



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
August, 2021

Statement of contributions

In this thesis I describe the research I performed during my PhD at the University of Cambridge, supervised by Professor Paul Shellard. Chapters 3, 4 and 5 describe my original work, except where specified below, and in the text.

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, however. 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 and **words**. This work will be published in [2].

Numpy, Scipy, Jupyter Notebooks, QUADPTS! Cite [3, 4].

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 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.

Acknowledgements

Write this!

My acknowledgements...

We make particular use of [5–8] throughout.

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.

I am grateful to my college for reasons.

Contents

1	Introduction I: Cosmology	15
1.1	The standard cosmological history	15
1.2	Initial conditions for ΛCDM	21
1.3	Statistical observables	32
1.4	Thesis outline	35
2	Introduction II: Probing Inflation	39
2.1	Perturbations about the background	41
2.2	The primordial power spectrum	46
2.3	The primordial bispectrum	48
2.4	A field guide to the definitions of f_{NL}	49
2.5	Step 1: the interaction Hamiltonian	52
2.6	Step 2: the primordial bispectrum	54
2.7	Step 3: the CMB bispectrum	61
2.8	Step 4: CMB bispectrum estimation	63
2.9	Previous work on in-in separability.	67
2.10	Configuration-by-configuration codes	67
2.11	Separable approximations to non-separable templates	69
2.12	CMB-BEST	70
3	Decomposing Primordial Shapes	71
3.1	Setting up the formalism	71
3.2	Testing basis sets	78
3.3	Building basis sets	79
3.4	The <i>scaling</i> basis	92
3.5	The <i>resonant</i> basis	93
3.6	A tradeoff between p_{\max} and k_{\max}/k_{\min}	98
3.7	Conclusions	98

4 Methods and Validation	101
4.1 Numerics of mode evolution	101
4.2 Integration weights	102
4.3 Decompositions	104
4.4 Starting the integration with a pinch	105
4.5 The interaction Hamiltonian	109
4.6 Stopping the integration	110
4.7 Validation	111
4.8 Conclusions	121
5 An Inflationary Constraint from the CMB Bispectrum	125
5.1 DBI sound speed constraint	125
5.2 Connecting to CMB-BEST	128
5.3 Convergence	131
5.4 Slow-roll effects	132
5.5 Conclusions	132
6 Conclusions and Future Work	133
6.1 Summary	133
6.2 Discussion	134
6.3 Future Work	135
References	139

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 very 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—whereas today we have distinct galaxies and the empty spaces in between, 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, the CMB) to constrain physical theories of the very early universe (the epoch of 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.

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, a straight line in spacetime—with spacetime being described in terms of curved geometry. 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 defines our notion of distance, and thus encodes the causal structure of the universe—which events can influence which other events.

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 to the Einstein equations. The metric of each universe is known as a Friedmann–Lemaître–Robertson–Walker metric (or more usually, and FLRW metric) and in this context the Einstein equations become the Friedmann equations, which we will soon present.

The form of the FLRW metric (we will always set the speed of light to unity¹) is

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a(t)^2 dS_3^2 \quad (1.1)$$

where $a(t)$ will have the interpretation as the scale factor which describes the evolution of the universe. The line element ds^2 then determines the distance between two spacetime points, or more intuitively, the proper time measured by an observer who freely falls between two events will be the integral of $d\tau = \sqrt{-g_{\mu\nu}dx^\mu dx^\nu}$. The homogeneity, isotropy and curvature of space are encoded in dS_3 , but a non-trivial time dependence of $a(t)$ breaks the homogeneity of time. We will only consider a flat universe, so $dS_3^2 = dx^2 + dy^2 + dz^2$.

The components of the matter and energy content of the universe enter the equations through their stress-energy tensors, which we approximate as that of a

¹We will also set $\hbar = M_P^{-1} = \sqrt{8\pi G} = 1$.

perfect fluid (denoted X)

$$T_X^{\mu\nu} = (\rho_X + P_X)U_X^\mu U_X^\nu + P_X g^{\mu\nu}, \quad (1.2)$$

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 \mathbf{x} , 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 quantity w takes the values 0, 1/3 and -1 for matter, radiation and dark energy (described by a cosmological constant) respectively. If we take the spatial curvature of the universe to be vanishingly small, the Einstein equation are then

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \quad (1.3)$$

where $G_{\mu\nu}$ is a complicated function of $g_{\mu\nu}$. Fortunately, when $g_{\mu\nu}$ takes the form (1.1), these equations simplify to the promised Friedmann equations (with $8\pi G$ temporarily reinstated)

$$H^2 = \frac{8\pi G\rho}{3}, \quad (1.4)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P), \quad (1.5)$$

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. A dot will refer to a derivative with respect to coordinate time t . We also have the continuity equation

$$\dot{\rho}_X + 3\frac{\dot{a}}{a}(\rho_X + P_X) = 0 \quad (1.6)$$

which is satisfied for each component separately.

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 . This defines the homogeneous background evolution of the universe—physical phenomena such as the redshifting of photons as they freestream through the universe. Photons travel

on geodesics. The geodesic equation that we use to determine their evolution is

$$\frac{dP^\mu}{d\lambda} + \Gamma_{\nu\sigma}^\mu P^\nu P^\sigma = 0 \quad (1.7)$$

where P^a is the photon's 4-momentum, and λ is an affine parameter for the path. The Christoffel symbols, $\Gamma_{\nu\sigma}^\mu$ are functions of the metric, and are used to build the $G_{\mu\nu}$ that appeared in the Einstein equations (1.3)

$$\Gamma_{\nu\sigma}^\mu = \frac{1}{2} g^{\mu\gamma} \left(\frac{\partial g_{\gamma\nu}}{\partial x^\sigma} + \frac{\partial g_{\gamma\sigma}}{\partial x^\nu} - \frac{\partial g_{\nu\sigma}}{\partial x^\gamma} \right) \quad (1.8)$$

Using the chain rule, (1.7) can be rewritten in terms of t

$$P^0 \frac{\partial P^\mu}{\partial t} + \Gamma_{\nu\sigma}^\mu P^\nu P^\sigma = 0. \quad (1.9)$$

Writing the components of the photon's null vector 4-momentum P^μ as (p, \mathbf{p}) , and evaluating (1.8) for the metric (1.1) we get that

$$p \propto \frac{1}{a}. \quad (1.10)$$

The interpretation of this result is that as the universe expands, the photon's measured energy decreases, and its measured wavelength increases. The geodesic equation (1.7) encodes many other physical effects when applied to the evolution of perturbations around the homogeneous background (i.e. allowing $g_{\mu\nu}$ to deviate from homogeneity) allowing us to understand how the statistics of the perturbations at the start of the ΛCDM evolution evolved into the perturbations we see today.

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.

Using (1.6) and the equation of state (w , summarised in table 1.1) for matter, radiation and dark energy (described as a cosmological constant Λ) we can solve for the dependence of ρ_X on a . 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. For Λ , we find $\rho_\Lambda \propto a^0$, by design. The Friedmann equation (1.4) then gives the time dependence of $a(t)$, again summarised in table 1.1.

Using this knowledge we can rewrite the Friedmann equation (1.4) as

$$H^2 = H_0^2 \left(\frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \Omega_\Lambda \right) \quad (1.11)$$

where we define the fractional density

$$\Omega_X = \frac{\rho_X}{\rho_{crit,0}} \quad (1.12)$$

for X being radiation, matter and dark energy. A subscript of 0 will denote a quantity evaluated today. The critical density $\rho_{crit,0}$ is the density for which the universe is flat, so for our universe $\sum \Omega_i = 1$. We will choose our scaling such that $a_0 = 1$.

The measured values for these quantities, as presented in [9], are

$$\Omega_{m,0} = 0.3111 \pm 0.0056, \quad (1.13)$$

$$\Omega_\Lambda = 0.6889 \pm 0.0056. \quad (1.14)$$

The temperature of the blackbody CMB radiation is also found to be $T_\gamma = 2.725 \pm 0.001 K$, from which we can calculate $\Omega_{\gamma,0} = 9.4 \times 10^{-5}$.

From these numbers, and (1.11), we can make some statements about the past and future of the universe, illustrated in figure 1.1. We can see that while Λ is dominant now, the transition from matter domination to Λ domination happened relatively recently in the history of the universe. 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. Running the clock back far enough, we see that despite $\Omega_{\gamma,0} \ll \Omega_{m,0}$, at one stage it was the case that $\Omega_\gamma \gg \Omega_m$, and the universe was dominated by radiation. In that early stage of the universe, the temperature was high enough that the photons were tightly coupled to the baryons. The photon pressure supported pressure waves in that nearly-homogeneous fluid, and the coupling meant that the baryons were dragged along with those waves. Once the temperature and density dropped sufficiently that this was no longer the case, the baryons were

dropped, and instead began to fall back into the gravitational wells that the dark matter had meanwhile been forming. The imprints of those pressure waves is thus seen in the positions of galaxies today (in the statistics of their separations) and in the acoustic oscillations of the CMB.

The ΛCDM model of the history of the universe assumes that at some point in the 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.

Radiation is one of the major known components of the present universe, in the form of the photons in the cosmic microwave background, the CMB. Another is matter, by which we mean *dark* matter, which is invisible to our telescopes (which depend on electromagnetic radiation) but can be mapped by the effect of its gravity[ref]. Another 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.

Evidence for this story comes in the form of predictions for the relative abundances of the lightest elements (which were forged in the first few minutes of the universe) and in correlations in the ΛCDM which can be directly linked to the transition between a universe so densely packed with radiation that it could support acoustic pressure waves, and a universe that had dropped below that density. Future work on ΛCDM includes the solution of the Hubble tension—how old is the universe really? All in all, it is incredible only a hundred years ago people were debating whether the universe had a beginning or not (and the “great debate”—were there other galaxies?) to now, where the debate rages over the second and third significant figures.

Before focusing our whole attention on the past, let us briefly look toward the future of the universe. Eventually a will be sufficiently large that Ω_Λ is the dominant contribution to the energy budget of the universe. Then

$$H^2 = H_0^2 \Omega_\Lambda \tag{1.15}$$

$$\implies \dot{a} = \pm H_0 \sqrt{\Omega_\Lambda} a \tag{1.16}$$

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

$$a(t) = a_0 e^{H(t-t_0)} \quad (1.17)$$

where H is a constant, having frozen in to a factor $\sqrt{\Omega_\Lambda}$ smaller than its value today.

One question we can ask is how far light will travel in the lifetime of the universe. Since $ds^2 = 0$ for a photon, if we define the conformal time τ such that

$$\frac{d\tau}{dt} = \frac{1}{a}, \quad (1.18)$$

we find that for the photon, $d\chi = d\tau$. This means that on a plot of χ vs. τ , the path of a photon will be a line of slope 1, even in an expanding universe. Integrating, we then find that

$$\tau(t) = \chi(t) = (Ha_0)^{-1} (1 - e^{-H(t-t_0)}) + \tau_0 \quad (1.19)$$

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, $\int ad\chi$) diverges. Integrating (1.11) in the same way that generated figure 1.5 we find that the points that we see in today's CMB are a comoving distance of 14.2Gpc away. We also find the asymptotic value of τ to be 19.2Gpc, indicating that we will never see the light that was emitted from the CMB at comoving distances greater than this. 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.

1.2 Initial conditions for ΛCDM

1.2.1 Motivations for inflation

The ΛCDM model does not posit a “Big Bang singularity”—instead, as we have discussed, it posits an early epoch in which the components that we have discussed are in thermal equilibrium, at an incredibly hot temperature. However, if one assumes that those components are all there are, and runs the equations backwards, one reaches $a = 0$ in finite conformal time.

These components are distributed incredibly uniformly. This simple scenario then evolves under gravitational collapse (and other interactions) and forms the

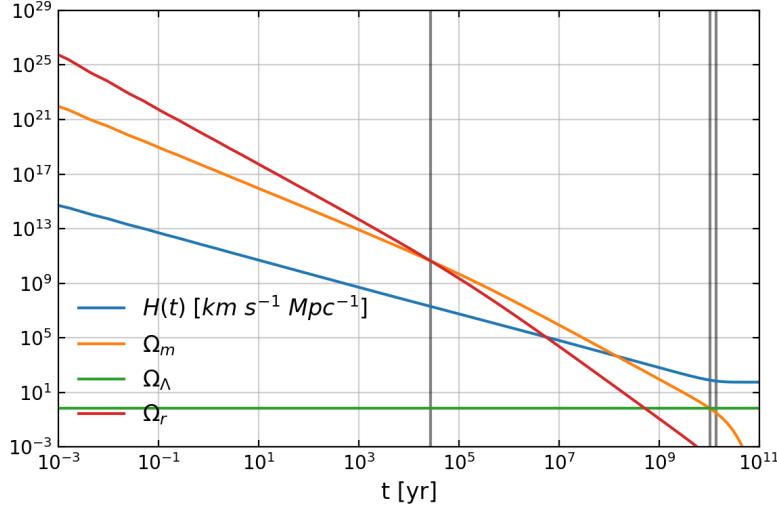


Figure 1.1: The evolution of the components of the universe up to the present, and slightly beyond. The vertical grey lines mark matter-radiation equality, matter-dark energy equality, and the present day, respectively. We measure these quantities at the present, and extrapolate into the past and future using the Friedmann equation (1.11). 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 dark energy comes to dominate, $H(t)$ will become constant.

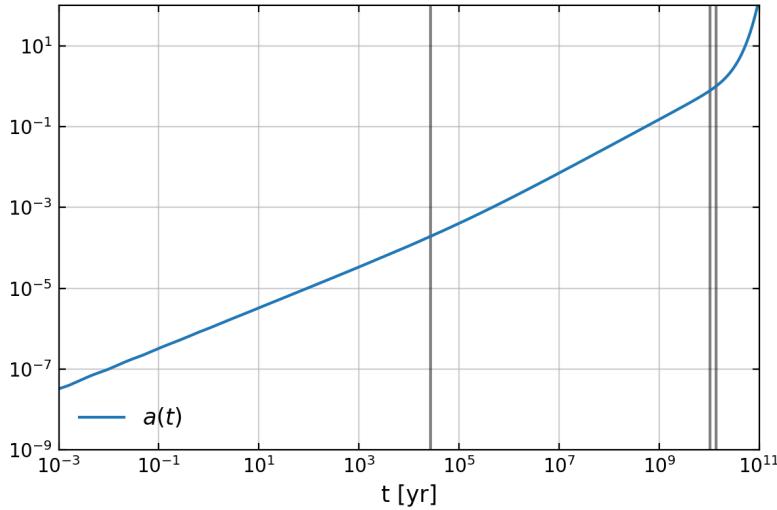


Figure 1.2: The evolution of the scale factor. For most of the ΛCDM history it evolves as some power of t (see table 1.1), however as Λ comes to dominate it will begin to grow exponentially.



Figure 1.3: During the radiation and matter dominated eras, the evolution of $a(t)$ has been slowing—this is decelerating expansion. However, as Λ comes to dominate, the \dot{a} will begin to increase, and the universe will enter an epoch of accelerated expansion.

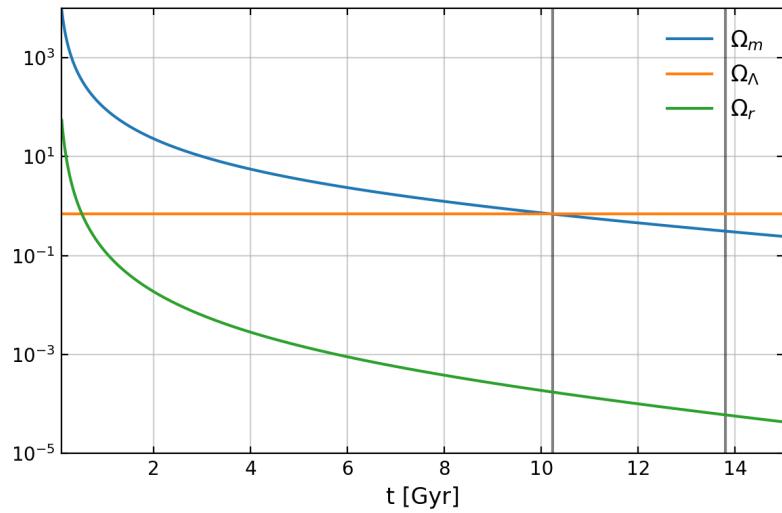


Figure 1.4: The evolution of the components of the universe, zoomed in to more clearly show the matter-dark energy transition. The first vertical grey line is matter-dark energy equality, the second is the present day.



Figure 1.5: The evolution of the conformal time τ since the ΛCDM singularity, if there is no inflation. We see that without inflation there is finite conformal time in the past and that due to eventual Λ dominance eventually τ will asymptote to a constant, denoted here by the horizontal grey line at $\tau = \tau_0 + (H_0 a_0)^{-1}$. Thus, without inflation, there is a maximum scale on which we would expect the CMB to be correlated, and a maximum comoving distance that we can ever expect to receive CMB photons from.

universe we know.

Is this story of ΛCDM enough to explain all the features of the universe that we observe? The answer is mostly yes, given the correct initial conditions, the correct starting point. But then that begs the question of how those initial conditions were chosen out of the other possibilities one could imagine. Two clues that we have are known as the horizon problem, and the flatness problem. Roughly speaking, the first 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 part of the sky so close to the temperature on the other side of the sky? The second is the problem that the curvature of the universe appears to be so small, despite the fact that there is no reason to believe it should be set to precisely zero, and one would expect the dynamics to increase its relative contribution to the energy density of the universe.

To fit observations, ΛCDM requires initial conditions with a particular set of properties. Some of these properties are surprising, and require an explanation. Since $a(\tau) \propto \tau$ during the radiation dominated era we can calculate 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

$$\chi_{rec} = \int_0^{\tau_{rec}} d\tau = \int_0^{a_{rec}} \frac{d\tau}{da} da \propto a_{rec}. \quad (1.20)$$

We can see that this is finite. Including the prefactor we can find that $\chi_{rec} = 280\text{Mpc}$. We can also calculate that the comoving distance that a photon could have freely streamed through the universe since last scattering is $\chi_0 = 14000\text{Mpc}$. This means that we would expect the homogeneous patches at recombination to span at most an angle of

$$\tan^{-1} \left(2 \frac{\chi_{rec} - 0}{\chi_0 - \chi_{rec}} \right) \approx 2^\circ \quad (1.21)$$

on the CMB sky today. We can also calculate the number of disconnected patches we should see as

$$\frac{4\pi(\chi_0 - \chi_{rec})^2}{\pi(\chi_{rec} - 0)^2} \approx 10^5. \quad (1.22)$$

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 to shrink—this would give large scales enough time to come into thermal equilibrium, then leave them outside 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 a cosmological constant—in that case the horizon $1/(aH)$ shrinks exponentially, and we find that $\tau = -1/(aH)$, so in the limit $a \rightarrow 0$ we see there is infinite conformal time in the past, to allow the entire observable universe to come into causal contact.

The flatness problem is the statement that our universe is much flatter than one would expect. In (1.1), the possibility exists that dS_3^2 is not the simple flat form we see observationally. One can write down this curvature as a contribution on the right hand side of (1.11), taking the form $\Omega_K = \frac{-K}{a^2 H_0^2}$. One then observes that this relative contribution would be expected to increase with time, as it drops off more slowly than either matter or radiation. **But now $\sum \Omega_i \neq 1$ check this!** This means that if the universe is very close to flat now (as we measure it

to be) then it must have been surprisingly close to flat at the beginning of the ΛCDM evolution. A natural explanation of this would be an attractive quality in a theory of the very early universe.

We will mention three more features of the ΛCDM initial conditions that we would like an explanation for. Those three features are the adiabaticity of the initial perturbations, the (as measured so far) Gaussianity of those perturbations, and finally the very slight deviation from perfect scale invariance that observations demand the initial power spectrum must have had.

Adiabaticity is the statement that for each of the components of the universe, their initial conditions $\delta\rho_i$ were related such that

$$\frac{\delta\rho_i}{\dot{\rho}_i} = \frac{\delta\rho_j}{\dot{\rho}_j}. \quad (1.23)$$

To phrase this another way, by performing a local reparametrisation of time $t \rightarrow t + \delta t(x)$

$$\bar{\rho}_i(t) + \delta\rho_i(t, x) = \bar{\rho}_i(t) + \delta t(x)\dot{\bar{\rho}}_i = \bar{\rho}_i(t + \delta t(x)) \quad (1.24)$$

we can absorb the perturbations in *all* of the components. The implication that we can draw is that the initial conditions of these separate components were generated by a process that was simpler than it could have been.

