UNIVERSITY OF CAMBRIDGE

Probing Inflation with Precision Bispectra

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 Preface 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 except as declared in the Preface and specified in the text. 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 except as declared in the Preface and specified in the text. This dissertation does not exceed the prescribed limit of 60 000 words.

Philip Clarke
July, 2021

Abstract

Probing Inflation with Precision Bispectra

Philip Clarke

This is just a placeholder outline, will be reduced to proper abstract size eventually. 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 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 directly confront specific models of inflation.

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 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.

Acknowledgements

My acknowledgements ...

Contents

List of Figures

List of Tables

Chapter 1

Introduction to cosmology

1.1 General introduction

1.1.1 Fundamentals

General relativity tells us how the expansion of the universe depends on its matter and energy content. In the context of a spatially homogeneous and isotropic universe (known as a FLRW universe) the Einstein equations become the Friedmann equations.

We begin with a homogeneous and isotropic metric with $c = 1$

$$ds^2 = a(t)^2 ds_3^2 - dt^2 \tag{1.1}$$

where $a(t)$ will have the interpretation as the scale factor which describes the evolution of the universe. The components of the universe enter the equations through their stress-energy tensor, which we approximate as that of a perfect fluid

$$T_{ab} = (\rho + P)U_a U_b + P g_{ab}. \tag{1.2}$$

The quantity $w = \frac{P}{\rho}$ is constant, and takes the values 0, 1/3 and -1 for matter, radiation and dark energy respectively. The Einstein equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{1.3}$$

then gives us the Friedmann equations for a flat universe with $\Lambda = 0$

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

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

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. We also have the continuity equation

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0. \tag{1.6}$$

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 $a(t)$. From this, one can understand how perturbations evolve, and understand how photons are redshifted as they freestream through the universe.

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

$$\frac{dP^a}{d\lambda} + \Gamma^a_{bc}P^bP^c = 0. \tag{1.7}$$

Using the chain rule, this can be rewritten as

$$P^0\frac{\partial P^a}{\partial t} + \Gamma^a_{bc}P^bP^c = 0. \tag{1.8}$$

REPHRASE ALL THIS Writing the components of the photon's 4-momentum P^a as $(p, \mathbf{p})$ we get that

$$p \propto \frac{1}{a}. \tag{1.9}$$

The interpretation of this result is that as the universe expands, the photon's energy decreases, and its wavelength increases. The geodesic equation encodes many other physical effects when applied to the evolution of perturbations around the homogeneous background, allowing us to understand how the statistics of the perturbations at the start of the ΛCDM evolution evolved into the perturbations we see today.

We can rewrite the Friedmann equation as

$$H^2 = H_0^2\left(\frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \Omega_\Lambda\right) \tag{1.10}$$

where we define the fractional density

$$\Omega_X = \frac{\rho_X}{\rho_{crit,0}} \tag{1.11}$$

for X being radiation, matter and dark energy, using $\rho_{crit,0}$, the critical density for which the universe is flat. Let us take the future of the universe as an example, where a is sufficiently large that Ω_Λ is the dominant contribution to the expansion. Then

$$H^2 = H_0^2 \Omega_\Lambda \tag{1.12}$$

$$\implies \dot{a} = \pm H_0 \sqrt{\Omega_\Lambda} a \tag{1.13}$$

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

$$a(t) = a_0 e^{H(t-t_0)}. \tag{1.14}$$

Then, using $a d\tau = dt$, we find that $\tau(t) = (Ha_0)^{-1}\left(1 - e^{-H(t-t_0)}\right) + \tau_0$ where as usual a 0 subscript denotes a quantity evaluated today. This is the behaviour we would also like in the very universe, as we will see.

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

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

Set-up the physical effects of transfer functions.

