as presented in table 5 of [60], with the CMB-BEST result, presented in [2]. We plot the constraints obtained by each method for f_{NL}^{equil}, along with their 68% uncertainty margin. Note that the estimation methods are optimal (at least to a very good approximation), hence their uncertainty margins are close to identical. We see that while the estimators do not agree, there is no discrepancy with any significance—in particular, we see that CMB-BEST is not an outlier. Possible contributions to the difference in the central values are convergence in each method (for their respective convergence parameters) and numerical issues. It is important to note that the differences between the different results are not due to statistical variance. The CMB-BEST result quoted here was obtained by decomposing the equilateral template (2.56) (using definition (2.47) for f_{NL}^{equil}) in the $\mathcal{P}_1^{n_s}$ basis for $p_{\max} = 30$.

free parameters, β_{IR}, λ_{DBI}, ϕ_0, and V_0. We use β_{IR} to parameterise our scan, which we take across the range

$$\beta_{IR} \in [0.1885, 0.58]. \tag{5.4}$$

with the parameters

$$\lambda_{\text{DBI}} = 2.00475 \times 10^{15} \tag{5.5}$$

$$V_0 = 5.2 \times 10^{-12} M_{Pl}^4 \tag{5.6}$$

$$\phi_0 = 0.46042 M_{Pl} \tag{5.7}$$

held fixed.

The number of e-folds to the end of inflation is denoted N_e. From [89] we can use slow-roll relations to determine the approximate value of N_e for our values of ϕ as

$$N_e = \frac{\sqrt{\lambda_{\rm DBI}}H}{\phi} \approx 10^2. \tag{5.8}$$

We can also then write down the expected relation between the sound speed and β_{IR}

$$c_s \approx \frac{3}{\beta_{IR}N_e} \tag{5.9}$$

which we indeed find to be approximately true for our numerical results. We start the background evolution on the slow-roll attractor, finding initial conditions which satisfy the Friedmann constraint, as discussed in Section 2.4. For each scenario in this scan we ensure that $\ln(10^{10}A_s)$ is within 3.044 ± 0.014, and that n_s^* is within 0.9649 ± 0.0042. In Table 5.1 we summarise the scenario parameters for the scan, and in Table 5.2 we summarise the resulting scaling indices.

β_{IR}	c_s^*	ε_s^*	ε^*
1.89×10^{-1}	1.39×10^{-1}	8.57×10^{-3}	7.44×10^{-5}
5.80×10^{-1}	4.50×10^{-2}	8.67×10^{-3}	2.31×10^{-4}
β_{IR}	η^*	ϕ^*	H^*
1.89×10^{-1}	2.63×10^{-2}	5.19×10^{-1}	1.31×10^{-6}
5.80×10^{-1}	2.71×10^{-2}	5.15×10^{-1}	1.30×10^{-6}

Table 5.1: Summary of scenario parameters at the upper and lower limits of the scan, evaluated at the horizon crossing of the pivot scale.

β_{IR}	n_s^*	n_{NG}^*
1.89×10^{-1}	9.650×10^{-1}	-8.72×10^{-2}
5.80×10^{-1}	9.637×10^{-1}	-8.99×10^{-2}

Table 5.2: Summary of the slow-roll predictions for the scaling indices at the upper and lower limits of the scan.

5.3 Convergence

It is not obvious how the convergence of the primordial bispectrum translates to convergence of the estimate for f_{NL}, as different momentum configurations will be processed and projected differently. For example, if the primordial bispectrum has equal absolute error in a given equilateral configuration v.s. a squeezed configuration, will the error in each have comparable effect on the final constraint, or will one matter more? We will now discuss the convergence at the primordial level and compare it to the convergence in the final constraint.

In Figure 5.3 we show convergence results for $\mathcal{P}_1^{n_s}$ with $p_{\rm max} = 30$. We quantify the convergence by measuring $\mathcal{E}$, as defined in (3.19), between $p_{\rm max} = 30$ and $p_{\rm max} = 25$. For this basis the match to the template is good in the equilateral limit, but quite poor in the squeezed limit. The convergence in the *scaling* basis is better, with $\mathcal{E}$ falling in the range $[10^{-4}, 10^{-3}]$. However, to connect to the CMB-BEst code we must use the $\mathcal{P}_1^{n_s}$ basis. We see from Figure 5.3 that it is sufficient for our purposes.

In Figure 5.4 we see how this convergence translates to convergence in the final constraint, at the CMB level. We see that across the β_{IR} range, comparing $p_{\rm max} = 25$ and $p_{\rm max} = 30$, the relative error in the quantity that determines the constraint, $f_{NL} + 2\sigma$, is sufficiently small that we can be confident our results have converged in $p_{\rm max}$. We also see that it is (for the majority of the scan) smaller that the corresponding relative error at the primordial level, plotted in Figure 5.3.

When we examine the convergence to the sum (2.60) and product (2.59) scaling templates, in Figure 5.3, we see that neither is obviously the better match to the numerical result. This is due to the numerical result having a non-zero squeezed limit coming from the usual slow-roll suppressed local-type contributions (as in (2.57)) which are neglected in the DBI templates.

5.4 Slow-roll effects

The main slow-roll corrections are a correction to the overall amplitude, a deviation from perfect scale-dependence, and a local-type contribution in the squeezed limit which eventually comes to dominate the primordial comoving curvature bispectrum for sufficiently squeezed triangles.

We can test the impact of the slow-roll effects on the constraint by decomposing the DBI template (2.58) in $\mathcal{P}_1^{n_s}$ and using the result to obtain a constraint on

c_s. The constraint obtained from such a decomposition is plotted as the dashed line in Figure 5.1. Comparing the template decomposition constraint to the constraint obtained using the PRIMODAL coefficients calculated from the in-in formalism (in the same figure) we see no significant difference.