The Gaussianity of the initial conditions is the statement that their statistics are completely described by their two-point function. We will discuss this in more detail in a later section. The deviation from perfect scale invariance has been measured to high confidence by the *Planck* satellite. Earlier measurements were made by the WMAP satellite [10]. Later in section 1.2.4 we will 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 this is not precisely respected.

1.2.2 Driving inflation with a scalar field

REORDER WHOLE SECTION. We will work with an inflaton action of the form

$$S = \int d^4x \sqrt{-g} P(X, \phi) \quad (1.25)$$

with $X = -\frac{1}{2}g^{ab}\nabla_a\phi\nabla_b\phi$. We work with the number of e-folds, N , as our time variable $x' = \frac{dx}{dN} = a\frac{dx}{da}$ and recall the definitions of the slow-roll parameters (1.39), though we make no assumption that these are actually small. c_s is the sound speed of the theory, which can vary with time [11]

$$c_s = \frac{P_{,X}}{P_{,X} + 2XP_{,XX}}. \quad (1.26)$$

The background quantities are evolved according to the Friedmann equations, which are set with consistent initial conditions. The expression for the energy density is [11]

$$\rho = 2XP_{,X} - P, \quad (1.27)$$

so then

$$\rho' = -6XP_{,X}. \quad (1.28)$$

We can also see that requiring $\rho_\phi = -p_\phi$ demands that $2XP_{,X} \ll P$.

We take $X = \frac{1}{2}(\frac{\partial\phi}{\partial t})^2$. The sound speed for the adiabatic perturbation is defined as

$$c_s^2 = \left. \frac{\partial P}{\partial \rho} \right|_S = \frac{\dot{P}_0}{\dot{\rho}_0} = \frac{P_{,X}}{\rho_{,X}}. \quad (1.29)$$

Since we also have $\rho = 2XP_{,X} - P$, we see that

$$c_s^2 = \frac{P_{,X}}{P_X + 2XP_{,XX}}. \quad (1.30)$$

For canonical inflation models, $P(X, \phi) = X - V(\phi)$ so $P_{XX} = 0$ and $c_s = 1$.

Then, using the Friedmann equation we obtain

$$\varepsilon = -\frac{1}{2}\phi'^2P_{,X}. \quad (1.31)$$

The equation of motion for ϕ is [12]

$$\phi'' + (3c_s^2 - \varepsilon)\phi' + H^{-2}\frac{\rho_\phi}{\rho_X} = 0. \quad (1.32)$$

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 (1.18):

$$\tau'_s = \frac{c_s}{aH}, \quad (1.33)$$

$$(1.34)$$

We desire a period of exponential expansion in the early universe to explain the scale-invariance of the initial perturbations. One simple way one could imagine driving this expansion is through a single scalar field. This scalar field would have

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad (1.35)$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi). \quad (1.36)$$

For successful inflation with a cosmological constant, we want $w = -1$, i.e. that $V(\phi) \gg \frac{1}{2}\dot{\phi}^2$. Since the kinetic term is required to be small, this is referred to as *slow-roll* inflation. The Friedmann equations then become

$$H^2 \approx \frac{8\pi G}{3}V(\phi), \quad (1.37)$$

$$\frac{\ddot{a}}{a} \approx -\frac{4\pi G}{3}(-2V(\phi)). \quad (1.38)$$

We define the Hubble parameter and the standard “slow-roll” parameters:

$$\begin{aligned} H &= \frac{d \ln a}{dt}, & \epsilon &= -\frac{d \ln H}{dN} \\ \eta &= \frac{d \ln \varepsilon}{dN}, & \epsilon_s &= +\frac{d \ln c_s}{dN}. \end{aligned} \quad (1.39)$$

From the continuity equation and the Friedmann equation

$$\rho' = -3(\rho + P) \quad (1.40)$$

$$\implies 6HH' = -3(2XP_{,X}) \quad (1.41)$$

$$\implies -H^2\varepsilon = -\left(\frac{1}{2}H^2\phi'^2\right)P_{,X} \quad (1.42)$$

$$\implies \varepsilon = \frac{1}{2}\phi'^2P_{,X}. \quad (1.43)$$

In the canonical case this simplifies to $\varepsilon = \frac{1}{2}\phi'^2$. In the DBI case we find $\varepsilon = \frac{1}{2}\frac{\phi'^2}{c_s}$.

This allows us to rephrase the Friedmann equation in the following useful

way. Note that this is still exact, the slow-roll approximation has not been used.

$$H^2 = \frac{1}{3} \left(\frac{\varepsilon H^2}{P_{,X}} + V(\phi) \right) \quad (1.44)$$

$$\implies H^2 = \frac{V(\phi)}{3 - \frac{\varepsilon}{P_{,X}}}. \quad (1.45)$$

1.2.3 Criteria for successful inflation

Rewrite entirely in terms of a , H ? And move earlier. 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

$$\frac{d}{dt} (aH)^{-1} = -\frac{\ddot{a}}{\dot{a}^2} < 0 \quad (1.46)$$

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

$$\frac{d}{dt} (aH)^{-1} = -\frac{1}{(aH)^2} (\dot{a}H + a\dot{H}) \quad (1.47)$$

$$= -\frac{1}{a} (1 - \varepsilon) \quad (1.48)$$

and so we need $\varepsilon < 1$ to successfully inflate the universe. This is necessary but not sufficient—we still need to generate the correct statistics for the primordial perturbations. This implies we must be close to a perfect de Sitter phase, i.e. $\varepsilon \ll 1$. To match the measured deviation from scale-invariance in standard slow-roll inflation we would need $\varepsilon \sim O(10^{-2})$.

For a model with a canonical kinetic term, (1.32) simplifies to

$$\phi'' + (3 - \varepsilon)\phi' + H^{-2}V'(\phi) = 0. \quad (1.49)$$

In the slow-roll approximation we take assume $\phi'' \ll \phi'$, and so

$$\phi' \approx \frac{V'(\phi)}{3H^2} \approx \frac{V'(\phi)}{V(\phi)}. \quad (1.50)$$

From this, we see that demanding that ε be small places a constraint on the flatness of the potential in a canonical model.

We also require $\eta \ll 1$ to maintain inflation for the sufficient number of e-folds.

1.2.4 Correlation functions

If you find yourself in a universe with... Expectation value of an ensemble. Monte-Carlo simulations are used to estimate covariance matrices $\langle a_{lm}a_{l'm'} \rangle$, which are not diagonal in the presence of a sky mask. Classical vs. quantum correlators (order matters), ensemble vs. spatial averages.

Cite [13, 14]. The power spectrum and other correlations are predicted by inflation. Define n-point correlations, their Fourier transforms, talk about them as observables.

A model of inflation will predict the statistical properties of the distribution of matter that forms the initial conditions for the ΛCDM evolution of our universe. From this, we can calculate the statistical properties of the CMB sky that we see. One such property is the two-point correlation of the temperature ϕ at a point x in space, $\langle \phi(x)\phi(x') \rangle$. When there is a large separation between x and x' **talk about cosmic variance of different modes. Angular?**

Scale invariance is the statement that

$$\langle f(\mathbf{x})f(\mathbf{x}') \rangle = \langle f(\lambda\mathbf{x})f(\lambda\mathbf{x}') \rangle. \quad (1.51)$$

From this, we can see that scale invariance needs $P(k) \propto k^{-3}$. In fact, it was found that $P(k) \propto k^{n_s-4}$, with $n_s = 0.9649 \pm 0.0042$.

The ensemble average (1.57) of a product of two fields is known as the two-point correlator, or the two-point function $\langle f(\mathbf{x})f(\mathbf{x}') \rangle$. There are two properties we wish to immediately impose on this function, and higher order correlators (products of more fields). Those conditions are statistical homogeneity and statistical isotropy. That is, we demand that the correlators are invariant under translations and rotations—that is, while some realisation of the ensemble might have special points (e.g. a maximum value), when averaged over the entire ensemble these special points are equally likely to appear anywhere in the space.

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

$$\langle f(\mathbf{k}_1)f(\mathbf{k}_2) \rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2)P(k_1), \quad (1.52)$$

$$\langle f(\mathbf{k}_1)f(\mathbf{k}_2)f(\mathbf{k}_3) \rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)B(k_1, k_2, k_3), \quad (1.53)$$

where the delta functions come from statistical homogeneity, and the fact that $P(k)$ (the power spectrum) 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 degrees of freedom, but $B(k_1, k_2, k_3)$ has only three. This is due to the delta function in demanding the \mathbf{k}_i sum to a triangle, and since the orientation of this triangle cannot matter, three parameters will suffice—we choose the magnitudes.

We will want to measure these correlations. How can we do that with only

access to one realisation of the distribution? It is useful to think about the ergodic theorem. Since an integral over space can be identified with the ensemble average, we can measure, say, the average temperature across space, which should give us an estimate for the average temperature of the ensemble (the quantity that will actually be predicted from theory). **Is this related to large cosmic variance at large scales? See pg 94 of Lyth and Liddle.**

1.3 Statistical observables

1.3.1 Checking if dice are fair

Move to after defining correlation functions? Organise this!!! Correlation functions are expectation values over ensembles of possible universes. 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? The answer, thankfully, is no.

One question we can ask is, if we magically had access to the entirety of one spatially infinite universe then would our measured averages have a cosmic variance, or would those spatial averages precisely match the ensemble averages? If the distribution obeys the assumptions of the ergodic theorem (see for example appendix D of [15]) they would match—despite the fact we only had one universe, we would still be able to beat cosmic variance.

Inflation predicts the statistics of an ensemble of universes. If we only have a finite amount of data from this ensemble, we cannot perfectly measure some given average, as we will be limited by cosmic variance. There are two ways to beat cosmic variance (to measure the ensemble average $\langle \delta_m(x)^2 \rangle$ at some fixed 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.

If you make the assumption that your universe was drawn from a Gaussian distribution, you could use Gaussian Monte-Carlo simulations to, for example, estimate your rate of false positives for non-Gaussianity. The assumed Gaussian distribution (values of C_l) is used to generate a_{lm} draws **check this**. But, of course, they will not allow you to get from observations to the true distribution.

Cosmic variance in our measurements is due to both being limited to one

universe, and to being limited to a finite spatial region within that universe. In practice, we have some other advantages. For example, the distribution from which the Fourier coefficients are drawn depends only on $|\mathbf{k}|$, so each new direction we measure will give us new information. So, even in the same spatial region in the one universe, we have access to multiple independent samples of the distribution. **Organise this!!!**

Everything we have discussed until this point has dealt with quantities with definite values, that have a definite evolution. 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.

Our result will be a constraint on an inflationary parameter, stated at e.g. 95% confidence. For example, if we flipped one fair coin 100 times, we would expect an even split of heads and tails. Should we be suspicious of the fairness of the coin if we instead get 60 heads and 40 tails? Or 80 heads and 20 tails? An event at least as extreme at a 60-40 split will occur 3.5% of the time. For a 80-20 split however, the probability of an event that extreme is around one part in 10^{10} , effectively impossible. This means that if we observed a trial where a coin came up heads in 80 out of 100 cases, we could suspect that our hypothesis of fairness was incorrect. One can calculate that, for a fair coin, 95% of the time the number of heads will take a value between 41 and 59. We call this the 95% confidence interval. What we wish to do is the opposite, however. Instead of calculating using known probabilities, we wish to take a series of observations of some random variable and use these observations to estimate the probability distribution that they were drawn from. Given an infinite number of observations this task has a straightforward solution, simply plotting the normalised histogram of the results. However, in the context of cosmology we will only have a finite amount of information available to us, and thus a limit to how well we can ever measure the relevant probability distributions. This limitation is known as cosmic variance.

We quantify this expected scatter around the mean using a quantity known as the standard deviation. The standard deviation of a random variable X is the square root of the expected value of the squared deviation from the mean μ ,

$$\sigma = \sqrt{E[(X - \mu)^2]}. \quad (1.54)$$

In our example above of flipping one coin 100 times, the total number of heads has $\mu = 50$ and $\sigma = 5$.

We have only one universe that we can observe, only one draw from the probability distribution we are trying to probe. However, there is a lot of information in that one draw. **Motivate and explain this all better!** This is related to the concept of ergodicity. This is the statement that $\langle \cdot \rangle$ as an expectation over ensembles at a fixed point is the same as the expectation over points for a fixed ensemble. This assumes homogeneity, stationarity, and that distant points are uncorrelated. In terms of our coin example, this is related to the statement that flipping one coin 100 times should probe the probability distribution in the same way as flipping 100 identical coins once each.

We can define the expectation value of a function of a discrete or continuous random variable x , or a functional F of a field configuration $f(x)$, respectively as

$$\langle f(x) \rangle = \sum_i x_i P(x_i) \quad (1.55)$$

$$\langle f(x) \rangle = \int dx x \rho(x) \quad (1.56)$$

$$\langle F[f(x)] \rangle = \int \mathcal{D}f F[f(x)] P[f(x)] \quad (1.57)$$

where the sum and integral over x are over the values that x can take, and the functional integral over f is over all the field configurations that $f(x)$ can take.

In our example of the 100 coin flips, we used (1.55) to calculate the mean and standard deviation. For a continuous variable, like the average height of a population or the average temperature in a given room, we would use (1.56). In this thesis, we will be working with the expectation value of field configurations, so we will use (1.57). **Talk about quantum to classical transition!**

$$\langle \hat{v}_{\mathbf{k}} \hat{v}_{\mathbf{k}'} \rangle = \langle 0 | \hat{v}_{\mathbf{k}} \hat{v}_{\mathbf{k}'} | 0 \rangle \quad (1.58)$$

$$= |v_k|^2 \left\langle 0 \left| \left[\hat{a}_{\mathbf{k}}, \hat{a}_{-\mathbf{k}}^\dagger \right] \right| 0 \right\rangle \quad (1.59)$$

$$= |v_k|^2 \delta(\mathbf{k} + \mathbf{k}') \quad (1.60)$$

$$= P_v(k) \delta(\mathbf{k} + \mathbf{k}') \quad (1.61)$$

1.4 Thesis outline

Merge this, make coherent. We are now in a position to give a more precise description of the goals, methods and results outlined in this thesis. Our main goal was to develop an efficient numerical pipeline for connecting inflation models directly to observations through the CMB bispectrum. The concrete results we aimed for were constraints on the parameters of inflation models, not constraints on phenomenological templates or summary f_{NL} parameters. We aimed to obtain the full bispectrum shape information from an inflation scenario, not point samples or a limit. The novelty of our goals comes from the separable and numerical methods granting access to more accurate, and in some cases 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. This will link to a CMB calculation that will be presented in [2]. That calculation is 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 developed methods to achieve this.

Our main result is the first development of a formalism for calculating the separable expansion we describe to sufficiently high orders as to describe inflationary features. These methods are implemented in the PRIMODAL code. Another result is the recognition and description of the central issue of the convergence of the expansion, and in particular how it is affected by the non-physical k -configurations. The result of our work in circumventing this problem is the development of multiple sets of basis functions, which efficiently capture a broad range of bispectra. This allowed the first validation of our separable methods on features. We also note that as a side-effect of this desire for a separable primordial bispectrum, the primordial calculation of the inflationary bispectrum that we present is much more efficient than previous methods, in the sense converges far faster in modes than previous methods would in point samples.

Our final highlighted result, which we obtained in collaboration, is the connection of our inflationary scenario to the CMB through the CMB-BEST code developed by Wuhyun Sohn. This allowed us to obtain constraint on the sound speed of DBI inflation, c_s .

The thesis is organised as follows. In chapter 2 we present brief reviews of the various parts of the pipeline that connects inflation scenarios to observations

through the bispectrum. We review the usual paradigm of bispectrum estimation in the CMB, and the motivation for separable bispectra. We review the in-in formalism, for calculating the tree level bispectrum for a given model of inflation. We review $P(X, \phi)$ models of inflation as an example, and some of the usual approximate bispectrum templates that we aim to bypass. We will draw our validation scenarios from these models. We discuss previous numerical codes for calculating the primordial bispectrum k -configuration by k -configuration, which contrasts our separable basis expansion. We review the previous work in achieving separability through modal expansions in [16], and we discuss methods of testing numerical bispectrum results, defining our relative difference measurement.

We begin chapter 3 by recasting the usual in-in calculation into an explicitly separable form, in terms of an expansion in an arbitrary basis. Since the paradigm we aim to present is only viable if we can find a basis that can efficiently represent a wide variety of bispectra, we devote the rest of chapter 3 to presenting this discussion, which is distinct and separate from our numerical methods but nonetheless vital. We discuss the effects of the non-physical k -configurations on the convergence of our expansion on the tetrapyd, and present multiple efficient basis sets. In chapter 4 we present the precise methods involved in the separable formalism and its numerical implementation, and detail our methods for carefully calculating the coefficients to high order. In particular in section 4.7 we validate our methods and implementation on inflation scenarios with varied features from the literature. We present a constraint on an inflation model in chapter 5, and finish with a discussion of future work in chapter 6.

The specific results involved can be categorised into two main lines of research. The first line of research is in determining the feasibility of the method. Our initial result here was identifying the contributions of the non-physical configurations as a novel problem to this analysis, 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. We find basis sets that 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 we follow is in developing 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 we use to overcome the difficulties encountered. These difficulties include accurately including the highly oscillatory

early-time contributions. We present a careful and comprehensive validation of our methods, showing that we can overcome the convergence problems and capture the oscillatory contributions efficiently. At the primordial level we do this by validating our results using three distinct tests. We validate on established templates with non-trivial phenomenology; we use the squeezed-limit consistency condition; and we compare our full bispectrum results to previous codes through point-tests.

At the level of the CMB, we validate the convergence of our basis by ensuring we can reproduce the *Planck* constraints (using the *Planck* data) on the DBI sound speed, and we translate this constraint into a constraint on a fundamental parameter in the context of a physically realistic scan.

Chapter 1 is a general introduction to cosmology, and chapter 2 is an introduction to the bispectrum as an observable.

Chapter 3 describes the first line of research mentioned above—we discuss why we need to take the non-physical configurations into account, how we build our basis sets, and present quantitative comparisons of the convergence of each basis set to relevant examples of bispectrum shapes.

In chapter 4 we detail the second line of research mentioned above—we discuss the details of recasting the in-in formalism in an explicitly separable form, making explicit each step of the calculation. We discuss how to efficiently deal with the early time contributions to the integrals, and other numerical issues that require considerable care and attention. We also present validation tests on a very broad range of types of non-Gaussianity.

In chapter 5 we present constraints (obtained in collaboration with Wuhyun Sohn) on the sound speed of the DBI model, which validate our pipeline against *Planck*. This chapter will describe the parametrisation of the scan, and the physical interpretation of the results.

In chapter 6 we discuss possible avenues of future work, and present our conclusions.

Chapter 2

Introduction II: Probing Inflation

Put this somewhere: In this chapter we will review details of the bispectrum and how it is obtained.

The standard cosmological history, the Λ *CDM* model, explains the structure we see in the distribution of matter (large-scale structure, LSS) and radiation (the cosmic microwave background, *CMB*) in our universe. However, this model still requires initial conditions. These initial conditions – adiabatic perturbations on a uniform background, close to Gaussian, with a power spectrum that is nearly (but not exactly) scale-invariant – are provided by an epoch of “inflation”, an epoch before the Big Bang of the Λ *CDM* model.

During inflation, the universe is supposed to have expanded rapidly, stretching quantum fluctuations from small scales to large. This seeded the formation of structure and encoded correlations that we would eventually measure today. Much information on fundamental physics has been gleaned from the two-point correlators of observations. For higher order correlators (which probe the non-linearities of interactions in the early universe) calculating the predictions of a model of inflation and comparing them to what we see in the sky is much more computationally intensive. Previous work, including that done by the Planck collaboration [17], has managed to constrain some models of inflation through their three-point correlations (or rather through their proxy, the bispectrum). The computational complexity was dealt with by making layers of approximations and the space of models which could be constrained was limited. One notable simplifying assumption is that of “separability”, which makes the process of *CMB* estimation much more tractable [18].

The formalism used to calculate inflationary non-Gaussianities within an inflation scenario is known as the in-in formalism [19]…

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 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 [20], 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. Ways this can happen include self-interactions [7, 21], interactions between multiple fields [22–34], sharp features [35–37] and periodic features [38–47]. The link between primordial non-Gaussianity and primordial black holes has also been studied [48–51]. Ultra slow-roll is a regime that has been explored [52–55] **Check those references!** However, constraining such imprints is extremely difficult observationally. Even once the data has been obtained, using existing methods it is extremely computationally intensive to translate this into constraints on specific inflation scenarios. Much progress has been made by coarse-graining the model space into a small number of approximate templates, and leveraging the simplifying characteristic of separability with respect to the three parameters of the bispectrum [18, 56].