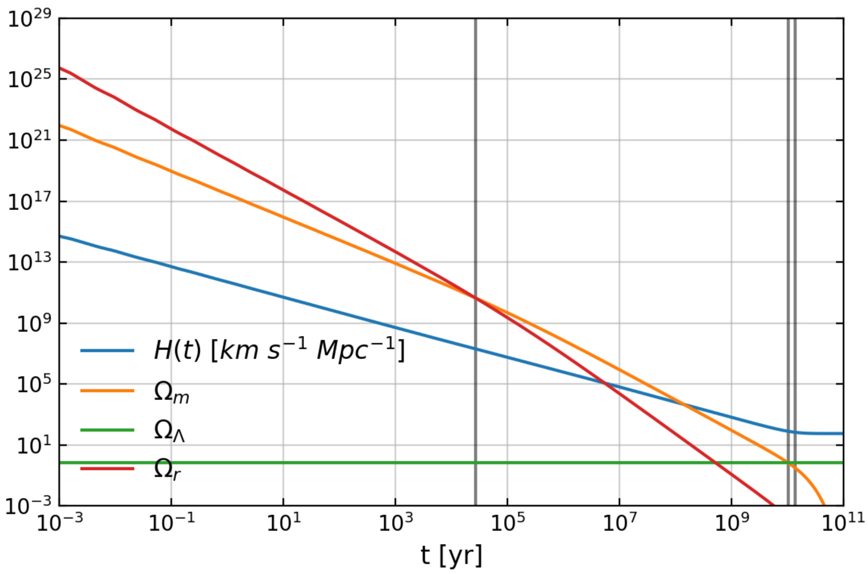


Figure 1.1: The evolution of the components of the universe up to the present, and slightly beyond. The vertical grey lines mark matter-radiation equality, matter-dark energy equality, and the present day, respectively. In the past, densities of matter and radiation were far higher, and the expansion of the universe ($H(t)$) was far stronger. In the future, as dark energy comes to dominate, $H(t)$ will become constant.

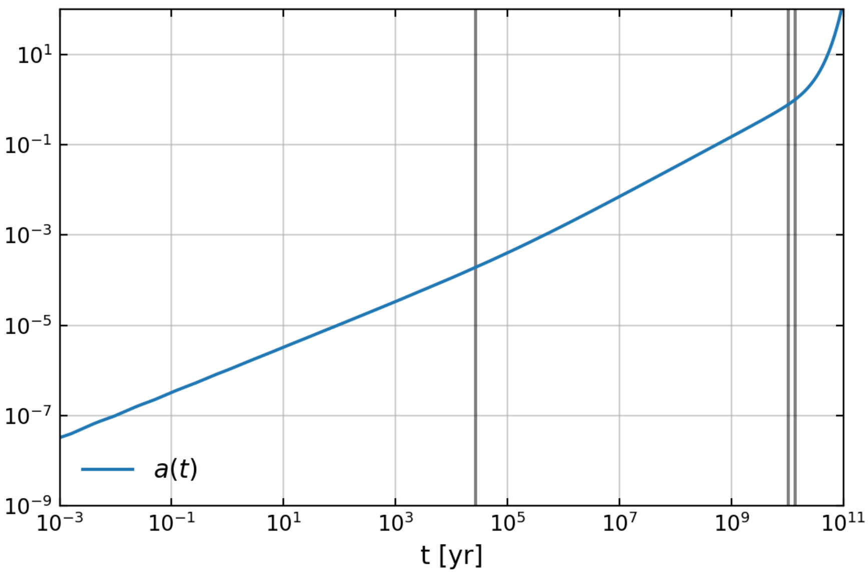


Figure 1.2: The evolution of the scale factor. For most of the ΛCDM history it evolves as some power of t, however as Λ comes to dominate it will begin to grow exponentially.