From this we can quantitatively see that the slow-roll suppressed effects, despite becoming dominant in the squeezed limit, do not appreciably affect the constraint on the DBI scenario. This also confirms that the difference between constraints (5.1) and (5.3) is not due to the more accurate shape function used in our method.

5.5 Conclusions

Through the methods described in this thesis (as implemented in the PRIMODAL code) we can calculate the shape coefficients α_{pqr} for each value of β_{IR} in the scan described in Section 5.2. In this scan, these coefficients are with respect to the $\mathcal{P}_1^{n_s}$ basis. Using these coefficients, we can use the methods described in [2] (as implemented in the CMB-BEST code) to place a constraint on the amplitude of that predicted bispectrum. This constraint uses the latest *Planck* temperature data. The constraint then directly translates to a constraint on the inflationary parameters of that scenario, without the need for further approximations. This is done by giving the bispectrum of each scenario an amplitude parameter, which we call f_{NL}—this is then scenario specific and template-free, in contrast to previous works. Since by definition the scenario predicts $f_{NL} = 1$, if it is found that $f_{NL} = 1$ is at 2σ tension for a given scenario, then that scenario is judged to be inconsistent with the data.

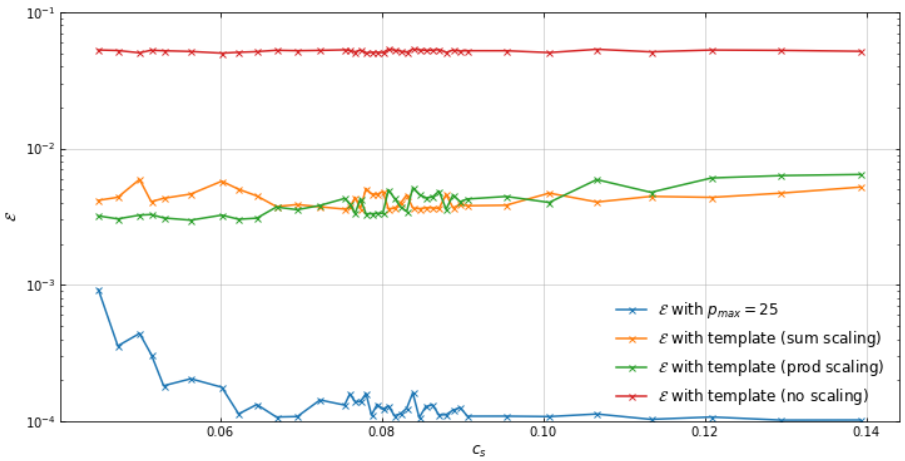
Using this method we investigated the scenarios in the scan described in (5.4) and (5.5). We found that $\beta_{IR} \leq 0.46$ is ruled out at 95% confidence—this is shown in Figure 5.1.

In this chapter we also quantified the convergence of this result, by comparing the full $p_{\max} = 30$ result with the same results calculated for $p_{\max} = 25$. We found that this did not affect the final constraint, and concluded that our result has converged satisfactorily in $p_{\max}$, both at the primordial level and the CMB level. We also quantified the effect of slow-roll corrections on the result, finding them to be negligible, as expected.

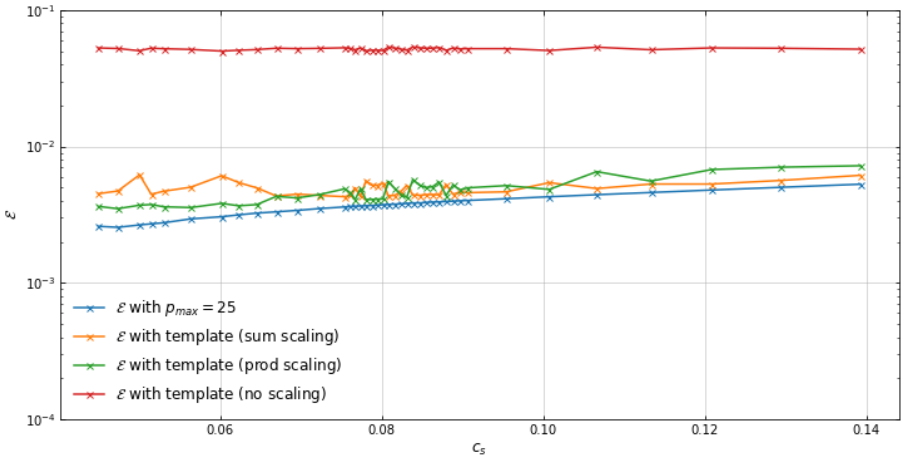
In the context of the same parameter scan, the *Planck* constraint (5.3) would instead imply that $\beta_{IR} \leq 0.33$ is ruled out at 95% confidence. The difference be-

tween our result and the *Planck* result cannot be explained by statistical variance, as they are based on the same *Planck* temperature data. Possible contributions to this disagreement include our limitation on $k_{\rm max}/k_{\rm min}$, convergence (for each method's respective convergence parameters), numerical errors, and differing numbers of Gaussian maps used. However, as we see for f_{NL}^{equil} in Figure 5.2, even between the estimation methods used within *Planck* there is some variance, and we judge the disagreement to be within acceptable bounds.

Thus we conclude that our constraint on the DBI sound speed validates our pipeline as a whole, laying a solid foundation to move forward to shapes that have no standard template, constraining previously unexplored models.