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 all primordial templates estimated thus far from the CMB are consistent with zero has already provided such constraints in certain scenarios [17, 57]. With this high-precision *Planck* data, and data from forthcoming experiments such as the Simons Observatory (SO) [58] and CMB-S4 [59], robust pipelines must be developed to circumvent the computational difficulties and extract the maximum amount of information possible. Due to the nature of bispectrum estimation in the CMB and LSS [60–63] constraining an arbitrary template is difficult. 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 [64–66] 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 cannot. In this work we exploit the intrinsic separability of the tree-level in-in formalism to apply these methods at the level of inflation. Expressing the primordial bispectrum in a separable basis expansion leads to vast increases in efficiency both at the primordial and late-universe parts of the calculation. The main advantage is that expressing the primordial shape function in this way reduces the process of bispectrum estimation in the CMB to a cost which is large, but need only be paid once per basis, not per scenario. A proof of concept of this approach at the primordial level was presented in [16], and the details of the bispectrum estimation part will be detailed in [2]. We go beyond the work of [16] 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 modal expansion, allowing us to apply the method to feature bispectra for the first time.

2.1 Perturbations about the background

We can get a lot of information on the evolution of the perturbations from observations, and that tells us about the history of universe, which tells us about its content, etc... We now turn our attention to the details of the perturbations.

The sound speed c_s (1.29) is the propagation speed of the perturbations. This defines a sound horizon, c_s/H —as the modes cross this sound horizon they freeze.

The action for the perturbations is derived by working in a general gauge (using the ADM formalism) and splitting the relevant quantities into a background part and a small perturbation

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

In the uniform density gauge we have (neglecting tensor modes)

$$\delta\phi = 0, \quad h_{ij} = a^2 e^{2\zeta} \delta_{ij} \quad (2.2)$$

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

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

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

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

where $z^2 = 2a^2\varepsilon/c_s^2 = a^2(\rho + P)/(c_s H)^2$. 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. Expanding the mass, we find

$$\frac{1}{z} \frac{d^2 z}{d\tau^2} \approx 2(aH)^2 \quad (2.5)$$

where we have neglected terms that are slow-roll suppressed. The evolution of the perturbations is therefore well approximated using $\frac{1}{z} \frac{d^2 z}{d\tau^2} = 2(aH)^2$ unless there are features that make the slow-roll parameters large.

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

$$\zeta_k'' + (3 - \varepsilon + \eta - 2\varepsilon_s)\zeta_k' + \left(\frac{kc_s}{aH} \right)^2 \zeta_k = 0. \quad (2.6)$$

We can see that at late times, when $kc_s \ll aH$, ζ_k will freeze. This conservation of ζ outside the horizon [67] is an important property¹. This 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.

¹Systems with multiple degrees of freedom have ζ evolving even outside the horizon. See [11] for a discussion on how this relates to non-adiabatic pressure in inflation models with non-canonical kinetic terms.

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

$$\frac{\partial^2 v_k}{\partial \tau^2} + \left(c_s^2 k^2 - \frac{1}{z} \frac{d^2 z}{d\tau^2} \right) v_k = 0 \quad (2.7)$$

however, this is not ideal as c_s depends on time. Instead, let us use τ_s as our time parameter. If we follow [12] and define

$$y_k = \sqrt{2k c_s} v_k, \quad (2.8)$$

we obtain

$$\frac{\partial^2 y_k}{\partial \tau_s^2} + \left(k^2 - 2 \left(\frac{aH}{c_s} \right)^2 (1 + \dots) \right) y_k \approx 0 \quad (2.9)$$

where we have neglected 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

$$\frac{\partial^2 y_k}{\partial \tau_s^2} + k^2 y_k \approx 0 \quad (2.10)$$

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

2.1.2 Early time behaviour

2.1.2.1 Quantizing the perturbations

We now wish to quantize the perturbations. We will quantize v , as at early times it should behave like a free field in Minkowski space. We expand

$$v(\tau, \mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \left[v_k(\tau) \hat{a}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} + v_k^*(\tau) \hat{a}_{\mathbf{k}}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}} \right] \quad (2.11)$$

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

where $v_k(\tau)$ and $v_k^*(\tau)$ satisfy (2.7). The creation and annihilation operators satisfy

$$\left[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger \right] = (2\pi)^3 \delta^{(3)}(\mathbf{k} - \mathbf{k}'). \quad (2.13)$$

Taking the commutator of v and its conjugate momentum $\pi = \frac{dv}{d\tau}$ **CHECK ALL OF THIS** we can obtain

$$[v(\tau, \mathbf{x}), \pi(\tau, \mathbf{y})] = \int \frac{d^3\mathbf{k}}{(2\pi)^3} (v_k \pi_k^* - \pi_k v_k^*) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \quad (2.14)$$

$$= i\delta^{(3)}(\mathbf{x} - \mathbf{y}) \quad (2.15)$$

if we demand that $v_k \pi_k^* - \pi_k v_k^* = 1$.

We can define the vacuum state by $\hat{a}_k |0\rangle = 0$. Demanding this to be an eigenstate of H forces $\pi_k = \pm ik v_k$. **How does this relate to before?**

2.1.2.2 Initial conditions for the perturbations

Make ζ vs. u_k vs. v_k consistent throughout! The Bunch-Davies vacuum leads us to the mode functions

$$\zeta_k(\tau) = \frac{H}{\sqrt{4\varepsilon c_s k^3}} (1 - ik c_s \tau) e^{ic_s k \tau}. \quad (2.16)$$

The power spectrum is enhanced by a factor of c_s ,

$$P(k) = \frac{H^2}{4\varepsilon c_s k^3} \quad (2.17)$$

due to the larger prefactor.

The vacuum is not unique. We choose it so that it asymptotes to the Minkowski vacuum in the limit $\tau \rightarrow -\infty$.

Demanding that the early time behaviour matches that of flat space gives us that

$$v_k = \frac{1}{\sqrt{2k}} e^{-ik\tau_s} \quad (2.18)$$

when $\tau_s \rightarrow -\infty$. **Is this really τ_s or τ ?** Assuming **assumptions** we can write the solution to the equations of motion in terms of the Hankel function $H_i^{(2)}$ as [7]

$$\zeta_k = \frac{1}{2a} \sqrt{\frac{\pi}{2}} \sqrt{\frac{-\tau_s}{zc_s}} H_\nu^{(2)}[-k\tau_s]. \quad (2.19)$$

where $\nu = \frac{3}{2} + \varepsilon + \frac{\eta}{2} + \frac{\varepsilon_s}{2}$ **Rewrite this in terms of τ_s . Citation?**

2.1.3 Late time behaviour

At late times we can use the limit

$$\lim_{x \rightarrow 0} H_\nu^{(1)}(x) \rightarrow -\frac{i}{\pi} \Gamma(\nu) \left(\frac{x}{2}\right)^{-\nu}. \quad (2.20)$$

From this limiting behaviour we can calculate the power spectrum predicted by a $P(X, \phi)$ model, as a function of time and k . What we find **EXPAND!!** is that the time dependence of the prefactor $H^2/(4zc_s^3)$ is precisely canceled out by the time dependence of the limiting behaviour of the Hankel function, as expected. This means we can obtain the final power spectrum by evaluating it separately for each k . In fact, we choose to evaluate the asymptotic expression at sound-horizon crossing $c_s k = aH$. Note that we are *not* evaluating $\zeta_k \zeta_k^*$ at horizon crossing—we are evaluating the asymptotic expression for that quantity there.

The equation of motion for the perturbations is (2.6), where $c_s = 1$ for standard canonical inflation. We use standard Bunch-Davies initial conditions, which leads us to impose the following condition deep in the horizon:

$$\zeta_k = \frac{i}{a} \sqrt{\frac{c_s}{4\varepsilon k}} e^{-ik\tau_s} \quad (2.21)$$

where τ_s is defined in (1.33). The solution in slow-roll (without features) is then approximately

$$\zeta_k \propto (1 + ik\tau_s) e^{-ik\tau_s}. \quad (2.22)$$

At leading order in slow-roll, using definition (1.52), the power spectrum is [68, 69]

$$P_\zeta(k) = \frac{H^2}{4\varepsilon c_s k^3}, \quad (2.23)$$

where the right hand side is evaluated at $c_s k = aH$. We also define a dimensionless version

$$\mathcal{P}_\zeta(k) = \frac{k^3}{2\pi^2} P_\zeta(k) \quad (2.24)$$

which is typically parameterised as

$$\mathcal{P}_\zeta(k) = A_s \left(\frac{k}{k_*} \right)^{n_s - 1}. \quad (2.25)$$

The spectral index can be found to be

$$n_s - 1 = -2\varepsilon - \eta - \varepsilon_s \quad (2.26)$$

at leading order in slow-roll.

Initially we consider the same basic models as in [16]; a quadratic potential

$$V_{\phi^2}(\phi) = \frac{1}{2}m^2\phi^2. \quad (2.27)$$

with a canonical kinetic term, and a non-canonical model, the DBI model described in section 2.5.2.

2.2 The primordial power spectrum

Define the CMB power spectrum, discuss constraints. Note we neglect ISW throughout. Mention polarisation, B-modes [70].

How does this rule out e.g. $m^2\phi^2$?

In general when projecting a three dimensional random field onto a sphere (the CMB sky) the statistics are projected as

$$\langle f_{lm} f_{l'm'}^* \rangle = 4\pi \delta_{ll'} \delta_{mm'} \int d\ln k \mathcal{P}_{\mathcal{F}}(k) j_l^2(kr) \quad (2.28)$$

where $j_l(x)$ is a Bessel function, and the $\delta_{ll'}$ and $\delta_{mm'}$ are demanded by statistical isotropy. The Bessel function encodes the fact that a three dimensional wave of a given wavelength intersects the sphere at multiple different angular separations.

The temperature anisotropies of the CMB are decomposed in spherical harmonics as

$$\frac{\Delta T}{T}(\hat{\mathbf{n}}) = \sum_{lm} a_{lm} Y_{lm}(\hat{\mathbf{n}}) \quad (2.29)$$

and the power spectrum is

$$C_l = \sum_m a_{lm} a_{l-m}. \quad (2.30)$$

The variance is given by

$$\text{var}(\hat{C}_l) = \frac{2}{2l+1} C_l^2, \quad (2.31)$$

which reflects the fact that for the distribution of C_l , there are $2l+1$ independently draws that we can observe.

The CMB power spectrum C_l is calculated from the primordial power spectrum \mathcal{P}_ζ using a transfer function

$$C_l \approx 4\pi \int d \ln k \mathcal{P}_\zeta(k) \Delta_l(k). \quad (2.32)$$

For large scales, where there was not enough time for oscillations before recombination, there is no processing of the primordial statistics and the transfer function simply encodes the projection effects

$$C_l \approx 4\pi \int d \ln k \mathcal{P}_\zeta(k) [j_l(k\chi_*)/5]^2. \quad (2.33)$$

For large l the spherical Bessel functions have a sharp peak when $k\chi_* = l$, so for large l we see that C_l is mostly determined by $\mathcal{P}_\zeta(l/\chi_*)$. For $\mathcal{P}_\zeta(k) = A_s$, i.e. perfectly scale invariant, we can use that

$$\int_0^\infty d \ln x j_l^2(x) = \frac{1}{2l(l+1)} \quad (2.34)$$

to find that $l(l+1)C_l \propto \text{const}$, i.e. that on large scales (so small l) this quantity is independent of l .

For intermediate scales, the acoustic oscillations and the Doppler effect becomes relevant, and the transfer function additionally encodes this processing

$$C_l \approx 4\pi \int d \ln k \mathcal{P}_\zeta(k) \left[\frac{(\Theta_0 + \psi)(\eta_*, \mathbf{k})}{\mathcal{R}(\mathbf{k})} j_l(k\chi_*) - \frac{v_\gamma(\eta_*, \mathbf{k})}{\mathcal{R}(\mathbf{k})} j'_l(k\chi_*) \right]^2. \quad (2.35)$$

Derivative of j with respect to its argument.

For small scales, diffusion of photons escaping **pressure waves??** damps out the sharp peaks of the acoustic oscillations.

2.3 The primordial bispectrum

2.3.1 The shape function

Shape versus scale dependence. The primordial bispectrum is usually written as:

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_1, k_2, k_3) \quad (2.36)$$

where $\zeta_{\mathbf{k}}$ is a Fourier mode of the standard gauge invariant curvature perturbation. 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:

$$S(k_1, k_2, k_3) = (k_1 k_2 k_3)^2 B(k_1, k_2, k_3). \quad (2.37)$$

As we saw in 1.2.4 the bispectrum is only defined where the triangle condition

$$\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0, \quad (2.38)$$

is satisfied, which implies that the triangle inequality must hold

$$k_1 + k_2 \geq k_3 \text{ and cyclic perms.} \quad (2.39)$$

The space of configurations we are interested in is therefore reduced from the full cube $[k_{\min}, k_{\max}]^3$ to a tetravdron (illustrated in figure 4.3), the intersection of that cube with the tetrahedron that satisfies (2.39). This has important implications that we will explore in chapter 3.

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

$$B^F(k_1, k_2, k_3) = f_{NL}^F \times F(k_1, k_2, k_3) \quad (2.40)$$

where F contains the dependence on the k -configuration. This definition coincides with the definitions of f_{NL}^{local} , f_{NL}^{equil} and f_{NL}^{ortho} when F is (respectively) the local (see (2.49)), equilateral (see (2.52)) and orthogonal templates, as defined in [57, 71].

2.4 A field guide to the definitions of f_{NL}

In our template-free pipeline, we take f_{NL} as a coefficient of our expansion of the numerically calculated primordial shape. Thus, for us, we rule out an inflation scenario by ruling out $f_{NL} = 1$. There are, however, many competing definitions and conventions for f_{NL} in the literature, especially between the observations-side and the primordial-side. We list some below, for convenience.

For the local form, we have **Check prefactor!**

$$\zeta(\mathbf{x}) = \zeta_G(\mathbf{x}) - \frac{3}{5} f_{NL}^{local} (\zeta_G(\mathbf{x})^2 - \langle \zeta_G(\mathbf{x})^2 \rangle) \quad (2.41)$$

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 (2.49).

In [71], f_{NL} is defined individually for each template, in such a way that for some template X , in the equilateral limit $k_1 = k_2 = k_3$

$$B_\Phi^X(k, k, k) = \frac{6A_\phi^2 f_{NL}^X}{k^6}, \quad (2.42)$$

for example equation (5) in [71], or

$$B_\Phi^X(k, k, k) = \frac{6A_\phi^2 f_{NL}^X}{k^{8-2n_s}}, \quad (2.43)$$

as in equation (6) in [71]. The parameter A_ϕ comes from

$$P_\Phi(k) = \frac{A_\phi}{k^{4-n_s}} \approx \left(\frac{3}{5}\right)^2 \frac{H^2}{4\varepsilon c_s k^{4-n_s}} \quad (2.44)$$

where we used $\Phi = \frac{3}{5}\zeta$ and the slow-roll result for \mathcal{P}_ζ . In [21, 69, 72] 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 (2.89). For now, we briefly state the definition of f_{NL}^{DBI} , from [71]

$$B_\Phi^{DBI}(k, k, k) = \frac{6A_\phi^2 f_{NL}^{DBI}}{k^6}. \quad (2.45)$$

Comparing to (2.89), we find that

$$f_{NL}^{DBI} = -\frac{35}{108} (c_s^{-2} - 1). \quad (2.46)$$

In [7, 73] we see f_{NL} defined as the reduced bispectrum

$$f_{NL}(k_1, k_2, k_3) = \frac{5}{6} \frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2) + P(k_1)P(k_3) + P(k_2)P(k_3)}, \quad (2.47)$$

which clearly (except for the local case) depends on k_1 , k_2 and k_3 .

For completeness, we also note another possible point of notational confusion. In [7] the definition $s^{there} = d \ln c_s / dN$ is used, however we refer to this quantity (as we defined in (1.39)) as ε_s . In [12], 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 (1.33). In [74], 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	[7]	[12]	[74]
τ_s	(1.33)		s	s
ε_s	(1.39)	s		σ_1
z		ε/c_s^2		

Table 2.1: Notation summary and comparison. Check N/A's. What about $X = \frac{1}{2}\dot{\phi}^2$ vs. $X = \dot{\phi}^2$?

2.4.1 Templates

The momentum dependence of the various possible bispectrum shapes can be complex and it is therefore useful to define a standard notation in which to express this dependence. For example in [75, 76] the elementary symmetric polynomials are used to compactly represent the results in a way that is manifestly symmetric².

²This 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 [75].

In this work we will use the notation employed in [65]:

$$\begin{aligned} K_p &= \sum_{i=1,2,3} k_i^p, \\ K_{pq} &= \frac{1}{\Delta_{pq}} \sum_{i \neq j} k_i^p k_j^q, \\ K_{prs} &= \frac{1}{\Delta_{prs}} \sum_{i \neq j \neq l} k_i^p k_j^r k_l^s, \end{aligned} \quad (2.48)$$

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 **Shouldn't be dimensionless, check this everywhere.**
Note on n_s , that's wrong here too. Connect to (2.41)

$$S^{local}(k_1, k_2, k_3) = \frac{6}{5} A_\zeta^2 f_{NL}^{local} \frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_3 k_1} + \frac{k_3^2}{k_1 k_2} \quad (2.49)$$

$$= \frac{6}{5} f_{NL}^{local} \left(\mathcal{P}_\zeta(k_2) \mathcal{P}_\zeta(k_3) + \mathcal{P}_\zeta(k_3) \mathcal{P}_\zeta(k_1) + \mathcal{P}_\zeta(k_1) \mathcal{P}_\zeta(k_2) \right) \quad (2.50)$$

$$= \frac{6}{5} A_\zeta^2 f_{NL}^{local} \frac{K_3}{6 K_{111}}. \quad (2.51)$$

This template peaks on squeezed triangles, i.e. when $k_1 \sim k_2 \gg k_3$. The local template is used to test for multi-field effects [57].

Non-Gaussianity arising from gravitational collapse **CHECK THIS** and from inflation with a non-canonical kinetic term can be described by the equilateral template

$$S^{equil}(k_1, k_2, k_3) = \frac{18}{5} A_\zeta^2 f_{NL}^{equil} \frac{(k_2 + k_3 - k_1)(k_3 + k_1 - k_2)(k_1 + k_2 - k_3)}{k_1 k_2 k_3} \quad (2.52)$$

$$= \frac{18}{5} A_\zeta^2 f_{NL}^{equil} \frac{K_3??}{6 K_{111}??}. \quad (2.53)$$

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

2.5 Step 1: the interaction Hamiltonian

2.5.1 Constraints

Since General Relativity has a diffeomorphism invariance, it has a redundancy which must be settled before the true degrees of freedom are manifest. One does this by decomposing the metric (typically in the *ADM* formalism) and expanding the action **is this what I want?**

$$S = \int dx^4 \sqrt{g} \left[\frac{R}{2} - \frac{1}{2}(\nabla\phi)^2 - V(\phi) \right] \quad (2.54)$$

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 interaction picture fields; and at third order, one obtains the cubic interaction Hamiltonian. **Check this!**

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 $\pi = \dot{\zeta}$. However if the Lagrangian contains other cubic terms with an explicit $\dot{\zeta}$, this relation will change and may not be easily invertible. However one can show that it will still be true that $\mathcal{H}_{int} = -\mathcal{L}_{int} + \mathcal{O}(\zeta^4, \dot{\zeta}^4)$, so calculating the cubic interaction Hamiltonian to this order from the cubic interaction Lagrangian is trivial.

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

$$H = H_b + H_0 + H_{int}, \quad (2.55)$$

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 , $H_{int}^{(3)}$, which we will use 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} .

2.5.2 Interactions

Simpler derivation neglecting the metric perturbations to begin with. Discuss DBI, then $P(X, \phi)$.

The DBI model has kinetic term

$$S_{DBI} = \int d^4x \sqrt{-g} \left(-\frac{1}{f(\phi)} \left((1 + f(\phi) \partial_\mu \phi \partial^\mu \phi)^{\frac{1}{2}} - 1 \right) - V(\phi) \right), \quad (2.56)$$

and we choose

$$f(\phi) = \frac{\lambda_{DBI}}{\phi^4}, \quad V(\phi) = V_0 - \frac{1}{2} m^2 \phi^2, \quad m = \sqrt{\beta_{IR}} H \quad (2.57)$$

in line with [77, 78].