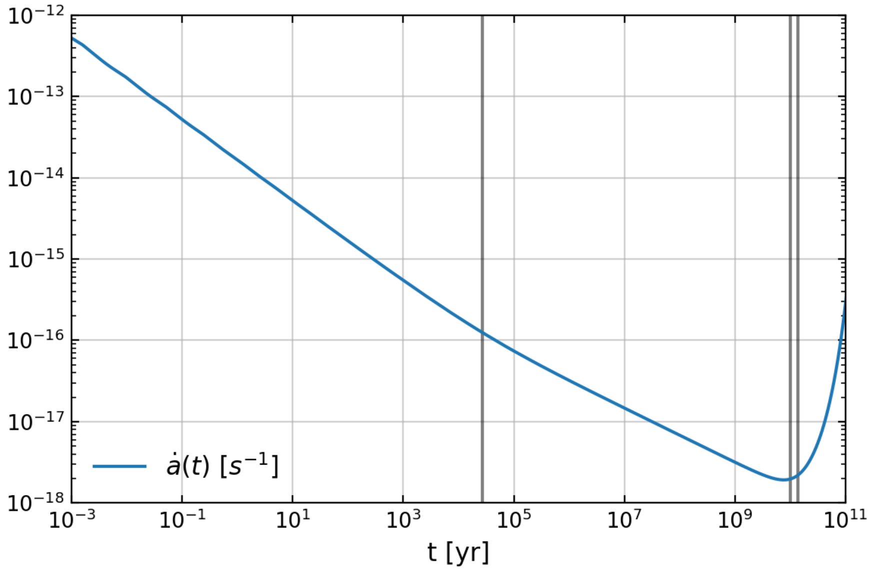


Figure 1.3: During the radiation and matter dominated eras, the evolution of $a(t)$ as been slowing. However, as Λ comes to dominate, the scale factor will begin to accelerate.

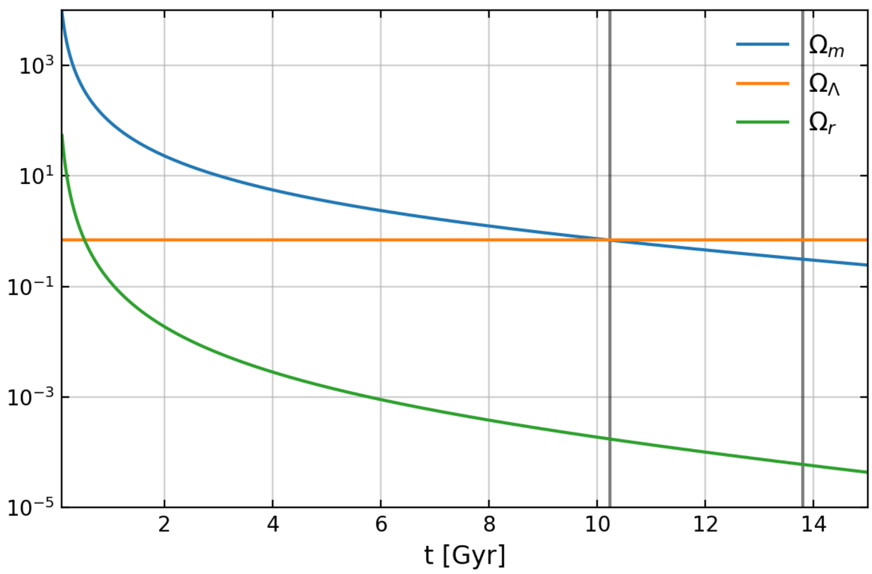


Figure 1.4: The evolution of the components of the universe, zoomed in to more clearly show the matter-dark energy transition.