(a) The *scaling* basis



(b) The $\mathcal{P}_1^{n_s}$ basis

Figure 5.3: The *scaling* basis converges well across the scan range. We see that the bare DBI template is a poor match to the true numerical result. This is mostly due to the error in the overall magnitude. Once this is corrected, we see that the numerical result matches the approximate template to better than 1%. As the convergence of the numerical result is better than 0.1% for the *scaling* basis we can see that sum scaling (2.60) and the product scaling (2.59) perform comparably in matching the numerical result. This is mostly due to those templates neglecting the usual slow-roll suppressed contributions (as in (2.57)), which do in fact become relevant to the primordial bispectrum deep enough into the squeezed limit, due to their local-type shape. The $\mathcal{P}_1^{n_s}$ basis is sufficiently convergent across the scan range to obtain the desired constraint. We see that the convergence error is only slightly better than the error in the slow-roll corrected templates. In each case we estimate the convergence at the primordial level by refitting the full result with $\mathcal{P}_1^{n_s}$ at $p_{\max} = 25$, and calculating the relative difference $\mathcal{E}$ between the two, as defined in (3.19).

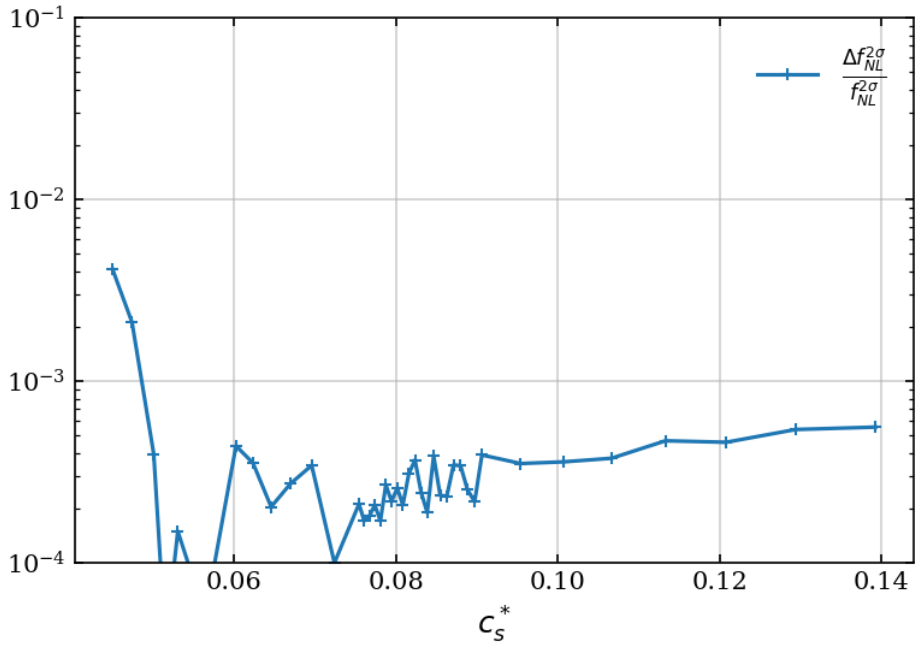


Figure 5.4: We plot the relative error in the value of $f_{NL}^{2\sigma} = f_{NL} + 2\sigma$, between $p_{\max} = 30$ and $p_{\max} = 25$, with f_{NL} obtained from the *Planck* map for each scenario in the scan. We see that the convergence across the majority of the scan is better than that of the convergence in the primordial bispectrum, as plotted in Figure 5.3. This validates the $p_{\max}$ convergence of our pipeline as a whole.

Chapter 6

Conclusions

6.1 Summary

The work detailed in this thesis enables the exploitation of the inherent separability of the tree level in-in formalism, using expansions in separable basis functions. Following the modal philosophy of [84–86], and building on the basic idea of [23], this method obviates the requirement of considering an approximate separable template, which was a limitation of previous analyses. Instead, the goal is to work with the full numerically calculated tree-level inflationary bispectrum of the scenario being considered. The output is not a grid of points, but a set of coefficients of the explicitly separable basis expansion. This preserves the separability and therefore allows us to skip the usual template-approximation step, opening the door to constraining models with non-trivial bispectra. We have:

- developed practical sets of such basis functions, showing the overall method to be feasible by testing them on realistic and interesting shapes.
- developed efficient and robust methods of applying the in-in formalism to an inflation scenario, to generate shape coefficients for some given basis.
- applied these methods to a particular inflation scenario, and used the separability of our result to connect to the methods of [2]. This allowed us to place a constraint on that scenario, using the *Planck* 2018 CMB temperature data.

Through this work we developed this separable approach into a practical and efficient numerical methodology. These methods and basis sets have been implemented in the Primodal code, which calculates the shape coefficients for a

given inflation scenario, in a given basis. We have explored and contrasted the advantages of the various basis sets we described, and thoroughly validated our methods on single-field inflation models with non-trivial phenomenology. This validation shows that our calculation of these coefficients is fast and accurate to high orders.

The latter part of this bispectrum estimation pipeline connects the shape coefficient output of PRIMODAL to the CMB. This part of the pipeline is described and validated in [2] and implemented in the CMB-BEST code. In Chapter 5 we tested the integration of the PRIMODAL and CMB-BEST parts of the pipeline. We scanned across a fundamental parameter, β_{IR}, of the DBI model of inflation. Our goal was to obtain a constraint on this parameter using our template-free analysis—we found $\beta_{IR} \leq 0.46$ at 95% confidence using *Planck* 2018 temperature data. This differs from the equivalent constraint found by the *Modal* estimator in the *Planck* analysis, but the results are sufficiently in agreement that we can take this as a broad validation of our pipeline.