The energy-momentum tensor in (1.3), in terms of the matter Lagrangian \mathcal{L}_m , is

$$T_{\mu\nu} = -2 \frac{\partial \mathcal{L}_m}{\partial g^{\mu\nu}} + g_{\mu\nu} \mathcal{L}_m \quad (2.58)$$

which for a $P(X, \phi)$ model evaluates to

$$T_{\mu\nu} = g_{\mu\nu} P(X, \phi) + P_{,X}(X, \phi) \partial_\mu \phi \partial_\nu \phi \quad (2.59)$$

$$\approx g_{\mu\nu} P(X, \phi) + 2X \delta_{0\mu} \delta_{0\nu} P_{,X}(X, \phi). \quad (2.60)$$

Following [11, 68], we can then calculate the useful quantities

$$P(X, \phi) = -\frac{1}{f(\phi)} \left(\sqrt{1 - f(\phi)X} - 1 \right) - V(\phi) \quad (2.61)$$

$$P_{,X}(X, \phi) = \frac{1}{2\sqrt{1 - f(\phi)X}} \quad (2.62)$$

$$\rho(X, \phi) = \frac{X}{\sqrt{1 - f(\phi)X}} - P(X, \phi) \quad (2.63)$$

$$= \frac{1}{\sqrt{1 - f(\phi)X}} (P + V) + V \quad (2.64)$$

$$\rho_{,X}(X, \phi) = \frac{1}{2} \frac{1}{(1 - f(\phi)X)^{\frac{3}{2}}} \quad (2.65)$$

$$c_s^2(X, \phi) = \frac{P_{,X}}{\rho_{,X}} \quad (2.66)$$

$$= 1 - f(\phi)X \quad (2.67)$$

Factor of two in $P_{,X}$? We see that, as expected, in the limit $\lambda_{DBI} \rightarrow 0$, the sound

speed tends to unity. From this, we can calculate the ϕ equation of motion (1.32)

$$\phi'' = -(3c_s^2 - \varepsilon)\phi' - \frac{3f_\phi(\phi)}{2f(\phi)}\phi'^2 + \frac{f_\phi(\phi)}{H^2 f(\phi)^2} - \left(V_\phi(\phi) + \frac{f_\phi(\phi)}{f(\phi)^2} \right) \frac{c_s^3}{H^2}. \quad (2.68)$$

This can be compared with the equations of motion presented in [79] and [74]. See also [80] for a brief discussion on setting consistent initial conditions.

We can now evolve τ_s , ϕ and H numerically using (1.33), (1.32) and (1.31) 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 (1.4), obtaining consistent initial conditions to start our numerical evolution. This procedure was described in [74]. **Is it exactly the same?** See [80] for a brief discussion on consistent initial conditions. We will discuss this further in section 2.5.2.

2.6 Step 2: the primordial bispectrum

2.6.1 Using the in-in formalism

Conditions for small non-Gaussianity,

so we need non-canonical kinetic term, or features, or something.

Cite [81–86].

The standard starting point for calculating higher-order correlators for models of inflation is the in-in formalism [19, 20]. The in-in formalism takes the time evolution of the interaction picture mode functions as an input for calculating the bispectrum. At tree-level, the in-in formalism gives us the following expression:

$$\langle \zeta_{\mathbf{k}_1}(\tau) \zeta_{\mathbf{k}_2}(\tau) \zeta_{\mathbf{k}_3}(\tau) \rangle = -i \int_{-\infty(1-i\varepsilon)}^{\tau} d\tau' a(\tau') \langle 0 | \zeta_{\mathbf{k}_1}(\tau) \zeta_{\mathbf{k}_2}(\tau) \zeta_{\mathbf{k}_3}(\tau) H_{int}(\tau') | 0 \rangle + c.c \quad (2.69)$$

where all the operators on the right-hand side are in the interaction picture and H_{int} is the interaction Hamiltonian, containing terms cubic in ζ . From this calculation we obtain the dimensionless shape function $S(k_1, k_2, k_3)$, defined in (2.37), which is then used as input into (2.97). As an example, if one takes

$H_{int} \propto \dot{\zeta}^3$, this set-up can produce the standard EFT shape [expand on this](#)

$$S(k_1, k_2, k_3) = \frac{k_1 k_2 k_3}{(k_1 + k_2 + k_3)^3}. \quad (2.70)$$

The central point, as noticed in [16], is that the integrand of (2.69) 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 (2.70), but can be regained (to arbitrary 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.

We now briefly outline the set-up of the standard calculation. The Lagrangian is expanded in the perturbations and used to obtain the Hamiltonian. The Hamiltonian is split into H_0 and H_{int} . The first part is used to evolve the interaction picture fields, ζ_I , which we will simply refer to as ζ , as in (2.69). The perturbations see an interaction Hamiltonian H_{int} , of which we will consider the part cubic in the perturbations, with time dependent coefficients due to the evolution of the background fields. The perturbations are assumed to be initially in the Bunch-Davies vacuum, but the non-linear evolution introduces correlations between the modes. As the modes cross the horizon they begin to behave classically and eventually freeze out.

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 [87]. This can be used, as pointed out in [16], 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 [7], 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.

The bispectrum arising from a single field inflation model, with a canonical kinetic term, slowly rolling, turns out to produce unobservably small non-Gaussianity [20]. However, by breaking these assumptions large signals can arise. These signals are usually calculated using (2.69) within tailored approximations. The results are not always separable, so further approximations must then be made to allow comparison with the CMB.

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

$$H_{int} = \int d^3x g(t) a^3 \left(\dot{\zeta}^I \right)^3 \quad (2.71)$$

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

$$\langle \zeta_{\mathbf{k}_1}(t) \zeta_{\mathbf{k}_2}(t) \zeta_{\mathbf{k}_3}(t) \rangle = \text{Re} \left(\left\langle -2i \zeta_{\mathbf{k}_1}^I(t) \zeta_{\mathbf{k}_2}^I(t) \zeta_{\mathbf{k}_3}^I(t) \int_{-\infty(1+i\varepsilon)}^t dt' H_{int}(t') \right\rangle \right). \quad (2.72)$$

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

$$\langle 0 | \zeta_{\mathbf{k}_1} \dots \zeta_{\mathbf{k}_n} | 0 \rangle = \left\langle 0 \left| : \zeta_{\mathbf{k}_1} \dots \zeta_{\mathbf{k}_n} : + \sum \text{pairwise contractions:} \right| 0 \right\rangle \quad (2.73)$$

where $: \ : :$ denotes normal ordering—ordering 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 $: \ : :$ contains uncontracted operators, the term vanishes inside $\langle \dots \rangle$.

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 be the term with every $\zeta_{\mathbf{k}}$ contracted in a pair.

To evaluate the pairwise contractions we will make use of the fact that,

from (2.11),

$$\langle \zeta_{\mathbf{k}_1}(t_1) \zeta_{\mathbf{k}_2}(t_2) \rangle = \left\langle 0 \left| \left(u_{k_1}(t_1) \hat{a}_{\mathbf{k}_1} + u_{k_1}^*(t_1) \hat{a}_{-\mathbf{k}_1}^\dagger \right) \left(u_{k_2}(t_2) \hat{a}_{\mathbf{k}_2} + u_{k_2}^*(t_2) \hat{a}_{-\mathbf{k}_2}^\dagger \right) \right| 0 \right\rangle \quad (2.74)$$

$$= \left\langle 0 \left| u_{k_1}(t_1) \hat{a}_{\mathbf{k}_1} u_{k_2}^*(t_2) \hat{a}_{-\mathbf{k}_2}^\dagger \right| 0 \right\rangle \quad (2.75)$$

$$= u_{k_1}(t_1) u_{k_2}^*(t_2) \left\langle 0 \left| [\hat{a}_{\mathbf{k}_1}, \hat{a}_{-\mathbf{k}_2}^\dagger] \right| 0 \right\rangle \quad (2.76)$$

$$= u_{k_1}(t_1) u_{k_2}^*(t_2) (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) \quad (2.77)$$

where we used (2.13).

So, inserting (2.71) into (2.72) and rewriting $\dot{\zeta}(\mathbf{x})$ in terms of its Fourier transform, we get

$$\langle \zeta_{\mathbf{k}_1}(t) \zeta_{\mathbf{k}_2}(t) \zeta_{\mathbf{k}_3}(t) \rangle \quad (2.78)$$

$$= \text{Re} \left(\left\langle -2i\zeta_{\mathbf{k}_1}(t) \zeta_{\mathbf{k}_2}(t) \zeta_{\mathbf{k}_3}(t) \int_{-\infty(1+i\varepsilon)}^t dt' \int d^3x \left(g(t) a^3 \dot{\zeta}^3(\mathbf{x}, t') \right) \right\rangle \right). \quad (2.79)$$

where the operators on the right hand side are in the interaction picture, but we have dropped the I for simplicity of notation. Continuing, we replace ζ with its Fourier transform

$$\left\langle -2i\zeta_{\mathbf{k}_1}(t) \zeta_{\mathbf{k}_2}(t) \zeta_{\mathbf{k}_3}(t) \int_{-\infty(1+i\varepsilon)}^t dt' \int d^3x \left(g(t) a^3 \dot{\zeta}^3(\mathbf{x}, t') \right) \right\rangle \quad (2.80)$$

$$= \left\langle -2i\zeta_{\mathbf{k}_1}(t) \zeta_{\mathbf{k}_2}(t) \zeta_{\mathbf{k}_3}(t) \int_{-\infty(1+i\varepsilon)}^t dt' (g(t) a^3) \cdot \int \frac{d^3p}{(2\pi)^3} \frac{d^3q}{(2\pi)^3} \frac{d^3r}{(2\pi)^3} \dot{\zeta}_{\mathbf{p}}(t') \dot{\zeta}_{\mathbf{q}}(t') \dot{\zeta}_{\mathbf{r}}(t') \int d^3x e^{i\mathbf{x} \cdot (\mathbf{p} + \mathbf{q} + \mathbf{r})} \right\rangle \quad (2.81)$$

$$= \left\langle -2i\zeta_{\mathbf{k}_1}(t) \zeta_{\mathbf{k}_2}(t) \zeta_{\mathbf{k}_3}(t) \int_{-\infty(1+i\varepsilon)}^t dt' (g(t) a^3) \cdot \int \frac{d^3p}{(2\pi)^3} \frac{d^3q}{(2\pi)^3} \frac{d^3r}{(2\pi)^3} \dot{\zeta}_{\mathbf{p}}(t') \dot{\zeta}_{\mathbf{q}}(t') \dot{\zeta}_{\mathbf{r}}(t') (2\pi)^3 \delta^{(3)}(\mathbf{p} + \mathbf{q} + \mathbf{r}) \right\rangle. \quad (2.82)$$

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. **CHECK THIS.**

The results of the contractions will therefore have the form

$$\left\langle \zeta_{\mathbf{k}_1}(t) \dot{\zeta}_{\mathbf{p}}(t') \right\rangle = u_{k_1}(t) \dot{u}_p^*(t') (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{p}). \quad (2.83)$$

and permutations. Using the delta functions to perform the momentum integrals, switching to conformal time, and recalling (2.36), we obtain

$$B(k_1, k_1, k_3) = \text{Re} \left(-2i u_{k_1}(0) u_{k_2}(0) u_{k_3}(0) \cdot \int_{-\infty(1+i\varepsilon)}^0 d\tau' \left(g(\tau') a^4(\tau') \right) \dot{u}_{k_1}^*(\tau') \dot{u}_{k_2}^*(\tau') \dot{u}_{k_3}^*(\tau') + \text{perms} \right). \quad (2.84)$$

If one wanted to numerically calculate the bispectrum resulting from (2.71), 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 u_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 (2.84) is explicitly separable in k_1 , k_2 and k_3 .

We will now continue the calculation within the slow-roll approximation, using (2.16). 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 sign?

$$\int_{-\infty(1+i\varepsilon)}^0 d\tau \tau^2 e^{-ic_s K\tau} = \frac{-2i}{c_s^3 K^3} \quad (2.85)$$

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

$$S(k_1, k_2, k_3) = -3 \frac{g H^5}{8\varepsilon^3} \frac{k_1 k_2 k_3}{(k_1 + k_2 + k_3)^3} \quad (2.86)$$

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 [10, 17] by setting $g(t) = -\frac{\varepsilon}{H c_s^2} (1 - c_s^2) \left(1 + \frac{2\tilde{c}_3}{3c_s^2} \right)$ and recalling $\Phi = \frac{3}{5}\zeta$. Check this again!

2.6.2 The squeezed limit consistency condition

The squeezed limit of canonical single-field bispectra will not cause observable deviations from a Gaussian universe, due to a cancellation when switching to physical coordinates [88]. 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 [89, 90]. With $\mathbf{k}_S \equiv (\mathbf{k}_2 - \mathbf{k}_3) / 2$:

$$S(k_1, k_2, k_3) = - \left[(n_s - 1)|_{k_S} + \mathcal{O}\left(\frac{k_1^2}{k_S^2}\right) \right] P_\zeta(k_1)P_\zeta(k_S), \quad k_1 \ll k_S \quad (2.87)$$

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 [90].

2.6.3 Basic shapes

With a canonical kinetic term, the slow-roll result for the shape is:

$$S^{Molda}(k_1, k_2, k_3) = -\frac{1}{32} \frac{H^4}{12\varepsilon^2} \left((3\varepsilon - 2\eta) \frac{K_3}{K_{111}} + \varepsilon \left(K_{12} + 8 \frac{K_{22}}{K} \right) \right) \quad (2.88)$$

with $\eta = 2\varepsilon$ for (2.27). At the primordial level, this is well approximated by the separable local template (2.49). However, the amplitude of this shape is expected to be tiny, and the dominant contributions (the squeezed configurations) are expected to have no observable effect [88].

For the featureless DBI scenario, the shape function is [21]:

$$S^{DBI}(k_1, k_2, k_3) = -\frac{35}{108} \left(\frac{1}{c_s^2} - 1 \right) 6 \left(\frac{H^2}{4\varepsilon c_s} \right)^2 \left(\frac{3}{5} \right) \left(-\frac{3}{7} \right) \frac{K_5 + 2K_{14} - 3K_{23} + 2K_{113} - 8K_{122}}{K_{111}K^2} \quad (2.89)$$

to leading order in slow-roll. Any constraint on the magnitude of this template can be translated into one on the effective sound speed which from *Planck* has a lower limit $c_s^{DBI} \geq 0.087$ at 95% significance [57]. The shape (2.89) can be approximated by the separable equilateral template (2.52).

2.6.4 Summary, scaling

There is an extensive literature on the calculation of bispectra from models of inflation [7, 35, 38, 69, 72, 91–94]. 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 [21, 74, 79, 95–98]; effectively single-field models with imaginary sound speeds can generate a bispectrum mostly orthogonal to the usual equilateral and local templates [99]. The methods outlined in this thesis have been implemented and tested for single-field models, with multi-field models being a prime target for future work.

These templates can be modified to be more physically realistic by including scaling consistent with the spectral index n_s [57]. For example, we can add some scale dependence to the DBI (2.89) in a reasonable first approximation by including a prefactor. We define the product scaling template

$$S^{DBI-n_s}(k_1, k_2, k_3) = \left(\frac{k_1 k_2 k_3}{k_*^3} \right)^{\frac{n_{NG}}{3}} S^{DBI}(k_1, k_2, k_3) \quad (2.90)$$

and the sum scaling template

$$S^{DBI-n_s}(k_1, k_2, k_3) = \left(\frac{k_1 + k_2 + k_3}{3k_*} \right)^{n_{NG}} S^{DBI}(k_1, k_2, k_3) \quad (2.91)$$

with $n_{NG} = 2(-2\varepsilon - \varepsilon_s - \eta) - 2\varepsilon_s = 2(n_s - 1) - 2\varepsilon_s$.

Much success has been had in constraining non-Gaussianity in the CMB using separable approximations to these approximate templates. Other methods target oscillations [100, 101], 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. Our motivation in this work for directly calculating the primordial bispectrum in a separable form is to build towards a pipeline to constrain a broader section of the model space, removing these layers of approximations, though these standard results provide useful validation tests.

2.6.5 Shapes from features during inflation

Explicit details of how resonance and features generate large NG. Refer to 2.6.

For our more stringent validation tests we work with feature model scenarios based on the above base models. To explore non-Gaussianity coming from sharp

features we include a kink [102]

$$V(\phi) = V_{\phi^2}(\phi) \left(1 - c \tanh \left(\frac{\phi_f - \phi}{d} \right) \right). \quad (2.92)$$

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

$$V(\phi) = V_{\phi^2}(\phi) \left(1 + bf \sin \left(\frac{\phi}{f} \right) \right). \quad (2.93)$$

For more details on these models, see [92].

We now turn to feature templates. The result of adding a feature of the form (2.92) is to add oscillatory features of the form

$$S^{\cos}(k_1, k_2, k_3) = \cos(w(k_1 + k_2 + k_3)) \quad (2.94)$$

though more realistically there is some phase, shape dependence and a modulating envelope, as detailed in [35]. The result of adding a resonant feature of the form (2.93) is to generate logarithmic oscillatory features in the shape function of the form

$$S^{\ln-\cos}(k_1, k_2, k_3) = \cos(w \ln(k_1 + k_2 + k_3)). \quad (2.95)$$

With a non-canonical kinetic term, this can also cause out-of-phase oscillations in the folded limit as well as a modulating shape, see [97].

2.7 Step 3: the CMB bispectrum

For a three dimensional statistical distribution projected onto a sphere, we expect the three-point statistics to be projected like (as in (2.106))

$$\langle f_{l_1 m_1} f_{l_2 m_2} f_{l_3 m_3} \rangle = B_{l_1 l_2 l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \quad (2.96)$$

where $\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ enforces statistical isotropy.

For the CMB in particular, we calculate the projected bispectrum using [65]

$$b_{l_1 l_2 l_3} = \left(\frac{2}{\pi}\right)^3 \int_0^\infty dr r^2 \int d^3 k (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3) \prod_{i=1}^3 [j_{l_i}(k_i r) \Delta_{l_i}(k_i)], \quad (2.97)$$

analogously to (2.32). 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 [65]. 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 ($l \ll 200$) then we can use the **Sachs-Wolfe** approximation

$$\Delta_l(k) = \frac{1}{3} j_l((\tau_0 - \tau_{dec})k). \quad (2.98)$$

How does this relate to χ_* , and why 3?

The integral becomes

$$b_{l_1 l_2 l_3} = \left(\frac{2}{3\pi}\right)^3 \int_0^\infty dr r^2 \int d^3 k \prod_{i=1}^3 [j_{l_i}(k_i r) j_l((\tau_0 - \tau_{dec}) k_i)]. \quad (2.99)$$

Now, using that

$$\int_0^\infty dk j_l(k) j_l(xk) = \frac{\pi}{2} \frac{x^{-(l+1)}}{2l+1} \quad \text{for } x > 1 \quad (2.100)$$

$$\frac{\pi}{2} \frac{x^l}{2l+1} \quad \text{for } x < 1 \quad (2.101)$$

we find that

$$b_{l_1 l_2 l_3} = \frac{f_{NL}}{27} \frac{1}{(2l_1 + 1)(2l_2 + 1)(2l_3 + 1)} \left[\int_0^1 dx x^{l_1 + l_2 + l_3 + 2} + \int_1^\infty dx x^{-l_1 - l_2 - l_3 - 1} \right] \quad (2.102)$$

$$= \frac{f_{NL}}{27} \frac{1}{(2l_1 + 1)(2l_2 + 1)(2l_3 + 1)} \left[\frac{1}{l_1 + l_2 + l_3 + 3} + \frac{1}{l_1 + l_2 + l_3} \right]. \quad (2.103)$$

Note that this example of constant S is not especially realistic as a shape that could come from a model of inflation. Note also that the acoustic oscillations 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 (2.99) and (2.102) even for this simple case.

2.8 Step 4: CMB bispectrum estimation

2.8.1 Linear regression

E.g. [18, 103–106].

Work on forecasting future constraints on features was performed in [107] **Where else?** For recent reviews, see [108–112]. While there are some standard shapes with very tight constraints, there are still regions of the model space that have not been constrained.