6.2 Discussion

In this work we recast the calculation of the tree-level primordial bispectrum (2.91) into a form that explicitly preserves its separability. We emphasise again that this work has two main advantages over previous numerical methods. The more immediate is that by calculating the primordial bispectrum in terms of an expansion in some basis, the full bispectrum can be obtained much more efficiently than through repetitive integration separately for each k-configuration. The second (and more important) advantage is the link to observations. Unlike previous numerical and semi-analytic methods, once the shape function is expressed in some basis as in (3.1), the integral (2.110) and other computationally intensive steps involved in estimating a particular bispectrum in the CMB can be precomputed. Since this large cost is only paid once per basis, once a basis which converges well for a broad range of models has been found, an extremely broad exploration of primordial bispectra becomes immediately feasible in the CMB.

Our work here goes beyond that of [23] in that our careful methodology allows us to accurately and efficiently go to much higher orders. This allows us to present this method for feature bispectra for the first time, with linear oscillations, logarithmic oscillations, and complicated shape dependence. We

also identified and addressed the effects of the non-physical k-configurations on convergence within the three-dimensional tetrapyd. We explored, for the first time, possible basis set choices in the context of those effects. We showed rapid convergence on a broad range of scenarios, including cases with oscillatory features with non-trivial shape dependence, using our augmented basis sets.

The immediate application of this work will be the efficient exploration of bispectrum phenomenology, as our methods can much more quickly converge to the full shape information than previous numerical methods, which relied on calculating the shape function point-by-point, for each k-configuration separately. We have implemented these methods for single field scenarios with a varying sound speed, scenarios which have a rich feature phenomenology. An important goal, however, will be extending these methods to the case of multiple-field inflation. These goals support those laid out in the recent community white paper [139], by improving predictions of the phenomenology of non-linearities in the very early universe.

While the phenomenology is of interest in its own right, the main advantage of our methods is that it will enable direct constraints of parameters of inflationary scenarios which do not have standard bispectrum templates, thus obtaining new constraints on inflationary parameters. These new constraints will come from exploiting the existing *Planck* data in new ways, and from data from future experiments. Using the basis sets and the separable formulation of the in-in calculation developed in this work, it will be possible to obtain constraints using the full shape dependence, strengthening the connection between the very early universe and the universe we find ourselves in today.

References

[1] P. Clarke and E. P. S. Shellard, *Probing inflation with precision bispectra*, *JCAP* **08** (2021) 002 [2012.08546].

[2] W. Sohn, P. Clarke, J. R. Fergusson and E. P. S. Shellard, *High-resolution CMB bispectrum estimator*, *(in preparation)* .

[3] D. H. Lyth and A. R. Liddle, *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*. Cambridge University Press, 2009, 10.1017/CBO9780511819209.

[4] D. Baumann, *Primordial Cosmology*, *Proceedings of Science* **TASI2017** (2018) 009 [1807.03098].

[5] C. Burrage, R. H. Ribeiro and D. Seery, *Large slow-roll corrections to the bispectrum of noncanonical inflation*, *Journal of Cosmology and Astro-Particle Physics* **2011** (2011) 032 [1103.4126].

[6] E. A. Lim, *Advanced cosmology : Primordial non-gaussianities*, 2012. https://nms.kcl.ac.uk/eugene.lim/AdvCos/lecture2.pdf.

[7] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau et al., *Array programming with NumPy*, *Nature* **585** (2020) 357.

[8] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* **17** (2020) 261.

[9] J. D. Hunter, *Matplotlib: A 2d graphics environment*, *Computing in Science & Engineering* **9** (2007) 90.

[10] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[11] D. J. Fixsen, *The temperature of the cosmic microwave background*, *The Astrophysical Journal* **707** (2009) 916–920.

[12] P. J. E. Peebles, *Primordial Helium Abundance and the Primordial Fireball. II*, *The Astrophysical Journal* **146** (1966) 542.

[13] P. J. E. Peebles and J. T. Yu, *Primeval Adiabatic Perturbation in an Expanding Universe*, *The Astrophysical Journal* **162** (1970) 815.

[14] L. Verde, T. Treu and A. G. Riess, *Tensions between the Early and the Late Universe*, *Nature Astron.* **3** (2019) 891 [1907.10625].

[15] W. L. Freedman, *Measurements of the Hubble Constant: Tensions in Perspective*, 2106.15656.

[16] L. Senatore, K. M. Smith and M. Zaldarriaga, *Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data*, *JCAP* **01** (2010) 028 [0905.3746].

[17] A. J. Christopherson and K. A. Malik, *The non-adiabatic pressure in general scalar field systems*, *Phys. Lett. B* **675** (2009) 159 [0809.3518].

[18] M. Alishahiha, E. Silverstein and D. Tong, *DBI in the sky: Non-Gaussianity from inflation with a speed limit*, *Physical Review D* **70** (2004) 123505 [hep-th/0404084].

[19] W. Hu, *Generalized slow roll for noncanonical kinetic terms*, *Physical Review D* **84** (2011) 027303 [1104.4500].

[20] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown et al., *Planck 2015 results. XX. Constraints on inflation*, *Astronomy and Astrophysics* **594** (2016) A20 [1502.02114].

[21] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. X. Constraints on inflation*, *arXiv e-prints* (2018) [1807.06211].

[22] S. Weinberg, *Cosmology*. 2008.

[23] H. Funakoshi and S. Renaux-Petel, *A Modal Approach to the Numerical Calculation of Primordial non-Gaussianities*, *JCAP* **1302** (2013) 002 [arxiv/1211.3086].

[24] J. Maldacena, *Non-gaussian features of primordial fluctuations in single field inflationary models*, *Journal of High Energy Physics* **5** (2003) 013 [astro-ph/0210603].

[25] C. T. Byrnes and K.-Y. Choi, *Review of local non-Gaussianity from multi-field inflation*, *Adv. Astron.* **2010** (2010) 724525 [1002.3110].