The bispectrum derived from theory is denoted

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle \quad (2.104)$$

whereas the bispectrum measured from the CMB is

$$B_{m_1 m_2 m_3}^{l_1 l_2 l_3} = a_{l_1 m_1}^{\text{obs}} a_{l_2 m_2}^{\text{obs}} a_{l_3 m_3}^{\text{obs}}. \quad (2.105)$$

We also have

$$B_{l_1 l_2 l_3} = \sum_{m_i} \binom{l_1 \ l_2 \ l_3}{m_1 \ m_2 \ m_3} B_{m_1 m_2 m_3}^{l_1 l_2 l_3}. \quad (2.106)$$

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:

$$\sum_{l_i, m_i} \left(\lambda \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} - \frac{a_{l_1 m_1}^{\text{obs}} a_{l_2 m_2}^{\text{obs}} a_{l_3 m_3}^{\text{obs}}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right)^2. \quad (2.107)$$

The solution to this is simply

$$\lambda = \frac{\frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \cdot \frac{a_{l_1 m_1}^{\text{obs}} a_{l_2 m_2}^{\text{obs}} a_{l_3 m_3}^{\text{obs}}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}}}{\left| \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right|^2} \quad (2.108)$$

where \cdot denotes summation over l_i and m_i . Defining [Change to \[113, 114\] !](#)

$$N = \left| \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right|^2 \quad (2.109)$$

we can rewrite this in the usual way (with λ identified as the result of the estimator \mathcal{E})

$$\mathcal{E} = \frac{1}{N} \sum_{l_i, m_i} \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle a_{l_1 m_1}^{obs} a_{l_2 m_2}^{obs} a_{l_3 m_3}^{obs}}{C_{l_1} C_{l_2} C_{l_3}}. \quad (2.110)$$

However, if we simply calculate this quantity for some theoretical model we will not be able to interpret the result. This is because even a CMB sky in a universe with initial conditions drawn from a purely Gaussian distribution could, through random chance, result in a non-zero \mathcal{E} . This problem is solved using Gaussian maps. Many CMB are generated from Gaussian initial conditions, and then the estimator is applied to each. This gives a distribution of \mathcal{E} , from which we can calculate 1σ and 2σ regions for \mathcal{E} . 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.

This procedure must be modified to account for experimental noise, beam effects and the presence of a mask (to exclude regions of the sky saturated by our galaxy). The assumption is also made that the covariance matrix is diagonal, and the following quantities are defined:

$$C_{l_1 m_1, l_2 m_2} \approx C_l \delta_{l_1 l_2} \delta_{m_1 - m_2}, \quad (2.111)$$

$$\tilde{C}_l = b_l^2 C_l + N_l, \quad (2.112)$$

$$\bar{b}_{l_1 l_2 l_3} = b_{l_1} b_{l_2} b_{l_3} b_{l_1 l_2 l_3}, \quad (2.113)$$

to incorporate noise from the experiment. The estimator can then be written as [\[113, 114\]](#)

$$\mathcal{E} = \frac{1}{\bar{N}^2} \sum_{l_i m_i} \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \bar{b}_{l_1 l_2 l_3}}{\tilde{C}_{l_1} \tilde{C}_{l_2} \tilde{C}_{l_3}} (a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} - 3C_{l_1 m_1, l_2 m_2}^{sim} a_{l_3 m_3}). \quad (2.114)$$

Here, the diagonal approximation is insufficient for the linear term, so the full covariance matrix $C_{l_1 m_1, l_2 m_2}^{sim}$ is calculated 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. We have here used the Gaunt integral

$$\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} = \int d\Omega Y_{l_1 m_1}(\hat{\mathbf{n}}) Y_{l_2 m_2}(\hat{\mathbf{n}}) Y_{l_3 m_3}(\hat{\mathbf{n}}) \quad (2.115)$$

$$= h_{l_1 l_2 l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \quad (2.116)$$

and we can then define

$$b_{l_1 l_2 l_3} = h_{l_1 l_2 l_3}^{-1} B_{l_1 l_2 l_3}. \quad (2.117)$$

2.8.2 Complexity of bispectrum estimation

Memory bound computation, so tabulation won't help. The time is due to moving data. Not much to do with the maps?

The bispectrum, like the power spectrum, is a quantity that describes the statistical distribution of which our universe is only one realisation. We use this one sky we have access to to estimate the amplitude of particular bispectrum templates, and use these estimates to constrain inflationary physics; see [108, 109] for recent reviews. There are two parts to the pipeline of bispectrum estimation. Firstly, calculating the primordial bispectrum at the end of some inflation scenario, and then calculating the effect this bispectrum has on some appropriate observable today. One well-developed example is the bispectrum of temperature fluctuations in the CMB, which uses transfer functions (which we saw previously in (2.32)) 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 second 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 [18, 56]. 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 [64–66] 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 [64–66] allows to go beyond this paradigm, efficiently constraining inflationary bispectra in the CMB using all of the shape information; essentially constraining an f_{NL} specific to a given bispectrum. This bypasses the approximation step at the level of the templates, of finding a separable approximation to the primordial bispectrum. In this work, our numerical methods remove the need for some of the approximations made before this, during inflation, directly linking the parameters of the inflation scenario with the relevant observable. In addition to this improvement in accuracy, calculating the modal decomposition directly from the model of inflation is far more efficient than numerically calculating the bispectrum configuration by configuration.

If the shape function (2.37) has the form:

$$S(k_1, k_2, k_3) = X(k_1)Y(k_2)Z(k_3), \quad (2.118)$$

or can be expressed as a sum of such terms, it is called separable. The link between the separability of the primordial bispectrum and the reduced CMB bispectrum can be seen from (2.97), where we also see that if the primordial bispectrum is separable then the overall dimension of the calculation can be reduced from seven to five, 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 [63].

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 difficult to see how constraints on existing features can be applied.

2.9 Previous work on in-in separability.

In [16] it was pointed out that one can compute using the tree-level in-in formalism in such a way as to preserve its intrinsic separability. In addition to making this point, [16] 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 [64–66] 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 results of these methods (not using the work of [16]) are constraints on the parameters of certain inflation models through approximate phenomenological templates. These constraints can be found in [17, 57]. The idea of [16] is an extension of that philosophy to the primordial level, and our work is in implementing that idea. In [64–66] 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 [16]. We restructure the methods, improving on the work of [16] 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 [16] 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 [16] 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 [16], and that includes the magnitude as well as the shape information on the full tetrapyd. This is discussed in section 3.2.

Summary of the achievements and limitations of [16], how I went beyond them.

2.10 Configuration-by-configuration codes

Cite [115]. Previous work on the numerical calculations of inflationary non-Gaussianity include the BINGO code [116], Chen et al [91, 92], the work of Horner et al [117–119] and the Transport Method [73, 120–122]. 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 (3.2). 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 previously 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 formalism³. 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 [99].

2.10.1 Usage in recent works

CHECK ALL THIS. In [99, 123] and especially in [124], a large ensemble of inflationary trajectories is considered using the transport method [121]. For these trajectories, the bispectrum is calculated for a certain number of k -configurations—in these papers, f_{NL}^{equil} and f_{NL}^{flat} refer to the definition given in (2.47) evaluated on equilateral and flat configurations respectively. In these works it was found that squeezed limit configurations were much more difficult to calculate using that software. While this analysis is very sophisticated, much information is lost due to taking the bispectrum at points only, and not using the full inflationary shape function.

2.10.2 Limitations

However despite the differences, 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

³Infinite hierarchy of coupled equations. Choose a basis to make it nice? Real vs Imag...? Sounds cool, not worth pursuing. 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 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 the mode expansion of $\zeta_k(\tau)$ into (2.6) are coupled in an infinite hierarchy.

of points which must be processed further to be usefully compared to observation. Secondly, they must carefully implement some variation of the *i.e.* prescription without affecting the numerical results. In [73] this is achieved in the initial conditions for the bispectra; other methods impose some non-trivial cutoff at early times.

2.11 Separable approximations to non-separable templates

Separability is required, so usually one approximates non-separable templates by separable ones. For example, the (2.89) is closely approximated by the equilateral template (2.52). [Quantitative comparison of equilateral template and DBI shape, with and without scaling.](#)

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 (2.49), equilateral (2.52) and orthogonal [71] 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 oscillations⁴, but is limited in that it cannot function for shapes whose shape dependence is not sum-separable.

The Modal method [125] 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.

[These methods have been used to obtain constraints...](#)

⁴Since $e^{iw(k_1+k_2+k_3)} = e^{iwk_1}e^{iwk_2}e^{iwk_3}$.

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. This basis is then projected forwards onto the CMB. This is a massively resource-intensive calculation with the naive implementation scaling as p_{\max}^6 and requiring terabytes of RAM for reasonable method parameters, which slow the calculation due to the need to move data back and forth **because data is really heavy and hard to carry**. 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.

The quantity in (2.111) is not assumed to be diagonal **CHECK THIS**, it is calculated by taking an ensemble average of 140 Gaussian simulations. f_{NL} estimates are made for all of these simulations, with the results found to be distributed normally. The variance of this distribution then provides us with 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_{\max} to k_{\min} was important in determining the success of the method. If k_{\max}/k_{\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 $k_{\max}/k_{\min} = 1000$ was sufficient, hence that is the value we will use in testing out basis sets **CHECK THIS!!!**.

Chapter 3

Decomposing Primordial Shapes

As we have seen, while at tree level the bispectrum is an integral over a separable integrand (see (2.69) and the example presented in (2.84)), 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 [16]. That work extended the separable methodology previously implemented for the CMB bispectrum [64–66]. 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 (2.37) using a set of basis functions as

$$S(k_1, k_2, k_3) = \sum_n \alpha_n Q_n(k_1, k_2, k_3), \quad (3.1)$$

where the basis functions $Q_n(k_1, k_2, k_3)$ are explicitly separable functions of their arguments. 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 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_{\max} . 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 (2.69) is inherently separable given the form of the cubic interaction Hamiltonian H_{int} . This can be clearly seen in the integrand of the example in (2.84). Consider indexing with $i = 1, 2, 3\dots$ the interaction vertices in H_{int} , so then after the contractions have been performed, the bispectrum (2.69) can be expressed as a sum over separable contributions of the form:

$$\begin{aligned} S(k_1, k_2, k_3) &= \sum_i I^{(i)}(k_1, k_2, k_3) \\ &= \text{Re} \sum_i \left[v^{(i)}(k_1, k_2, k_3) \int_{-\infty(1-i\varepsilon)}^0 d\tau w^{(i)}(\tau) F^{(i)}(\tau, k_1) G^{(i)}(\tau, k_2) J^{(i)}(\tau, k_3) \right. \\ &\quad \left. + \text{cyclic perms} \right] \end{aligned} \tag{3.2}$$

where $w^{(i)}(\tau)$ is a function of the scale factor and the other background parameters (1.39) 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 (2.38). 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 (2.84). 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 then give

$$F^{(i)}(\tau, k) = G^{(i)}(\tau, k) = J^{(i)}(\tau, k) = k^2\zeta_k(0)\dot{\zeta}_k^*(\tau) \quad (3.3)$$

where the k^2 comes from the definition of the shape function (2.37), and we obtain $\zeta_k(\tau)$ by numerically solving the equation of motion (2.6).

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. 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 $F^{(i)}(k, \tau) = k^2\zeta_k(0)\zeta_k^*(\tau)$, 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, and indeed in dealing with the

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

spatial derivatives in section 4.5 we will find it useful to change to a different basis than the one used to perform the time integrals of the decompositions.

Jarring, move earlier. Using the approximate mode functions (2.22), an explicit example for the first interaction term in (4.8), i.e. $H_{\text{int}}^{(1)} = \zeta'^2 \zeta$, takes the form

$$F^{(1)}(\tau, k) = G^{(1)}(\tau, k) = c_s k^2 \tau \frac{H^2}{4\varepsilon c_s} e^{ic_s k \tau}, \quad (3.4)$$

$$J^{(1)}(\tau, k) = (1 - ikc_s \tau) \frac{H^2}{4\varepsilon c_s} e^{ic_s k \tau}. \quad (3.5)$$

In the simple mode approximation (2.22), such terms in (3.2) 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 [20]. However, for high precision bispectrum predictions we must incorporate the full time-dependence, while solving (2.6) to find accurate mode functions $\zeta_{\mathbf{k}}(\tau)$ numerically. Obtaining the full 3D bispectrum directly is computationally demanding at high resolution because it requires repetitive integration of (3.2) at each specific point for the wavenumbers (k_1, k_2, k_3) , a problem which is drastically compounded by bispectrum parameter searches e.g. for oscillatory models.

We will now link (3.2) to (3.1). Consider representing the primordial shape function $S(k_1, k_2, k_3)$ in (3.2) as a mode expansion for each interaction term $I^{(i)}(k_1, k_2, k_3)$ as

$$S(k_1, k_2, k_3) = \sum_i I^{(i)}(k_1, k_2, k_3) = \sum_i \sum_n \alpha_n^{(i)} Q_n(k_1, k_2, k_3), \quad (3.6)$$

where $Q_n(k_1, k_2, k_3)$ is separable, built out of some orthonormal set $q_p(k)$ as in (3.21). 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 (3.2) as:

$$F^{(i)}(\tau, k) = \sum_p f_p^{(i)}(\tau) q_p(k), \quad (3.7)$$

$$\text{where } f_p^{(i)}(\tau) = \int_{k_{\min}}^{k_{\max}} dk F^{(i)}(\tau, k) q_p(k). \quad (3.8)$$

Substituting these expansions into (3.2), we obtain the following decomposi-

tion for the i -th vertex contribution,

$$\begin{aligned} I^{(i)}(k_1, k_2, k_3) &= \text{Re} \left[v^{(i)}(k_1, k_2, k_3) \int_{-\infty(1-i\varepsilon)}^0 d\tau w^{(i)}(\tau) \right. \\ &\quad \cdot \sum_p f_p^{(i)}(\tau) q_p(k_1) \sum_r g_r^{(i)}(\tau) q_r(k_2) \sum_s h_s^{(i)}(\tau) q_s(k_3) + \text{cyclic perms} \left. \right] \end{aligned} \quad (3.9)$$

$$\begin{aligned} &= v^{(i)}(k_1, k_2, k_3) \sum_{prs} \left(\text{Re} \int_{-\infty}^0 d\tau w^{(i)}(\tau) \right. \\ &\quad \cdot f_p^{(i)}(\tau) g_r^{(i)}(\tau) h_s^{(i)}(\tau) \left. \right) q_p(k_1) q_r(k_2) q_s(k_3) + \text{cyclic perms}. \end{aligned} \quad (3.10)$$

For the sake of compactness of notation we 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 [16], 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,

$$\begin{aligned} I^{(i)}(k_1, k_2, k_3) &= v^{(i)}(k_1, k_2, k_3) \sum_P \tilde{\alpha}_P^{(i)} q_P(k_1, k_2, k_3) + \text{cyclic perms} \\ &= \sum_P \tilde{\alpha}_P^{(i)} \sum_{\tilde{P}} V_{P\tilde{P}}^{(i)} q_{\tilde{P}}(k_1, k_2, k_3) + \text{cyclic perms} \\ &= \sum_{\tilde{P}} \alpha_{\tilde{P}}^{(i)} q_{\tilde{P}}(k_1, k_2, k_3) + \text{cyclic perms}, \end{aligned} \quad (3.11)$$

where we have written

$$\tilde{\alpha}_P^{(i)} = \tilde{\alpha}_{prs}^{(i)} = \text{Re} \int_{-\infty}^0 d\tau w^{(i)}(\tau) f_p^{(i)}(\tau) g_r^{(i)}(\tau) h_s^{(i)}(\tau), \quad (3.12)$$

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

$$v^{(i)}(k_1, k_2, k_3) q_P(k_1, k_2, k_3) = \sum_{\tilde{P}} V_{P\tilde{P}}^{(i)} q_{\tilde{P}}(k_1, k_2, k_3), \quad (3.13)$$

and

$$\alpha_P^{(i)} = \sum_{\tilde{P}} \tilde{\alpha}_{\tilde{P}}^{(i)} V_{\tilde{P}P}^{(i)} + \text{cyclic perms}. \quad (3.14)$$

The numerical calculation of $V_{P\tilde{P}}^{(i)}$ (as defined by (3.13)) is highly efficient as

$v^{(i)}(k_1, k_2, k_3)$ 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 (3.7)) 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 (3.6) 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 (3.12) we need to calculate.

Note the basis sets on the left and right hand side of (3.13) 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 \mathcal{P}_0 , 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 (3.13) to be an expanded version of that on the left. For example, if the left hand basis was \mathcal{P}_0 of size p_{\max} , the right hand basis would be \mathcal{P}_1 of size $p_{\max} + 3$.

We can see from (3.12) that the number of time integrals needed is controlled by $N_V \times p_{\max}^3$ 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 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 (3.6) is then simply

$$S(k_1, k_2, k_3) = \sum_n \left(\sum_i \alpha_n^{(i)} \right) Q_n(k_1, k_2, k_3) = \sum_n \alpha_n Q_n(k_1, k_2, k_3). \quad (3.15)$$

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_{\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

²In fact the number is not quite p_{\max}^3 . 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 (3.12) will always be (at least) symmetric in p and r .

integrals (3.12) are now independent of the k -configuration³. 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 (3.15) 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_{\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_{\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 (2.88) and (2.89), 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 $Q_{pqr} \rightarrow \tilde{Q}_{pqr}$ can be trivially and cheaply achieved through a linear transformation Γ with the new expansion coefficients given by $\tilde{\alpha}_{abc} = \Gamma_{ap}\Gamma_{bq}\Gamma_{cr}\alpha_{pqr}$.

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.

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

3.2 Testing basis sets

We define the inner product of two bispectrum shape functions as

$$S_1 \cdot S_2 = \langle S_1, S_2 \rangle = \int_{T_k} d^3k S_1(k_1, k_2, k_3) S_2(k_1, k_2, k_3), \quad (3.16)$$

where T_k refers to the tetrapyd, the region of the cube $[k_{\min}, k_{\max}]^3$ that obeys the triangle inequality. Following [126] we define the two correlators:

$$\mathcal{S}(S_1, S_2) = \frac{S_1 \cdot S_2}{\sqrt{(S_1 \cdot S_1)(S_2 \cdot S_2)}}, \quad \mathcal{A}(S_1, S_2) = \sqrt{\frac{S_1 \cdot S_1}{S_2 \cdot S_2}}. \quad (3.17)$$

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 [64–66], 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:

$$\begin{aligned} \mathcal{E}(S_1, S_2) &= \sqrt{\frac{|S_1 - S_2|^2}{|S_2|^2}} = \sqrt{\frac{|S_1|^2 - 2S_1 \cdot S_2 + |S_2|^2}{|S_2|^2}} \\ &= \sqrt{\mathcal{A}(S_1, S_2)^2 - 2\mathcal{A}(S_1, S_2)\mathcal{S}(S_1, S_2) + 1}. \end{aligned} \quad (3.18)$$

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

$$\mathcal{E}^{(\mathcal{A}=1)}(S_1, S_2) = \sqrt{2(1 - \mathcal{S}(S_1, S_2))}. \quad (3.19)$$

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 [126] is a far better representation of actual convergence between two shape functions than the correlation used in [16], as it is easier to interpret and more stringent.

The full relative difference (3.18) 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 tetrabyd, relative to the representative value

$$S^{rep} = \sqrt{\frac{S \cdot S}{\int_{T_k} d^3k}}. \quad (3.20)$$

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 this work however we focus on calculating the primordial shape function accurately and efficiently 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. 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

$$Q_n(k_1, k_2, k_3) \equiv q_p(k_1) q_r(k_2) q_s(k_3). \quad (3.21)$$

Here, n labels the integer triplet $n \leftrightarrow \{prs\}$ in some appropriate manner. This modal expansion is terminated at some p_{\max} for which $\max(p, r, s) < p_{\max}$. 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 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 (3.21), thus determining the nature of the numerical errors in the calculated bispectrum. However, since we are going beyond the featureless examples of [16]

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 $[k_{\min}, k_{\max}]^3$ ³. This is because the only decomposition we actually perform is a one-dimensional integral over $[k_{\min}, k_{\max}]$ (as we saw in (3.8)). 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 (2.88) 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 (2.89), this effect can be disastrous if not handled properly. This can be easily seen from (2.52), 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 $1/k$ behaviour⁴.