[26] X. Gao, D. Langlois and S. Mizuno, *Oscillatory features in the curvature power spectrum after a sudden turn of the inflationary trajectory*, *Journal of Cosmology and Astroparticle Physics* **10** (2013) 023 [1306.5680].

[27] A. Achúcarro and G. A. Palma, *The string swampland constraints require multi-field inflation*, *JCAP* **02** (2019) 041 [1807.04390].

[28] A. Achúcarro, S. Céspedes, A.-C. Davis and G. A. Palma, *Constraints on Holographic Multifield Inflation and Models Based on the Hamilton-Jacobi Formalism*, *Phys. Rev. Lett.* **122** (2019) 191301 [1809.05341].

[29] J. Torrado, B. Hu and A. Achúcarro, *Robust predictions for an oscillatory bispectrum in Planck 2015 data from transient reductions in the speed of sound of the inflaton*, *Physical Review D* **96** (2017) 083515 [1611.10350].

[30] A. Achúcarro, V. Atal, M. Kawasaki and F. Takahashi, *The two-field regime of natural inflation*, *Journal of Cosmology and Astroparticle Physics* **12** (2015) 044 [1510.08775].

[31] A. Achúcarro, V. Atal and Y. Welling, *On the viability of $m^2\phi^2$ and natural inflation*, *Journal of Cosmology and Astroparticle Physics* **7** (2015) 008 [1503.07486].

[32] A. Achúcarro, V. Atal, B. Hu, P. Ortiz and J. Torrado, *Inflation with moderately sharp features in the speed of sound: Generalized slow roll and in-in formalism for power spectrum and bispectrum*, *Physical Review D* **90** (2014) 023511 [1404.7522].

[33] A. Achúcarro, V. Atal, P. Ortiz and J. Torrado, *Localized correlated features in the CMB power spectrum and primordial bispectrum from a transient reduction in the speed of sound*, *Physical Review D* **89** (2014) 103006 [1311.2552].

[34] A. Achúcarro, J.-O. Gong, G. A. Palma and S. P. Patil, *Correlating features in the primordial spectra*, *Physical Review D* **87** (2013) 121301 [1211.5619].

[35] A. Achúcarro, V. Atal, S. Céspedes, J.-O. Gong, G. A. Palma and S. P. Patil, *Heavy fields, reduced speeds of sound, and decoupling during inflation*, *Physical Review D* **86** (2012) 121301 [1205.0710].

[36] A. Achúcarro, J.-O. Gong, S. Hardeman, G. A. Palma and S. P. Patil, *Effective theories of single field inflation when heavy fields matter*, *Journal of High Energy Physics* **5** (2012) 66 [1201.6342].

[37] C. T. Byrnes, M. Cortês and A. R. Liddle, *Comprehensive analysis of the simplest curvaton model*, *Physical Review D* **90** (2014) 023523 [1403.4591].

[38] P. Adshead, C. Dvorkin, W. Hu and E. A. Lim, *Non-Gaussianity from step features in the inflationary potential*, *Physical Review D* **85** (2012) 023531 [1110.3050].

[39] C. Dvorkin and W. Hu, *Generalized slow roll approximation for large power spectrum features*, *Physical Review D* **81** (2010) 023518 [0910.2237].

[40] C. P. Novaes, M. Benetti and A. Bernui, *Primordial Non-Gaussianities of inflationary step-like models*, *arXiv e-prints* (2015) [1507.01657].

[41] R. Flauger and E. Pajer, *Resonant Non-Gaussianity*, *JCAP* **1101** (2011) 017 [1002.0833].

[42] E. Pajer and M. Peloso, *A review of Axion Inflation in the era of Planck*, *Class. Quant. Grav.* **30** (2013) 214002 [1305.3557].

[43] P. D. Meerburg and E. Pajer, *Observational Constraints on Gauge Field Production in Axion Inflation*, *JCAP* **1302** (2013) 017 [1203.6076].

[44] P. D. Meerburg, *Oscillations in the primordial bispectrum: Mode expansion*, *Physical Review D* **82** (2010) 063517 [1006.2771].

[45] P. D. Meerburg, *Oscillations in the primordial bispectrum*, *J. Phys. Conf. Ser.* **259** (2010) 012049 [1010.2234].

[46] N. Barnaby, E. Pajer and M. Peloso, *Gauge Field Production in Axion Inflation: Consequences for Monodromy, non-Gaussianity in the CMB, and Gravitational Waves at Interferometers*, *Phys. Rev.* **D85** (2012) 023525 [1110.3327].

[47] H. Peiris, R. Easther and R. Flauger, *Constraining Monodromy Inflation*, *JCAP* **1309** (2013) 018 [1303.2616].

[48] R. Easther and R. Flauger, *Planck Constraints on Monodromy Inflation*, *JCAP* **1402** (2014) 037 [1308.3736].

[49] G. Cabass, E. Pajer and F. Schmidt, *Imprints of Oscillatory Bispectra on Galaxy Clustering*, 1804.07295.

[50] S. R. Behbahani, A. Dymarsky, M. Mirbabayi and L. Senatore, *(Small) Resonant non-Gaussianities: Signatures of a Discrete Shift Symmetry in the Effective Field Theory of Inflation*, *JCAP* **1212** (2012) 036 [1111.3373].

[51] C. T. Byrnes, E. J. Copeland and A. M. Green, *Primordial black holes as a tool for constraining non-Gaussianity*, *Physical Review D* **86** (2012) 043512 [1206.4188].

[52] S. Young and C. T. Byrnes, *Primordial black holes in non-Gaussian regimes*, *Journal of Cosmology and Astro-Particle Physics* **2013** (2013) 052 [1307.4995].