⁴There are results in the literature that describe generic $K = k_1 + k_2 + k_3$ poles in correlators—see for example [127]. A simple example can be understood by recalling that in standard

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

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

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

$$\bar{k} = \frac{2k - k_{\max} - k_{\min}}{k_{\max} - k_{\min}}. \quad (3.23)$$

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 $r = 0, 1, \dots, p_{\max} - 1$. These were considered also in [16], 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 [16] to calculate primordial bispectra were as follows. The first was the Legendre polynomials taken with a log-mapping between k and the polynomial argument as

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

$$\ln k = \frac{2 \ln k - \ln k_{\max} - \ln k_{\min}}{\ln k_{\max} - \ln k_{\min}}. \quad (3.25)$$

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

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.

an unnormalised basis set of the Legendre polynomials divided by k :

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

where the rescaled \bar{k} is defined in (3.23). This can also be thought of as expanding the bispectrum $k_1 k_2 k_3 S(k_1, k_2, k_3)$ 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 (3.16). However, as shown in [16], 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 [16], 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 (2.97), 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 $r = 0, \dots, N_{\text{initial}} - 1$ (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

$$\tilde{f}(k) = \text{Orth}[f(k), B_0] \equiv f(k) - \sum_{q \in B_0} \frac{\langle f, q \rangle}{\langle q, q \rangle} q(k) \quad (3.27)$$

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

$$q_{p_{\max}-1}(k) = \text{Orth}[1/k, \mathcal{P}_0^{p_{\max}-1}] . \quad (3.28)$$

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

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

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 $q_{p_{\max}}(k) = \text{Orth}[1/k, \mathcal{F}_0]$ 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} .

In order to compare the efficacy of the different basis function sets, we have investigated their convergence on Maldacena's shape (2.88) and the DBI shape (2.89). To mimic the in-in calculation, we expand the shape on the cube, but test the result on the tetrabyd using (3.18). The results are shown in figure 3.1.

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)
<i>scaling</i>	Legendre polynomials	$k^{-1}, \ln(k)k^{-1}$	
<i>resonant</i>	$\frac{P_n(\bar{k})}{\sqrt{\bar{k}}}$, 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.

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 (2.94) and the resonance model (2.95) 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 (2.89) 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 DBI template (2.89) 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 exam-



(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 (2.88) and (b) the DBI template (2.89). The pure Legendre \mathcal{P}_0 basis requires many terms to fit the $1/k$ behaviour in both Maldacena's template (2.88) and the DBI template (2.89). 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 $p_{\max} = 5$. The Fourier \mathcal{F}_1 basis performs well, but converges more slowly than the \mathcal{P}_1 basis. Note that the convergence errors for (2.89) are larger than (2.88) because of the larger contributions outside the tetrapyd dominating the fit. In this plot and the following, unless otherwise stated, $k_{\max} = 1000k_{\min}$.



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 (2.94) (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 (2.89) and (2.94)), we see that again the *scaling* basis converges with the fewest modes. Note that before the expansion has fully converged, the fit on the tetraipyd can actually degrade slightly when the basis set is extended. This is an artifact of fitting on the cube and restricting (3.18) to the physical configurations on the tetraipyd; 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 (2.89) and (2.95). 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.



Figure 3.4: For the scale-dependent DBI template (2.90), 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 $p_{\max} = 30$. 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_{\max}/k_{\min} . The convergence power of these basis sets will allow us to efficiently capture the scale dependence of the numerically calculated shape function.



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.



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^{ns} 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.

ples 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 (2.90) 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),

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

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

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

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.

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

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

This is motivated by

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

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

which is valid when $\epsilon \ln(k_{\max}/k_{\min})$ 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 3.6) we have $\ln\left(\frac{k_{\max}}{k_{\min}}\right) \approx 6.91$. 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 [17]. 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 [16] in that it also includes a factor of $\frac{1}{\sqrt{k}}$ to retain orthogonality. We define the n th basis element as

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

We then see that

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

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 $p_{\max} = 30$) 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 $p_{\max} = 30$.

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.



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.

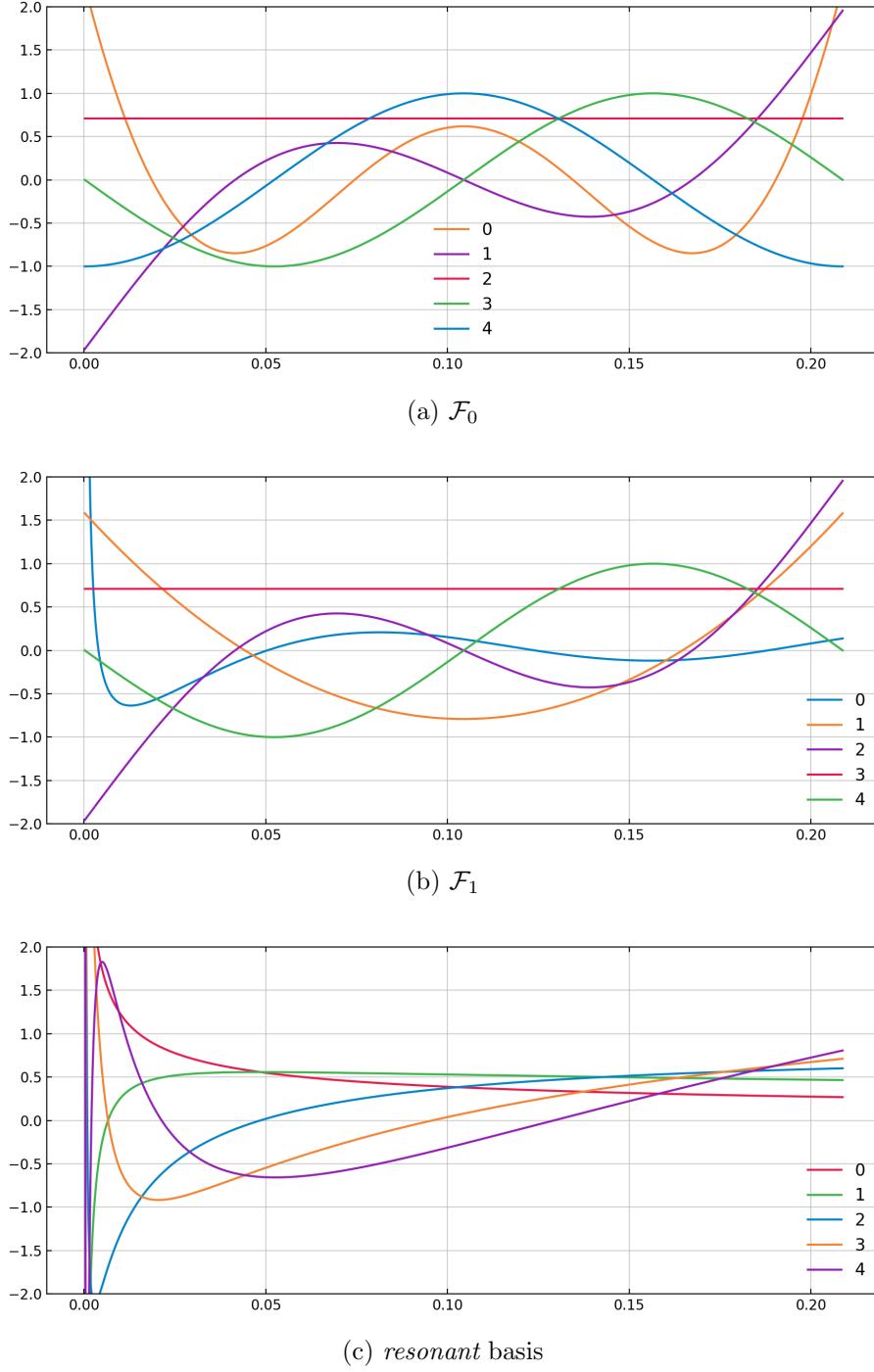


Figure 3.8: We plot the \mathcal{F}_0 , \mathcal{F}_1 and *resonant* basis sets from table 3.1, for $p_{\max} = 5$. Note that \mathcal{F}_0 and \mathcal{F}_1 have Fourier basis elements in common, but differ in which functions they are augmented by. The *resonant* basis differs in all of its basis elements,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

3.5.1 Large non-physical contributions

When a shape is dominated by its non-physical configurations (those that do not obey the triangle inequality (2.39)) then we can see slow convergence on the tetrabyd, despite possibly fast convergence on the cube as a whole. We will refer to convergence on the tetrabyd 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 tetrabyd, 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 tetrabyd 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 (3.18), $S_2 \gg S_1$ 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 tetrabyd. For the equilateral template, this factor becomes -300 . This illustrates why the tetrabyd-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.

3.6 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 (2.90), 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 (2.89) 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.

3.7 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^{ns} , 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, [128]. 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 (3.8) and (3.12), and the numerical evolution of ζ_k according to (2.6). 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

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 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 (3.12). This happens as the integrand of (3.12) 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 [16], at early times we extract the factor of $e^{-ik\tau_s}$ from the mode functions¹ 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, (3.12) and (3.8). 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 (3.8), 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 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 $[x_0, x_1, x_2, \dots, x_{N-1}]$. 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

¹In fact [16] extracts a factor of $e^{-ikc_s(\tau)\tau}$, losing efficiency due to slow-roll corrections.

map it to $[\bar{x}_{i-1}, \bar{x}_i, \bar{x}_{i+1}]$, where

$$\bar{x}_{i-1} = -1, \quad (4.1)$$

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

$$\bar{x}_{i+1} = 1. \quad (4.3)$$

We then calculate

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

for the p th 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

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

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.

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 (3.7). 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 (3.15) converges within our p_{\max} , we need only compute the first p_{\max} of the coefficients² in the expansion (3.7) of the $F^{(i)}(\tau, k)$, not until the actual sum (3.7) converges, since for high enough orders the integrals in (3.12) 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

²This is slightly complicated by the later inclusion of the spatial derivatives in (3.9) and (3.13), which mixes the modes—see later discussion in that section.

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. 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 (3.12) 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 (3.8), 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 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 (3.12) 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 [16], 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 plane wave expansion (e.g. [129]):

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

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 (2.21) expanded in Legendre polynomials (and similar) give us an extra suppression of $1/\tau^3$ in (3.12).

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 [91] 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 [92] 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 [91, 92], 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 [130]). 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

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

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.

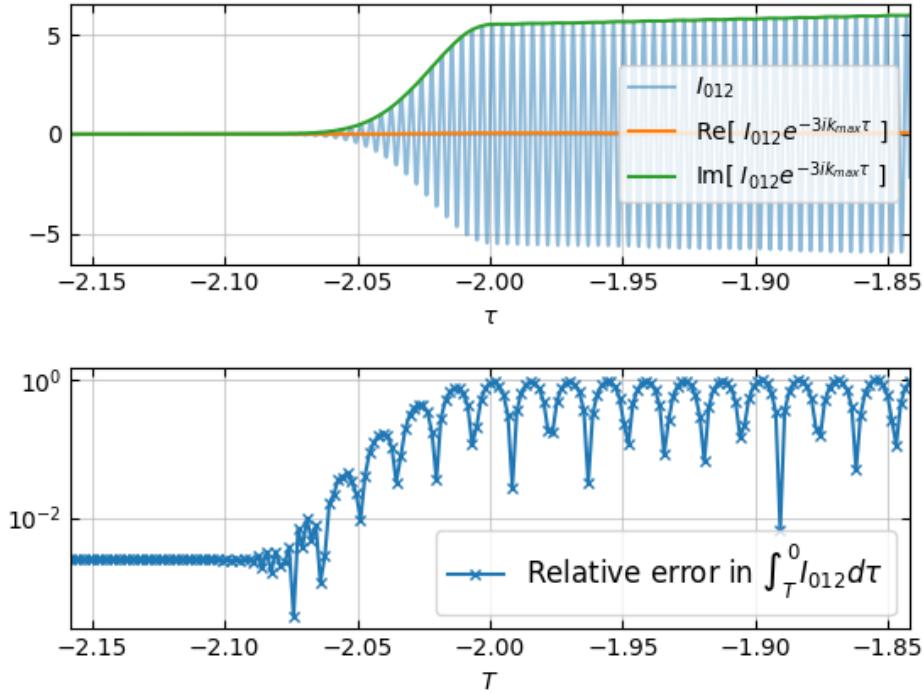


Figure 4.1: A toy example demonstrating the considerations involved in performing the time integrals (3.12). By carefully starting the time integrations, using the form (4.7), we can avoid errors that would otherwise swamp our result. The coefficient being calculated is the α_{012} coefficient of the \mathcal{P}_0 expansion of (2.70).

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 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 $H_{int} = (-1/\tau)\zeta^3$, which produces the shape (2.70).

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



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.

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 naive 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 (2.93) we have $c_s = 1$, so we expect (from [38]) the point of resonance τ_s^k for each mode k to be approximately given by $\tau_s^k = \frac{-|\phi'|}{2fk}$. For our example, we have $\phi' \approx -0.14$, $f = 0.02$, $k_{\min} = 2.088 \times 10^{-4}$ and $k_{\max} = 2.088 \times 10^{-1}$. This means that we expect the earliest resonance time to be $\tau_s \approx -\frac{0.14}{2(0.02)(2.088 \times 10^{-4})} \approx -1.7 \times 10^4$. Indeed, in figure 4.2 we see that we must have $k_{\max}\tau_s^{start} < -10^3$ to correctly capture the feature. **Double check this, reduce numbers.**

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 [7, 69, 72]

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

with $\Sigma = \frac{H^2\varepsilon}{c_s^2}$ and $\lambda = X^2 P_{,XX} + \frac{2}{3}X^3 P_{,XXX}$. See [7] for further details.

This is commonly quoted with a term proportional to the equation of motion, but this will never contribute [7, 94, 131, 132]. We do not need to make a slow-roll approximation (the quantities defined in (1.39) 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 [7] (see also [94, 131, 132]) 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

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = \left\langle \tilde{\zeta}_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_2} \tilde{\zeta}_{\mathbf{k}_3} \right\rangle + \lambda \left\langle \tilde{\zeta}_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_2} \right\rangle \left\langle \tilde{\zeta}_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_3} \right\rangle + cyclic \quad (4.9)$$

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 calculation. By integrating by parts and using the equation of motion, the interaction Hamiltonian can be rewritten without picking up boundary terms [87]. Using (3.7) from [87], with $f = -\varepsilon/(c_s^2 H)$, we obtain the following form:

$$H_{int}(t) = \int d^3x \left\{ -\frac{a^3 \varepsilon}{H c_s^4} \left(-c_s^2 - 2c_s^2 \frac{\lambda}{\Sigma} \right) \dot{\zeta}^3 + \frac{a^3 \varepsilon}{c_s^4} (-3c_s^2) \zeta \dot{\zeta}^2 - \frac{a \varepsilon}{c_s^2} (-c_s^2) \zeta (\partial \zeta)^2 - \frac{a \varepsilon}{H c_s^2} \dot{\zeta} (\partial \zeta)^2 - \frac{a^3 \varepsilon^2}{2c_s^4} (\varepsilon - 4) \dot{\zeta} \partial \zeta \partial (\partial^{-2} \dot{\zeta}) - \frac{a^3 \varepsilon^3}{4c_s^4} \partial^2 \zeta (\partial (\partial^{-2} \dot{\zeta}))^2 \right\}. \quad (4.10)$$

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 [16] 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 [16], 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 $V_{P\tilde{P}}^{(i)}$ defined in (3.13) 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 (3.18), 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 (3.18), 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 (2.87). 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

$$\frac{k_L}{k_S} = \frac{2k_{\min}}{k_{\max}}, \quad (4.11)$$

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 (2.27) and equilateral-type (2.57). 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 (3.18). We compare the converged results against analytic templates (2.88) and (2.89), using the full shape information (3.18), 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 (2.92). 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 [35], 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 (2.93). 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 [97], 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 (3.20). 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



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 in the bottom plot. Horizontal lines on this plot have constant k_3 . The bottom edge is $k_3 = k_{\min}$, the top edge is $k_3 = k_{\max}$. The right edge is $k_1 = k_2 = k_{\max}$, the left edge is $k_1 = k_2 = k_3/2$, i.e. the limit imposed by the triangle condition. Plotted in red (in the bottom plot) from top-left to bottom-right, are the flattened, equilateral and squeezed limits. For comparison, half of the tetrapyd is shown in the three-dimensional plot above (which was made by WHOM?).

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 (2.27). 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 (2.88) as a precision test. Indeed, we find that it has a scaled relative difference (3.19) of 2.7×10^{-5} with this shape, contrasting a scaled relative difference of 0.077 with the local template (2.49). 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 (2.88) or the local template (2.49), as for $\varepsilon \approx 0.8 \times 10^{-2}$ 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 (2.88) and (2.49), corresponding (in the sense of (3.19)) 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 (3.12), 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 (2.92), 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 (2.56) and (2.57). 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 (2.89), and 13% relative to the equilateral template (2.52). 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 (2.90), 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 (2.92) 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 [35], presented there in equations (48), (54), (55). We plot a comparison of the result of [35] 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 (2.57) 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 $p_{max} = 45$ and $p_{max} = 15$. The lower plot shows a DBI model that saturates the *Planck* limit on c_s . We set λ_{DBI} in (2.57) 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 $p_{max} = 45$ and $p_{max} = 15$.

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 [16]. 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 (2.93). 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 (2.92), 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 [35], $\mathcal{C} = 6c/(\varepsilon + 3c)$, 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 [35]. 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 (3.18), between the modal result and the analytic template of [35], as a function of the template parameter $C = \frac{6c}{\varepsilon_0 + 3c}$ (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 (2.92) 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 [93, 97]. 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 (3.12) and (3.8) 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 (2.93), 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 $p_{\max} = 65$ and $p_{\max} = 35$.



Figure 4.10: Non-Gaussianity generated by periodic features in a DBI model, including a phase difference in the flattened limit as described in [97]. 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 [16]. 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 the same constraint obtained in the *Planck* analysis [17], with the aim of validating the full PRIMODAL-CMB-BEST pipeline. The *Planck* result was obtained by constraining the amplitude of the DBI template (2.89), and using a slow-roll relation (2.46) 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, instead found by numerical 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} .

Using the scenario coefficients generated by PRIMODAL, the CMB-BEST code can use the *Planck* temperature data to place a constraint on β_{IR} , in the context of the scan described in 5.2. As shown in figure 5.1, we find

$$\beta_{IR} \leq 0.5 \quad (95\%, T \text{ only}). \quad (5.1)$$

This translates onto a constraint on the sound speed at horizon crossing of

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

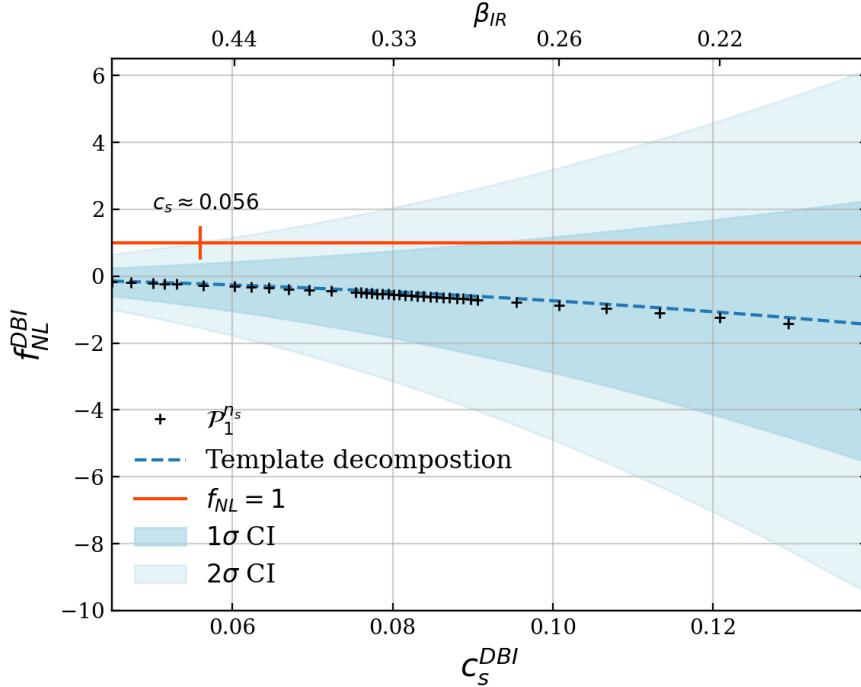


Figure 5.1: Our constraint on DBI inflation. Here c_s^{DBI} is not an input parameter (unlike in the template case), instead it is time dependent, and the plotted value is taken from the horizon crossing of the pivot scale. We find that $\beta_{IR} < 0.5$ is outside of our 2σ confidence interval. This 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 $f_{NL}^{DBI} = 1$. We can see that the CMB-BEST result does not agree with the *Modal* result (5.3), but as we see in figure 5.2, the discrepancy is not unexpected. We discuss possible explanations in the text. Note that the β_{IR} is approximately proportional to $(c_s^*)^{-2}$.