[53] G. Franciolini, A. Kehagias, S. Matarrese and A. Riotto, *Primordial black holes from inflation and non-Gaussianity*, *Journal of Cosmology and Astro-Particle Physics* **2018** (2018) 016 [1801.09415].

[54] S. Passaglia, W. Hu and H. Motohashi, *Primordial black holes and local non-Gaussianity in canonical inflation*, *Physical Review D* **99** (2019) 043536 [1812.08243].

[55] P. Creminelli and M. Zaldarriaga, *A single-field consistency relation for the three-point function*, *Journal of Cosmology and Astroparticle Physics* **10** (2004) 006 [astro-ph/0407059].

[56] S. Garcia-Saenz and S. Renaux-Petel, *Flattened non-Gaussianities from the effective field theory of inflation with imaginary speed of sound*, *ArXiv e-prints* (2018) [1805.12563].

[57] J. Fumagalli, S. Garcia-Saenz, L. Pinol, S. Renaux-Petel and J. Ronayne, *Hyper-Non-Gaussianities in Inflation with Strongly Nongeodesic Motion*, *Phys. Rev. Lett.* **123** (2019) 201302 [1902.03221].

[58] A. J. Tolley and M. Wyman, *The Gelaton Scenario: Equilateral non-Gaussianity from multi-field dynamics*, *Phys. Rev. D* **81** (2010) 043502 [0910.1853].

[59] E. Komatsu, D. N. Spergel and B. D. Wandelt, *Measuring primordial non-Gaussianity in the cosmic microwave background*, *Astrophys. J.* **634** (2005) 14 [astro-ph/0305189].

[60] Planck collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [1905.05697].

[61] E. Silverstein and D. Tong, *Scalar speed limits and cosmology: Acceleration from D-cceleration*, *Physical Review D* **70** (2004) 103505 [hep-th/0310221].

[62] R. Flauger, L. McAllister, E. Pajer, A. Westphal and G. Xu, *Oscillations in the CMB from axion monodromy inflation*, *Journal of Cosmology and Astroparticle Physics* **6** (2010) 009 [0907.2916].

[63] R. Flauger, L. McAllister, E. Silverstein and A. Westphal, *Drifting Oscillations in Axion Monodromy*, *JCAP* **1710** (2017) 055 [1412.1814].

[64] X. Chen, R. Easther and E. A. Lim, *Large non-Gaussianities in single-field inflation*, *Journal of Cosmology and Astroparticle Physics* **6** (2007) 023 [astro-ph/0611645].

[65] X. Chen, R. Easther and E. A. Lim, *Generation and Characterization of Large Non-Gaussianities in Single Field Inflation*, *Journal of Cosmology and Astroparticle Physics* **0804** (2008) 010 [0801.3295].

[66] X. Chen, M.-x. Huang, S. Kachru and G. Shiu, *Observational signatures and non-Gaussianities of general single-field inflation*, *Journal of Cosmology and Astroparticle Physics* **1** (2007) 002 [hep-th/0605045].

[67] D. Seery and J. E. Lidsey, *Primordial non-Gaussianities in single-field inflation*, *Journal of Cosmology and Astroparticle Physics* **6** (2005) 003 [astro-ph/0503692].

[68] N. Bartolo, D. Cannone and S. Matarrese, *The effective field theory of inflation models with sharp features*, *Journal of Cosmology and Astro-Particle Physics* **2013** (2013) 038 [1307.3483].

[69] S. Passaglia and W. Hu, *Scalar bispectrum beyond slow-roll in the unified EFT of inflation*, *Physical Review D* **98** (2018) 023526 [1804.07741].

[70] P. Adshead, W. Hu and V. Miranda, *Bispectrum in single-field inflation beyond slow-roll*, *Physical Review D* **88** (2013) 023507 [1303.7004].

[71] V. Miranda, W. Hu and P. Adshead, *Warp features in DBI inflation*, *Physical Review D* **86** (2012) 063529 [1207.2186].

[72] V. Miranda and W. Hu, *Inflationary steps in the Planck data*, *Physical Review D* **89** (2014) 083529 [1312.0946].

[73] X. Chen, *Folded resonant non-Gaussianity in general single field inflation*, *Journal of Cosmology and Astroparticle Physics* **12** (2010) 003 [1008.2485].

[74] S. Ávila, J. Martin and D. A. Steer, *Superimposed oscillations in brane inflation*, *Journal of Cosmology and Astro-Particle Physics* **2014** (2014) 032 [1304.3262].

[75] S. Garcia-Saenz, S. Renaux-Petel and J. Ronayne, *Primordial fluctuations and non-Gaussianities in sidetracked inflation*, *JCAP* **07** (2018) 057 [1804.11279].

[76] M. Münchmeyer, F. Bouchet, M. G. Jackson and B. Wandelt, *The Komatsu Spergel Wandelt estimator for oscillations in the cosmic microwave background bispectrum*, *Astron. Astrophys.* **570** (2014) A94 [1405.2550].

[77] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown et al., *Planck 2015 results. XVII. Constraints on primordial non-Gaussianity*, *Astronomy and Astrophysics* **594** (2016) A17 [1502.01592].

[78] Simons Observatory collaboration, *The Simons Observatory: Science goals and forecasts*, *JCAP* **02** (2019) 056 [1808.07445].

[79] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold et al., *Cmb-s4 science book, first edition*, 2016.

[80] T. Baldauf, U. Seljak and L. Senatore, *Primordial non-Gaussianity in the bispectrum of the halo density field*, *Journal of Cosmology and Astro-Particle Physics* **2011** (2011) 006 [1011.1513].

[81] D. Karagiannis, A. Lazanu, M. Liguori, A. Raccanelli, N. Bartolo and L. Verde, *Constraining primordial non-Gaussianity with bispectrum and power spectrum from upcoming optical and radio surveys*, *Monthly Notices of the Royal Astronomical Society* **478** (2018) 1341 [1801.09280].

[82] X. Chen, C. Dvorkin, Z. Huang, M. H. Namjoo and L. Verde, *The future of primordial features with large-scale structure surveys*, *Journal of Cosmology and Astroparticle Physics* **11** (2016) 014 [1605.09365].

[83] R. Scoccimarro, L. Hui, M. Manera and K. C. Chan, *Large-scale Bias and Efficient Generation of Initial Conditions for Non-Local Primordial Non-Gaussianity*, *Phys. Rev. D* **85** (2012) 083002 [1108.5512].

[84] J. R. Fergusson and E. P. S. Shellard, *Primordial non-Gaussianity and the CMB bispectrum*, *Physical Review D* **76** (2007) 083523 [astro-ph/0612713].

[85] J. R. Fergusson and E. P. S. Shellard, *Shape of primordial non-Gaussianity and the CMB bispectrum*, *Physical Review D* **80** (2009) 043510 [0812.3413].

[86] J. R. Fergusson, M. Liguori and E. P. S. Shellard, *The CMB bispectrum*, *Journal of Cosmology and Astroparticle Physics* **12** (2012) 032 [arxiv/1006.1642].

[87] D. H. Lyth and D. Wands, *Conserved cosmological perturbations*, *Phys. Rev. D* **68** (2003) 103515 [astro-ph/0306498].

[88] R. Bean, X. Chen, H. Peiris and J. Xu, *Comparing Infrared Dirac-Born-Infeld Brane Inflation to Observations*, *Phys. Rev. D* **77** (2008) 023527 [0710.1812].

[89] X. Chen, *Running non-Gaussianities in DBI inflation*, *Phys. Rev. D* **72** (2005) 123518 [astro-ph/0507053].

[90] J. Garriga and V. F. Mukhanov, *Perturbations in k-inflation*, *Physics Letters B* **458** (1999) 219 [hep-th/9904176].

[91] M. J. Mortonson, C. Dvorkin, H. V. Peiris and W. Hu, *CMB polarization features from inflation versus reionization*, *Physical Review D* **79** (2009) 103519 [0903.4920].

[92] Planck collaboration, *Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **571** (2014) A24 [1303.5084].

[93] M. Dias, J. Frazer, D. J. Mulryne and D. Seery, *Numerical evaluation of the bispectrum in multiple field inflation—the transport approach with code*, *JCAP* **1612** (2016) 033 [1609.00379].

[94] J. R. Fergusson, M. Liguori and E. P. S. Shellard, *General CMB and primordial bispectrum estimation: Mode expansion, map making, and measures of F_{NL}*, *Physical Review D* **82** (2010) 023502 [0912.5516].

[95] E. Pajer, *Building a Boostless Bootstrap for the Bispectrum*, *JCAP* **01** (2021) 023 [2010.12818].

[96] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan and L. Senatore, *The Effective Field Theory of Inflation*, *JHEP* **03** (2008) 014 [0709.0293].

[97] D. Baumann and D. Green, *Equilateral Non-Gaussianity and New Physics on the Horizon*, *JCAP* **09** (2011) 014 [1102.5343].

[98] G. Cabass, E. Pajer and F. Schmidt, *How Gaussian can our Universe be?*, *JCAP* **1701** (2017) 003 [1612.00033].

[99] J. Adams, B. Cresswell and R. Easther, *Inflationary perturbations from a potential with a step*, *Physical Review D* **64** (2001) 123514 [astro-ph/0102236].

[100] S. Weinberg, *Quantum contributions to cosmological correlations*, *Physical Review D* **72** (2005) 043514 [hep-th/0506236].

[101] P. Adshead, R. Easther and E. A. Lim, *Cosmology With Many Light Scalar Fields: Stochastic Inflation and Loop Corrections*, *Phys. Rev. D* **79** (2009) 063504 [0809.4008].

[102] X. Chen, *Primordial Non-Gaussianities from Inflation Models*, *Adv. Astron.* **2010** (2010) 638979 [1002.1416].

[103] D. Babich, P. Creminelli and M. Zaldarriaga, *The Shape of non-Gaussianities*, *JCAP* **08** (2004) 009 [astro-ph/0405356].

[104] S. Renaux-Petel, *Primordial non-Gaussianities after Planck 2015: an introductory review*, *Comptes Rendus Physique* **16** (2015) 969 [1508.06740].

[105] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz and X. Chen, *Prospects for Cosmological Collider Physics*, *JCAP* **1703** (2017) 050 [1610.06559].

[106] S. Renaux-Petel, *On the redundancy of operators and the bispectrum in the most general second-order scalar-tensor theory*, *Journal of Cosmology and Astro-Particle Physics* **2012** (2012) 020 [1107.5020].

[107] P. Creminelli, G. D'Amico, M. Musso and J. Noreña, *The (not so) squeezed limit of the primordial 3-point function*, *Journal of Cosmology and Astroparticle Physics* **11** (2011) 038 [1106.1462].

[108] J. R. Fergusson, H. F. Gruetjen, E. P. S. Shellard and B. Wallisch, *Polyspectra searches for sharp oscillatory features in cosmic microwave sky data*, *Phys. Rev. D* **91** (2015) 123506 [1412.6152].

[109] J. R. Fergusson, *Efficient optimal non-Gaussian CMB estimators with polarisation*, *Phys. Rev. D* **90** (2014) 043533 [1403.7949].

[110] D. Babich, *Optimal estimation of non-Gaussianity*, *Phys. Rev. D* **72** (2005) 043003 [astro-ph/0503375].