Equation (55) in [17] presents the constraint on the sound speed during DBI inflation derived from the CMB bispectrum using the *Modal* estimator

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

using the definition (2.46). While our goal is to validate our pipeline by reproducing this constraint, we would not expect to reproduce it exactly. This is due to numerous effects, including **Lack of maps? Different approximations? Numerics?**, as discussed in detail in [2].

We can however understand the expected variance of such estimators by simply examining the scatter between the estimators used in the *Planck* analysis.

Not that this is not statistical variance, but simply method variance. We show this in figure 5.2 for the equilateral template (2.52) which is closely related to the DBI template. The result obtained by CMB-BEST is different, but not surprisingly different. One possible reason for this is the fact we are using slightly different scales—since we are limited to $k_{\max}/k_{\min} = 1000$ for convergence reasons, we cannot use all the scales in the *Planck* data. Another possible reason is convergence in the number of maps—*Modal* uses $\tilde{300}$ maps while CMB-BEST uses 140. Numerical issues and implementation error could also contribute. See [2] for an in-depth discussion of these possibilities.

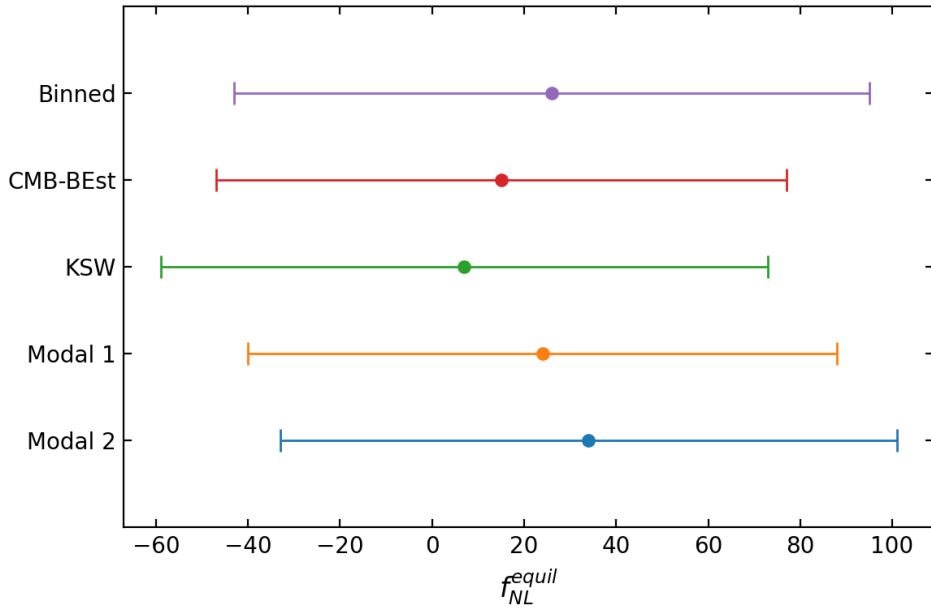


Figure 5.2: A comparison of the various estimation methods used in *Planck*, as presented in table 5 of [17], with the CMB-BEST result, presented in [2]. We plot the constraints obtained by each method for f_{NL}^{equil} , along with their 68% uncertainty margin. Note that all the estimation methods use the same data and maps, and are near optimal, 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), implementation error, and numerical stability. 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 (2.52) (using definition (2.43) for f_{NL}^{equil}) in the $\mathcal{P}_1^{n_s}$ basis for $p_{\max} = 30$. Which is the abstract one in *Planck*?

5.2 Connecting to CMB-BEst

The potential we use is the IR one discussed in [77] as discussed [earlier](#). See also [74, 78] for useful discussions. [Reasons](#).

We use equation (2.57), which has four free parameters, β_{IR} , λ_{DBI} , ϕ_0 , and V_0 . We use β_{IR} to parameterise our scan, $\beta_{IR} \in [0.1885, 0.58]$. We hold $\lambda_{DBI} = 2.00475 \times 10^{15}$, $V_0 = 5.2 \times 10^{-12}$ and $\phi_0 = 0.46042$ fixed. We start the background evolution on the slow-roll attractor, finding initial conditions which satisfy the Friedman constraint, as discussed in section 2.5.2. 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 .

We find that $\beta_{IR} = 0.331$ produces $c_s^* = 0.0794$, which is close to the *Planck* constraint (5.3). Thus, we roughly expect to find that $\beta_{IR} < 0.331$ is ruled out (for all the other scenario parameters held fixed at the above values).

In table 5.1 we summarise the slow-roll parameters for the scan, in table 5.2 we summarise the background parameters, and in table 5.3 we summarise the resulting scaling indices.

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

Table 5.1: Summary of slow-roll parameters across the scan.

β_{IR}	ϕ^*	H^*
1.89×10^{-1}	5.19×10^{-1}	1.31×10^{-6}
5.80×10^{-1}	5.15×10^{-1}	1.30×10^{-6}

Table 5.2: Summary of scenario parameters across the scan.

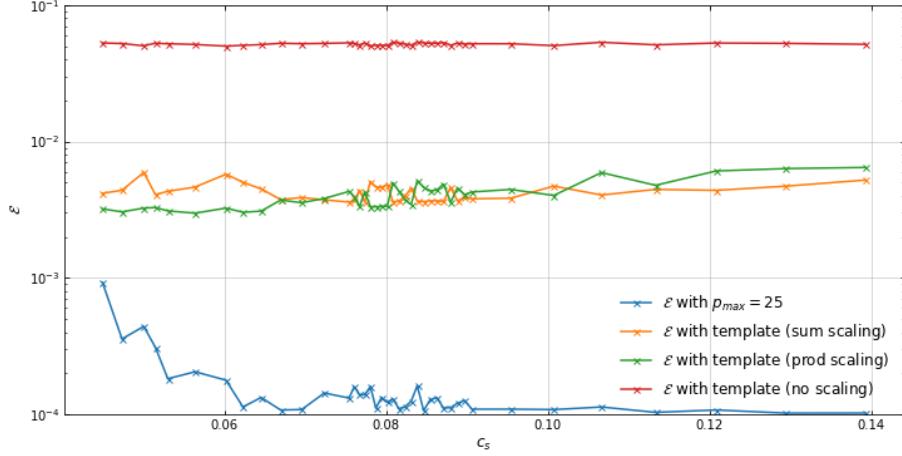


Figure 5.3: The *scaling* basis converges well across the scan range. We see that the bare DBI template is a poor match to the true numerical result. This is mostly due to the error in the overall magnitude. Once this is corrected, we see that the numerical result matches the approximate template to 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 (2.91) and the product scaling (2.90) perform comparably in matching the numerical result. This is mostly due to those templates neglecting the usual slow-roll suppressed contributions (as in (2.88)), which do in fact become relevant to the primordial bispectrum deep enough into the squeezed limit, due to their local-type shape.

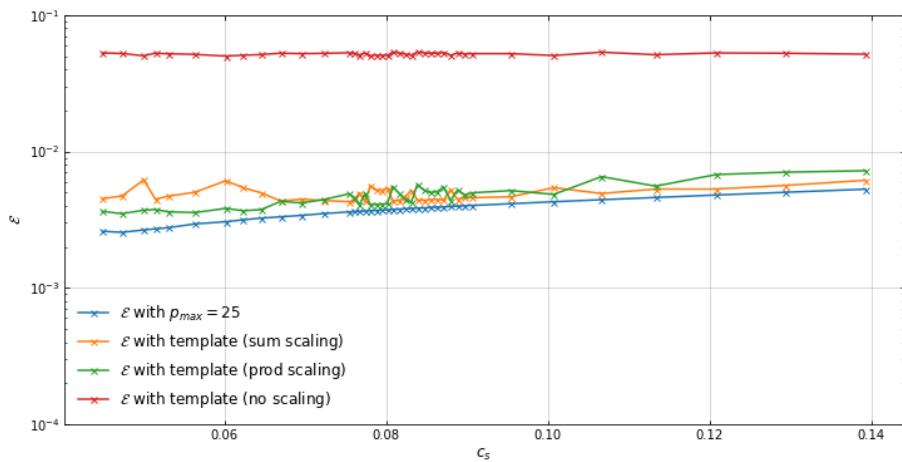


Figure 5.4: The \mathcal{P}_1^{ns} basis is sufficiently convergent across the scan range to obtain the desired constraint. We see that the convergence error is only slightly better than the error in the slow-roll corrected templates.

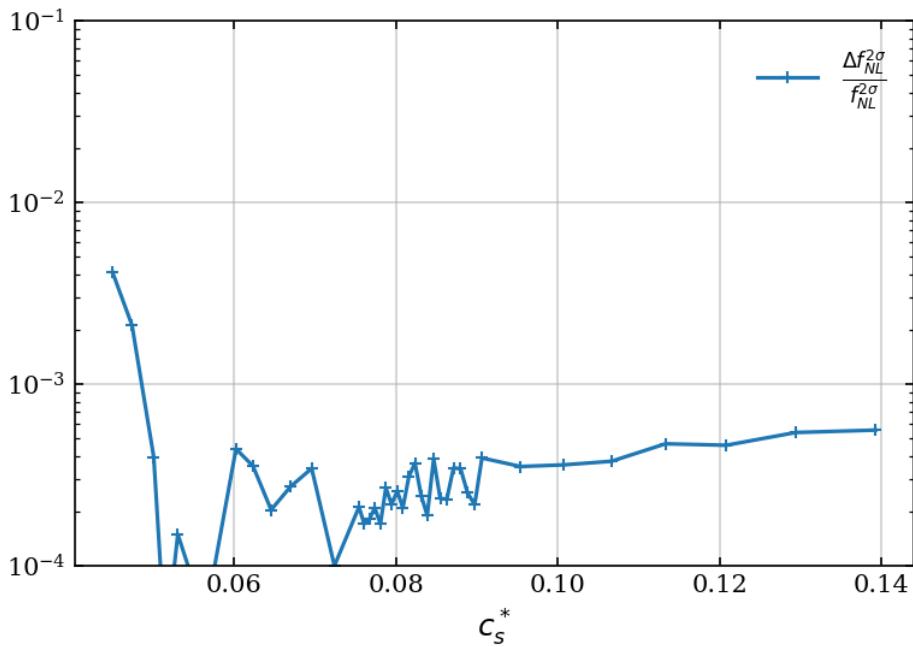


Figure 5.5: We plot the relative error in the value of $f_{NL}^{2\sigma} = f_{NL} + 2\sigma$, between $p_{\max} = 30$ and $p_{\max} = 25$, with f_{NL} obtained from *Planck* map for each scenario in the scan. We see that the convergence across the majority of the scan is better than that of the convergence in the primordial bispectrum, as plotted in figure 5.5. This validates the p_{\max} convergence of our pipeline as a whole.

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 5.3: Summary of scaling indices across 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.

Comparing primordial convergence to CMB convergence, we typically find that the CMB convergence is slightly better. This is a result of the fact that the squeezed limit (which is where the primordial shape converges most slowly) is suppressed by **by what?** At the CMB level, the convergence is $O(10^{-5})$, found by comparing the $p_{\max} = 29$ and $p_{\max} = 30$ results.

In table 5.4 we show convergence results for $\mathcal{P}_1^{n_s}$ with $p_{\max} = 30$. For this basis the match to the template is good in the equilateral limit, but quite poor in the squeezed limit.

β_{IR}	Sum Template	Product Template	Bare Template	With $p_{\max} = 25$
1.89×10^{-1}	6.15×10^{-3}	7.22×10^{-3}	5.16×10^{-2}	5.3×10^{-3}
5.80×10^{-1}	4.51×10^{-3}	3.63×10^{-3}	5.28×10^{-2}	2.6×10^{-3}

Table 5.4: Comparing the $\mathcal{P}_1^{n_s}(p_{\max} = 30)$ result compared to templates (2.89), (2.91) and (2.90). We also refit the result with $\mathcal{P}_1^{n_s}(p_{\max} = 25)$, and compare with the full result to estimate the convergence at the primordial level.

The convergence in the *scaling* basis is better, falling in the range $[1.02 \times 10^{-4}, 9.24 \times 10^{-4}]$. However, for this analysis the CMB-BEST code had only been run for the $\mathcal{P}_1^{n_s}$ basis. We see that it is sufficient in any case. When we examine the convergence to the sum (2.91) and product (2.90) 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 (2.88)) which are neglected in the DBI templates.

5.4 Slow-roll effects

The main slow-roll corrections are a correction to the amplitude, a deviation from perfect scale-dependence, and a local-type contribution in the squeezed limit which eventually comes to dominate the comoving curvature primordial bispectrum for sufficiently squeezed triangles. This is obvious because of that cancellation? From these results we can quantitatively see that slow-roll suppressed effects, despite becoming dominant in the squeezed limit, do not appreciably affect the constraint on the DBI scenario. We can see this since the template decompositions do not include these effects, but the numerical PRIMODAL results do, and both result in the same constraint.

5.5 Conclusions

We conclude that this result validates our pipeline as a whole, laying a solid foundation to move forward to shapes that have no standard template, obtaining new constraints.

Chapter 6

Conclusions and Future Work

6.1 Summary

In my PhD, I have exploited the inherent separability of the tree level in-in formalism, using expansions in separable basis functions. Following the modal philosophy of [66], and building on the basic idea of [16], 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 theory being considered—the output is not a grid of points however, but a set of coefficients of an explicitly separable basis expansion. This preserves the separability and therefore allows us to skip the usual template-approximation step.

I developed this separable approach into a practical and efficient numerical methodology which can be applied to a wider and more complicated range of bispectrum phenomenology, making an important step forward towards observational pipelines which can directly confront specific models of inflation using the full bispectrum information in a template-free analysis.

The latter part of this bispectrum estimation pipeline (CMB-BEST [2]) was recently completed by my collaborators (Wuhyun Sohn, Dr James Fergusson). This code constrains the inflationary output of my “PRIMODAL” code using *CMB* data.

In my recent paper (with E. P. S. Shellard) [1], which has been submitted to *JCAP*, we explored and described various basis choices and their advantages, and thoroughly validated the methods on single-field inflation models with non-trivial phenomenology, showing that our calculation of these coefficients is fast and accurate to high orders.

More recently I have reached the key exploitation phase of my research. My collaborators and I have tested the integration of the PRIMODAL and CMB-BEST parts of the pipeline, and at present, my focus is on running PRIMODAL on the Cambridge HPC facility, scanning models of inflation with the goal of obtaining some improved constraints on parameters of fundamental physics using our pipeline. Specifically, we are looking at models of DBI inflation [21] with sharp or oscillatory features, which are reviewed in, for example, [109].

6.2 Discussion

Link back to bigger picture. In this work we have extended the modal methods of [64–66] to recast the calculation of the tree-level primordial bispectrum (2.69) 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, 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 (2.37), the integral (2.97) 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. Making explicit the k -dependence in this way also opens the door to vast increases in efficiency in connecting to other observables, by precomputation using the basis set, then performing a (relatively) cheap scan over inflation parameters.

Our work here goes beyond that of [16] in that our careful methodology allows us to accurately and efficiently go to much higher orders, in particular our methods of starting the time integrals (3.12) and of including the spatial derivative terms in the calculation. This allowed us to present this method for feature bispectra for the first time, demonstrating the efficient exploration of much more general primordial bispectrum phenomenology. We have 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 Legendre polynomial basis sets.

The immediate application of this work is 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 will be extending these methods to the case of multiple-field inflation.

The next immediate application will be to directly constrain parameters of inflationary scenarios through modal bispectrum results from the *Planck* satellite [17]. The details of the work required to directly connect our coefficients to the observed data, and the large but once-per-basis cost of this calculation, will be detailed in a forthcoming paper [2]. CMB and LSS data from forthcoming surveys will be able to use these separable primordial bispectra to even more precisely constrain the parameters of inflationary scenarios.

These goals support those laid out in the recent community white paper [108], by improving predictions of the phenomenology of non-linearities in the very early universe. The ability to efficiently compare the phenomenology of these scenarios to observation and thus constrain the space of theories is a very worthwhile goal, though the phenomenology is also of interest in its own right.

[Move to interactions?](#)

6.3 Future Work

[Extension to non-Bunch-Davies effects \[133\], more Agocs papers?](#)

Non-linearities from inflaton self-interactions

If the fundamental physical model of inflation is of a single field, then the non-linearities of the self-interactions of that field may have produced observably large non-Gaussian signatures [33, 134]. Detecting these signatures in the CMB bispectrum requires an understanding of their form [18], 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 [17], include string theory-inspired models such as DBI inflation [79] and axion-monodromy [135, 136] (which gives large non-Gaussianity through the

resonance mechanism). Also within reach with the current implementation of PRIMODAL is the exploration of the effects of sudden features in the inflationary dynamics, and of more general time-dependence within the formalism of the effective field theory (EFT) of inflation.

Using the basis sets and the separable formulation of the in-in calculation developed in my PhD, We will reproduce and improve upon the established constraints on single-field inflation models. An initial target will be exploring how the scale-dependence of the numerically calculated DBI bispectrum affects the constraint on the sound speed, which was found in [17] using a scale-invariant approximate template. We will also explore oscillatory models, obtaining constraints using the full shape dependence, complementing the searches done in [17] which again used approximate templates (which exclude, for example, drifting frequencies).

Beyond this, we will expand the focus beyond DBI inflation, resonance features and sharp features and fully exploit the current implementation of my methods and code by comprehensively reviewing the single-field inflation models popular in the literature. We will explore the effects of including the full shape information in a template-free analysis, improving bispectrum constraints on parameters of single-field inflation models. Categorising models into distinguishability classes would be a useful initial goal.

Non-linearities from multi-field inflation

Another source of possibly observable non-Gaussian signatures can be found in inflation scenarios with more than one active field. In these models, which are motivated by string theory considerations [24], 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 [89]. The extra degrees of freedom in these scenarios can generate bispectrum shapes not usually seen in single-field models (for example, see recent work in [123, 137]), making the exploration of the case of multiple fields a prime target.

Our methods should extend well to multi-field. While some aspects of multi-field scenarios have been previously explored numerically [123] using the transport method [121] we can go beyond these explorations in two ways. Firstly, the transport method is a point-by-point method of calculating the bispectrum, and is thus significantly less efficient than our modal expansion method, and less suited to parameter scans. Secondly, the resulting grid of points is not easily

comparable to observations, which is the main motivator of the separability of our method.

I will extend the methods I have already developed to the case of multiple active fields during inflation, and implement them in the PRIMODAL code. I will validate my established basis sets on standard templates for multi-field bispectra, and test their ability to capture the expected scale-dependence of the shapes, a property usually excluded in standard templates. Once these validations are passed, I will recast the multi-field in-in calculation into a separable form (as I have already done for the single-field case) and implement that calculation in code. Using that implementation, I will explore the phenomenology of multi-field inflation, understanding which theories and scenarios can be constrained through the bispectrum, and obtaining those constraints.

Optimising the separable decomposition of primordial bispectra

The convergence of our separable expansion is a vital issue in determining the feasibility of our methods. This convergence depends centrally on the basis choice, as we explored in [1]. While in that work we described basis sets that could efficiently capture both the basic shapes and some more complicated oscillatory shapes, the question remains as to how to find the optimal separable description of a given bispectrum.

This is complementary to the philosophy of the main pipeline, which is to find a large general basis which covers a wide range of models. Instead, it may be possible to reduce the separable description of any given bispectrum to a sufficiently compact form that template-free estimation tailored to a scenario could be computationally tractable.

Additionally, the most relevant metric for convergence is the convergence of the observables, not the bispectrum at the end of inflation—exploring the relative efficiencies of different basis sets in this context could be very fruitful in expanding the range of models which could be constrained.

To pursue efficient separable decomposition, I intend to research techniques which have not previously been exploited in the setting of bispectrum calculation and estimation. One possible avenue is tensor rank decomposition techniques and other methods of tensor approximation, which are of use in, for example, signal processing.

By using these techniques to approximate the coefficient matrix (which is the output of PRIMODAL) we can determine a linear combination of our basis

functions which can more efficiently describe that particular bispectrum. I will apply these techniques to explore the general limits of separable approximations to primordial bispectra, with a focus on convergence at the level of observations. This will make clear the widest possible set of models that can be constrained through the bispectrum.

Public release of Primodal

The public release of PRIMODAL is a clear and worthwhile deliverable of this project. While other publicly released software for the calculation of inflationary non-Gaussianities does exist, they are based on point-by-point methods, and are thus inefficient in their calculation and description of the smooth bispectrum. PRIMODAL, which calculates the same quantity in an intrinsically smooth way, and whose result is tailored to comparison with observations, would thus fill an important gap.

My intention is to follow best practices when preparing PRIMODAL for public release. I will document the code, prepare example test cases for validation by users on their local machines, and release the code within a container to package the code along with its dependencies, libraries, and system settings to ensure reproducability.

References

- [1] P. Clarke and E. P. S. Shellard, *Probing inflation with precision bispectra*, *JCAP* **08** (2021) 002 [2012.08546].
- [2] W. Sohn, J. R. Fergusson, P. Clarke and E. P. S. Shellard, *High-resolution CMB bispectrum estimator, (in preparation)* .
- [3] 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.
- [4] 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.
- [5] 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.
- [6] D. Baumann, *Primordial Cosmology*, *Proceedings of Science TASI2017* (2018) 009 [1807.03098].
- [7] 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].
- [8] E. A. Lim, *Advanced cosmology : Primordial non-gaussianities*, 2012.
<https://nms.kcl.ac.uk/eugene.lim/AdvCos/lecture2.pdf>.
- [9] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

- [10] 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](#)].
- [11] A. J. Christopherson and K. A. Malik, *The non-adiabatic pressure in general scalar field systems*, *Phys. Lett. B* **675** (2009) 159 [[0809.3518](#)].
- [12] W. Hu, *Generalized slow roll for noncanonical kinetic terms*, *Physical Review D* **84** (2011) 027303 [[1104.4500](#)].
- [13] 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](#)].
- [14] 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](#)].
- [15] S. Weinberg, *Cosmology*. 2008.
- [16] H. Funakoshi and S. Renaux-Petel, *A Modal Approach to the Numerical Calculation of Primordial non-Gaussianities*, *JCAP* **1302** (2013) 002 [[arxiv/1211.3086](#)].
- [17] PLANCK collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [[1905.05697](#)].
- [18] 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](#)].
- [19] S. Weinberg, *Quantum contributions to cosmological correlations*, *Physical Review D* **72** (2005) 043514 [[hep-th/0506236](#)].
- [20] J. Maldacena, *Non-gaussian features of primordial fluctuations in single field inflationary models*, *Journal of High Energy Physics* **5** (2003) 013 [[astro-ph/0210603](#)].
- [21] 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](#)].

- [22] C. T. Byrnes and K.-Y. Choi, *Review of local non-Gaussianity from multi-field inflation*, *Adv. Astron.* **2010** (2010) 724525 [[1002.3110](#)].
- [23] 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](#)].
- [24] A. Achúcarro and G. A. Palma, *The string swampland constraints require multi-field inflation*, *JCAP* **02** (2019) 041 [[1807.04390](#)].
- [25] 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](#)].
- [26] 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](#)].
- [27] 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](#)].
- [28] 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](#)].
- [29] 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](#)].
- [30] 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](#)].
- [31] 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](#)].

- [32] 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](#)].
- [33] 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](#)].
- [34] C. T. Byrnes, M. Cortês and A. R. Liddle, *Comprehensive analysis of the simplest curvaton model*, *Physical Review D* **90** (2014) 023523 [[1403.4591](#)].
- [35] 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](#)].
- [36] C. Dvorkin and W. Hu, *Generalized slow roll approximation for large power spectrum features*, *Physical Review D* **81** (2010) 023518 [[0910.2237](#)].
- [37] C. P. Novaes, M. Benetti and A. Bernui, *Primordial Non-Gaussianities of inflationary step-like models*, *arXiv e-prints* (2015) [[1507.01657](#)].
- [38] R. Flauger and E. Pajer, *Resonant Non-Gaussianity*, *JCAP* **1101** (2011) 017 [[1002.0833](#)].
- [39] E. Pajer and M. Peloso, *A review of Axion Inflation in the era of Planck*, *Class. Quant. Grav.* **30** (2013) 214002 [[1305.3557](#)].
- [40] P. D. Meerburg and E. Pajer, *Observational Constraints on Gauge Field Production in Axion Inflation*, *JCAP* **1302** (2013) 017 [[1203.6076](#)].
- [41] P. D. Meerburg, *Oscillations in the primordial bispectrum: Mode expansion*, *Physical Review D* **82** (2010) 063517 [[1006.2771](#)].
- [42] P. D. Meerburg, *Oscillations in the primordial bispectrum*, *J. Phys. Conf. Ser.* **259** (2010) 012049 [[1010.2234](#)].
- [43] 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](#)].

- [44] H. Peiris, R. Easter and R. Flauger, *Constraining Monodromy Inflation*, *JCAP* **1309** (2013) 018 [[1303.2616](#)].
- [45] R. Easter and R. Flauger, *Planck Constraints on Monodromy Inflation*, *JCAP* **1402** (2014) 037 [[1308.3736](#)].
- [46] G. Cabass, E. Pajer and F. Schmidt, *Imprints of Oscillatory Bispectra on Galaxy Clustering*, [1804.07295](#).
- [47] 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](#)].
- [48] 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](#)].
- [49] 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](#)].
- [50] 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](#)].
- [51] 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](#)].
- [52] D. Chowdhury, J. Martin, C. Ringeval and V. Vennin, *Assessing the scientific status of inflation after Planck*, *Phys. Rev. D* **100** (2019) 083537 [[1902.03951](#)].
- [53] C. Pattison, V. Vennin, H. Assadullahi and D. Wands, *The attractive behaviour of ultra-slow-roll inflation*, *Journal of Cosmology and Astro-Particle Physics* **2018** (2018) 048 [[1806.09553](#)].
- [54] K. Dimopoulos, *Ultra slow-roll inflation demystified*, *Physics Letters B* **775** (2017) 262 [[1707.05644](#)].

- [55] J. Martin, H. Motohashi and T. Suyama, *Ultra slow-roll inflation and the non-Gaussianity consistency relation*, *Physical Review D* **87** (2013) 023514 [1211.0083].
- [56] 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].
- [57] 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].
- [58] SIMONS OBSERVATORY collaboration, *The Simons Observatory: Science goals and forecasts*, *JCAP* **02** (2019) 056 [1808.07445].
- [59] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold et al., *Cmb-s4 science book, first edition*, 2016.
- [60] 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].
- [61] 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].
- [62] 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].
- [63] 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].
- [64] J. R. Fergusson and E. P. S. Shellard, *Primordial non-Gaussianity and the CMB bispectrum*, *Physical Review D* **76** (2007) 083523 [astro-ph/0612713].

- [65] 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](#)].
- [66] 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](#)].
- [67] D. H. Lyth and D. Wands, *Conserved cosmological perturbations*, *Phys. Rev. D* **68** (2003) 103515 [[astro-ph/0306498](#)].
- [68] J. Garriga and V. F. Mukhanov, *Perturbations in k-inflation*, *Physics Letters B* **458** (1999) 219 [[hep-th/9904176](#)].
- [69] 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](#)].
- [70] D. Baumann, D. Green and R. A. Porto, *B-modes and the Nature of Inflation*, *JCAP* **01** (2015) 016 [[1407.2621](#)].
- [71] PLANCK collaboration, *Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **571** (2014) A24 [[1303.5084](#)].
- [72] 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](#)].
- [73] 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](#)].
- [74] V. Miranda, W. Hu and P. Adshead, *Warp features in DBI inflation*, *Physical Review D* **86** (2012) 063529 [[1207.2186](#)].
- [75] 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](#)].
- [76] E. Pajer, *Building a Boostless Bootstrap for the Bispectrum*, *JCAP* **01** (2021) 023 [[2010.12818](#)].

- [77] 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](#)].
- [78] X. Chen, *Running non-Gaussianities in DBI inflation*, *Phys. Rev. D* **72** (2005) 123518 [[astro-ph/0507053](#)].
- [79] E. Silverstein and D. Tong, *Scalar speed limits and cosmology: Acceleration from D-cceleration*, *Physical Review D* **70** (2004) 103505 [[hep-th/0310221](#)].
- [80] 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](#)].
- [81] N. Bartolo, E. Komatsu, S. Matarrese and A. Riotto, *Non-Gaussianity from inflation: Theory and observations*, *Phys. Rept.* **402** (2004) 103 [[astro-ph/0406398](#)].
- [82] X. Chen, *Primordial Non-Gaussianities from Inflation Models*, *Adv. Astron.* **2010** (2010) 638979 [[1002.1416](#)].
- [83] D. Babich, P. Creminelli and M. Zaldarriaga, *The Shape of non-Gaussianities*, *JCAP* **08** (2004) 009 [[astro-ph/0405356](#)].
- [84] D. Baumann and D. Green, *Equilateral Non-Gaussianity and New Physics on the Horizon*, *JCAP* **09** (2011) 014 [[1102.5343](#)].
- [85] S. Renaux-Petel, *Primordial non-Gaussianities after Planck 2015: an introductory review*, *Comptes Rendus Physique* **16** (2015) 969 [[1508.06740](#)].
- [86] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz and X. Chen, *Prospects for Cosmological Collider Physics*, *JCAP* **1703** (2017) 050 [[1610.06559](#)].
- [87] 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](#)].
- [88] G. Cabass, E. Pajer and F. Schmidt, *How Gaussian can our Universe be?*, *JCAP* **1701** (2017) 003 [[1612.00033](#)].

- [89] 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](#)].
- [90] 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](#)].
- [91] 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](#)].
- [92] 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](#)].
- [93] 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](#)].
- [94] S. Passaglia and W. Hu, *Scalar bispectrum beyond slow-roll in the unified EFT of inflation*, *Physical Review D* **98** (2018) 023526 [[1804.07741](#)].
- [95] P. Adshead, W. Hu and V. Miranda, *Bispectrum in single-field inflation beyond slow-roll*, *Physical Review D* **88** (2013) 023507 [[1303.7004](#)].
- [96] V. Miranda and W. Hu, *Inflationary steps in the Planck data*, *Physical Review D* **89** (2014) 083529 [[1312.0946](#)].
- [97] X. Chen, *Folded resonant non-Gaussianity in general single field inflation*, *Journal of Cosmology and Astroparticle Physics* **12** (2010) 003 [[1008.2485](#)].
- [98] 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](#)].
- [99] S. Garcia-Saenz, S. Renaux-Petel and J. Ronayne, *Primordial fluctuations and non-Gaussianities in sidetracked inflation*, *JCAP* **07** (2018) 057 [[1804.11279](#)].

- [100] 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](#)].
- [101] P. D. Meerburg and M. Münchmeyer, *Optimal CMB estimators for bispectra from excited states*, *Phys. Rev.* **D92** (2015) 063527 [[1505.05882](#)].
- [102] J. Adams, B. Cresswell and R. Easther, *Inflationary perturbations from a potential with a step*, *Physical Review D* **64** (2001) 123514 [[astro-ph/0102236](#)].
- [103] 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](#)].
- [104] J. Byun and R. Bean, *Non-Gaussian Shape Recognition*, *JCAP* **1309** (2013) 026 [[1303.3050](#)].
- [105] 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](#)].
- [106] T. Battfeld 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](#)].
- [107] W. Sohn and J. Fergusson, *CMB-S4 forecast on the primordial non-Gaussianity parameter of feature models*, *Phys. Rev. D* **100** (2019) 063536 [[1902.01142](#)].
- [108] P. D. Meerburg and et al, *Primordial Non-Gaussianity*, *arXiv e-prints* (2019) arXiv:1903.04409 [[1903.04409](#)].
- [109] A. Slosar, X. Chen, C. Dvorkin, D. Green, P. D. Meerburg, E. Silverstein et al., *Scratches from the Past: Inflationary Archaeology through Features in the Power Spectrum of Primordial Fluctuations*, *arXiv e-prints* (2019) arXiv:1903.09883 [[1903.09883](#)].
- [110] M. Ballardini, F. Finelli, R. Maartens and L. Moscardini, *Probing primordial features with next-generation photometric and radio surveys*, *JCAP* **1804** (2018) 044 [[1712.07425](#)].

- [111] S. Sypsas, *Cross-tests of cosmic microwave background features in the primordial spectra*, *Astron. Nachr.* **338** (2017) 1025 [[1711.09674](#)].
- [112] G. A. Palma, D. Sapone and S. Sypsas, *Constraints on inflation with LSS surveys: features in the primordial power spectrum*, *JCAP* **1806** (2018) 004 [[1710.02570](#)].
- [113] 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](#)].
- [114] J. R. Fergusson, *Efficient optimal non-Gaussian CMB estimators with polarisation*, *Phys. Rev. D* **90** (2014) 043533 [[1403.7949](#)].
- [115] C. Ringeval, *The exact numerical treatment of inflationary models*, *Lect. Notes Phys.* **738** (2008) 243 [[astro-ph/0703486](#)].
- [116] 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](#)].
- [117] J. S. Horner and C. R. Contaldi, *Non-Gaussianity Unleashed*, *ArXiv e-prints* (2013) [[1303.2119](#)].
- [118] 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](#)].
- [119] 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](#)].
- [120] 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](#)].
- [121] 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](#)].
- [122] 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](#)].

- [123] 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](#)].
- [124] K. Marzouk, A. Maraio and D. Seery, *Non-Gaussianity in D3-brane inflation*, [2105.03637](#).
- [125] 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](#)].
- [126] 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](#)].
- [127] 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](#)].
- [128] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, *SIAM Review* **51** (2009) 455.
- [129] A. Dixit, L. Jiu, V. H. Moll and C. Vignat, *The finite Fourier transform of classical polynomials*, *ArXiv e-prints* (2014) [[1402.5544](#)].
- [130] A. Iserles and S. P. Norsett, *Efficient quadrature of highly efficient integrals using derivatives*, *Royal Society Publishing, Volume 461, Issue 2057* (2005) .
- [131] 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](#)].
- [132] G. Rigopoulos, *Gauge invariance and non-Gaussianity in inflation*, *Physical Review D* **84** (2011) 021301 [[1104.0292](#)].
- [133] W. I. J. Haddadin and W. J. Handley, *Rapid numerical solutions for the Mukhanov-Sasaki equation*, *ArXiv e-prints* (2018) [[1809.11095](#)].
- [134] A. J. Tolley and M. Wyman, *The Gelaton Scenario: Equilateral non-Gaussianity from multi-field dynamics*, *Phys. Rev. D* **81** (2010) 043502 [[0910.1853](#)].

- [135] 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].
- [136] R. Flauger, L. McAllister, E. Silverstein and A. Westphal, *Drifting Oscillations in Axion Monodromy*, *JCAP* **1710** (2017) 055 [1412.1814].
- [137] 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].

List of Figures

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. We measure these quantities at the present, and extrapolate into the past and future using the Friedmann equation (1.11). 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 dark energy comes to dominate, $H(t)$ will become constant.	22
1.2	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.	22
1.3	During the radiation and matter dominated eras, the evolution of $a(t)$ has been slowing—this is decelerating expansion. However, as Λ comes to dominate, the \dot{a} will begin to increase, and the universe will enter an epoch of accelerated expansion.	23
1.4	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.	23

1.5 The evolution of the conformal time τ since the ΛCDM singularity, if there is no inflation. We see that without inflation there is finite conformal time in the past and that due to eventual Λ dominance eventually τ will asymptote to a constant, denoted here by the horizontal grey line at $\tau = \tau_0 + (H_0 a_0)^{-1}$. Thus, without inflation, there is a maximum scale on which we would expect the CMB to be correlated, and a maximum comoving distance that we can ever expect to receive CMB photons from.	24
3.1 Convergence comparisons for the Legendre and Fourier basis functions for (a) the Maldacena template (2.88) and (b) the DBI template (2.89). The pure Legendre \mathcal{P}_0 basis requires many terms to fit the $1/k$ behaviour in both Maldacena's template (2.88) and the DBI template (2.89). 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 $p_{\max} = 5$. The Fourier \mathcal{F}_1 basis performs well, but converges more slowly than the \mathcal{P}_1 basis. Note that the convergence errors for (2.89) are larger than (2.88) because of the larger contributions outside the tetrapyd dominating the fit. In this plot and the following, unless otherwise stated, $k_{\max} = 1000k_{\min}$	86
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 (2.94) (that is periodic in the k -range) perfectly. For this special case, the \mathcal{P}_0 , <i>scaling</i> and <i>resonant</i> 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 (2.89) and (2.94)), we see that again the <i>scaling</i> 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 (3.18) to the physical configurations on the tetrapyd; when considered over the entire cube the fit improves monotonically. The <i>resonant</i> basis, naturally, does not converge well to a linear oscillation.	87

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 (2.89) and (2.95). We choose $f = 4.55$ (so we obtain 5 whole oscillations in the k -range). (a) On the tetrapyd, we see that the <i>scaling</i> 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 <i>resonant</i> 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.	88
3.4 For the scale-dependent DBI template (2.90), 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 $p_{\max} = 30$. The <i>scaling</i> basis performs far better again, even without the extra power spectrum information. By far the fastest convergence, however, results from the <i>resonant</i> basis. The convergence of this template also has a dependence on k_{\max}/k_{\min} . The convergence power of these basis sets will allow us to efficiently capture the scale dependence of the numerically calculated shape function.	89
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 <i>scaling</i> 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 <i>scaling</i> 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.	90

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 <i>scaling</i> basis converge for a significantly larger frequency range than the <i>resonant</i> 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 <i>scaling</i> 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 <i>resonant</i> 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.	91
3.7 We plot the \mathcal{P}_0 , \mathcal{P}_1 and <i>scaling</i> 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.	95
3.8 We plot the \mathcal{F}_0 , \mathcal{F}_1 and <i>resonant</i> 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 <i>resonant</i> 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.	96
4.1 A toy example demonstrating the considerations involved in performing the time integrals (3.12). 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 (2.70).	107

- 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.

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 in the bottom plot. Horizontal lines on this plot have constant k_3 . The bottom edge is $k_3 = k_{\min}$, the top edge is $k_3 = k_{\max}$. The right edge is $k_1 = k_2 = k_{\max}$, the left edge is $k_1 = k_2 = k_3/2$, i.e. the limit imposed by the triangle condition. Plotted in red (in the bottom plot) from top-left to bottom-right, are the flattened, equilateral and squeezed limits. For comparison, half of the tetrapyd is shown in the three-dimensional plot above (which was made by WHOM?).

4.4 A canonical single-field model on a quadratic potential (2.92), 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 $p_{\max} = 15$

- 4.5 The upper plot shows the shape function for a DBI model deep in slow-roll. We set λ_{DBI} in (2.57) 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 $p_{\max} = 45$ and $p_{\max} = 15$. The lower plot shows a DBI model that saturates the *Planck* limit on c_s . We set λ_{DBI} in (2.57) 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 $p_{\max} = 45$ and $p_{\max} = 15$ 117
- 4.6 The tree-level shape function of a feature model (2.92), 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 [35], $\mathcal{C} = 6c/(\varepsilon + 3c)$, 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 $p_{\max} = 35$ 119
- 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 [35]. 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. 120

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 (3.18), between the modal result and the analytic template of [35], 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 (2.92) occurs, which is kept constant across this scan. The width of the feature was also kept constant.	121
4.9	Resonance on a quadratic potential (2.93), 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 $p_{\max} = 65$ and $p_{\max} = 35$	122
4.10	Non-Gaussianity generated by periodic features in a DBI model, including a phase difference in the flattened limit as described in [97]. 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 $p_{\max} = 35$	123

5.3	The <i>scaling</i> 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 <i>scaling</i> basis we can see that sum scaling (2.91) and the product scaling (2.90) perform comparably in matching the numerical result. This is mostly due to those templates neglecting the usual slow-roll suppressed contributions (as in (2.88)), which do in fact become relevant to the primordial bispectrum deep enough into the squeezed limit, due to their local-type shape.	129
5.4	The \mathcal{P}_1^{ns} 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.	129
5.5	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 <i>Planck</i> 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 5.5. This validates the p_{\max} convergence of our pipeline as a whole.	130

List of Tables

1.1	How the scale factor, $a(t)$, evolves in the different epochs of the universe.	18
2.1	Notation summary and comparison. Check N/A's. What about $X = \frac{1}{2}\dot{\phi}^2$ vs. $X = \dot{\phi}^2?$	50
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.	85
5.1	Summary of slow-roll parameters across the scan.	128
5.2	Summary of scenario parameters across the scan.	128
5.3	Summary of scaling indices across the scan.	131
5.4	Comparing the $\mathcal{P}_1^{n_s}(p_{\max} = 30)$ result compared to templates (2.89), (2.91) and (2.90). We also refit the result with $\mathcal{P}_1^{n_s}(p_{\max} = 25)$, and compare with the full result to estimate the convergence at the primordial level.	131