[111] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark and M. Zaldarriaga, *Limits on non-gaussianities from wmap data*, *JCAP* **05** (2006) 004 [astro-ph/0509029].

[112] K. M. Smith and M. Zaldarriaga, *Algorithms for bispectra: Forecasting, optimal analysis, and simulation*, *Mon. Not. Roy. Astron. Soc.* **417** (2011) 2 [astro-ph/0612571].

[113] M. Münchmeyer, P. D. Meerburg and B. D. Wandelt, *Optimal estimator for resonance bispectra in the CMB*, *Physical Review D* **91** (2015) 043534 [1412.3461].

[114] P. D. Meerburg and M. Münchmeyer, *Optimal CMB estimators for bispectra from excited states*, *Phys. Rev.* **D92** (2015) 063527 [1505.05882].

[115] J. R. Fergusson, H. F. Gruetjen, E. P. S. Shellard and M. Liguori, *Combining power spectrum and bispectrum measurements to detect oscillatory features*, *Phys. Rev.* **D91** (2015) 023502 [1410.5114].

[116] J. Byun and R. Bean, *Non-Gaussian Shape Recognition*, *JCAP* **1309** (2013) 026 [1303.3050].

[117] J. Byun, N. Agarwal, R. Bean and R. Holman, *Looking for non-Gaussianity in all the right places: A new basis for nonseparable bispectra*, *Phys. Rev.* **D91** (2015) 123518 [1504.01394].

[118] T. Battefeld and J. Grieb, *Anatomy of bispectra in general single-field inflation — Modal expansions*, *Journal of Cosmology and Astro-Particle Physics* **2011** (2011) 003 [1110.1369].

[119] D. K. Hazra, L. Sriramkumar and J. Martin, *BINGO: a code for the efficient computation of the scalar bi-spectrum*, *Journal of Cosmology and Astroparticle Physics* **5** (2013) 026 [1201.0926].

[120] J. S. Horner and C. R. Contaldi, *Non-Gaussianity Unleashed*, *ArXiv e-prints* (2013) [1303.2119].

[121] J. S. Horner and C. R. Contaldi, *Non-gaussian signatures of general inflationary trajectories*, *Journal of Cosmology and Astroparticle Physics* **9** (2014) 001 [1311.3224].

[122] J. S. Horner and C. R. Contaldi, *The bispectrum of single-field inflationary trajectories with $c_s \neq 1$*, *ArXiv e-prints* (2015) [1503.08103].

[123] D. J. Mulryne and J. W. Ronayne, *PyTransport: A Python package for the calculation of inflationary correlation functions*, *J. Open Source Softw.* **3** (2018) 494 [1609.00381].

[124] J. W. Ronayne and D. J. Mulryne, *Numerically evaluating the bispectrum in curved field-space— with PyTransport 2.0*, *JCAP* **1801** (2018) 023 [1708.07130].

[125] S. Butchers and D. Seery, *Numerical evaluation of inflationary 3-point functions on curved field space—with the transport method & CppTransport*, *JCAP* **07** (2018) 031 [1803.10563].

[126] K. Marzouk, A. Maraio and D. Seery, *Non-Gaussianity in D3-brane inflation*, 2105.03637.

[127] J. Hung, J. R. Fergusson and E. P. S. Shellard, *Advancing the matter bispectrum estimation of large-scale structure: a comparison of dark matter codes*, *arXiv e-prints* (2019) [1902.01830].

[128] N. Arkani-Hamed, D. Baumann, H. Lee and G. L. Pimentel, *The Cosmological Bootstrap: Inflationary Correlators from Symmetries and Singularities*, *JHEP* **04** (2020) 105 [1811.00024].

[129] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, *SIAM Review* **51** (2009) 455.

[130] C. Ringeval, *The exact numerical treatment of inflationary models*, *Lect. Notes Phys.* **738** (2008) 243 [astro-ph/0703486].

[131] W. I. J. Haddadin and W. J. Handley, *Rapid numerical solutions for the Mukhanov-Sazaki equation*, *ArXiv e-prints* (2018) [1809.11095].

[132] F. J. Agocs, W. J. Handley, A. N. Lasenby and M. P. Hobson, *Efficient method for solving highly oscillatory ordinary differential equations with applications to physical systems*, *Phys. Rev. Res.* **2** (2020) 013030 [1906.01421].

[133] F. J. Agocs, M. P. Hobson, W. J. Handley and A. N. Lasenby, *Dense output for highly oscillatory numerical solutions*, *arXiv e-prints* (2020) arXiv:2007.05013 [2007.05013].

[134] A. Dixit, L. Jiu, V. H. Moll and C. Vignat, *The finite Fourier transform of classical polynomials*, *ArXiv e-prints* (2014) [1402.5544].

[135] A. Iserles and S. P. Norsett, *Efficient quadrature of highly efficient integrals using deriatives*, *Royal Society Publishing, Volume 461, Issue 2057* (2005) .

[136] F. Arroja and T. Tanaka, *A note on the role of the boundary terms for the non-Gaussianity in general k-inflation*, *Journal of Cosmology and Astro-Particle Physics* **2011** (2011) 005 [1103.1102].

[137] G. Rigopoulos, *Gauge invariance and non-Gaussianity in inflation*, *Physical Review D* **84** (2011) 021301 [1104.0292].

[138] Planck collaboration, *Planck 2015 results. XII. Full Focal Plane simulations*, *Astron. Astrophys.* **594** (2016) A12 [1509.06348].

[139] P. D. Meerburg and et al, *Primordial Non-Gaussianity*, *arXiv e-prints* (2019) arXiv:1903.04409 [1903.04409].

List of Figures

List of Tables