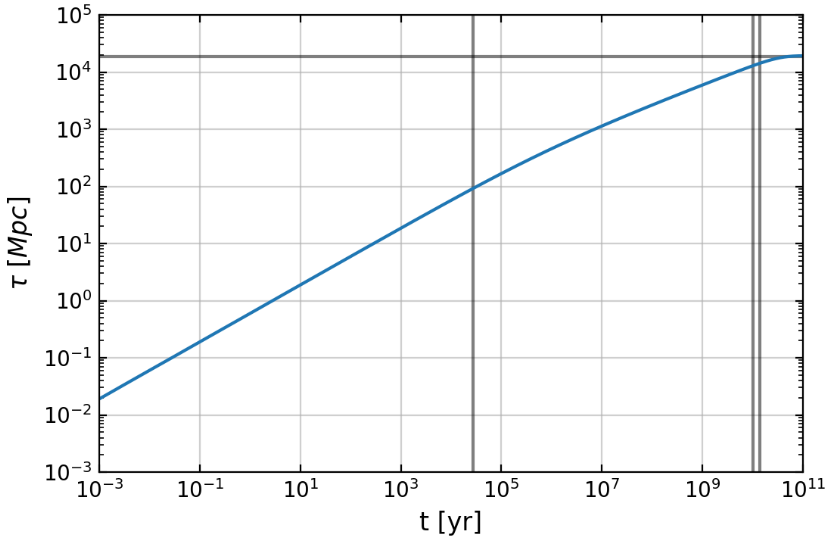


Figure 1.5: The evolution of the conformal time τ since the ΛCDM singularity, if there is no inflation. Eventually τ will asymptote to a constant. Physical interpretation at early and late times!!

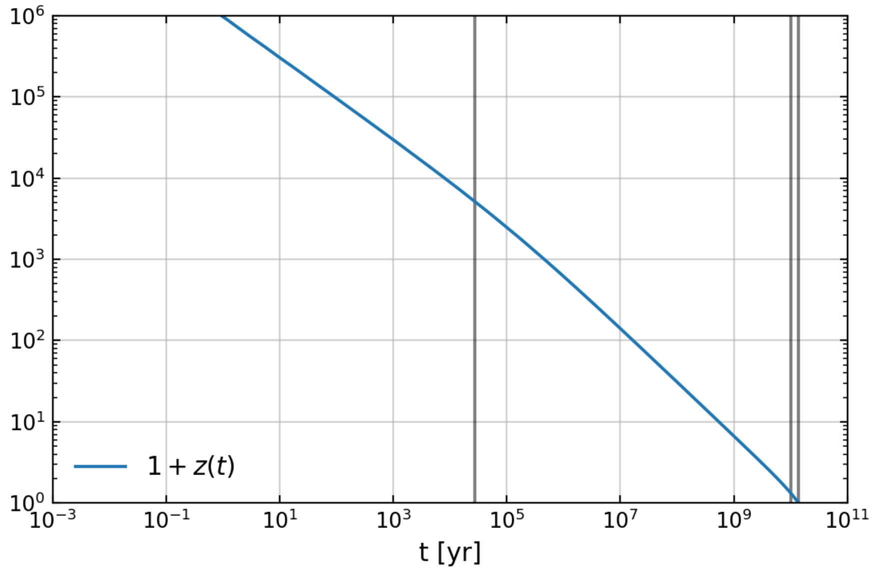


Figure 1.6: The correspondence between redshift z and t. Add temperature plot!!

The history of the universe as a story of falling temperature.

1.1.2 ΛCDM

Add references! Don't explain in detail, but do reference. 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.

Recombination, what the CMB is. Plot of $a(t)$, $\tau(t)$ over all epochs. Path of a photon in $t - x$? Plot of $E(t)$?

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.

What is the evidence for ΛCDM? Two main pieces of evidence are the relative abundances of primordial elements, and the baryon acoustic oscillations (BAOs).

While it is thought that the heavier elements are made in incredibly energetic supernovae and neutron star collisions (such as lead[ref], or gold[ref]) the lightest elements are instead thought to have been forged in the very first minutes of the universes history. The ΛCDM model makes a detailed prediction as to the relative abundances of these elements, hydrogen, helium and lithium. It works well, but there is the "lithium problem".

Baryon acoustic oscillations are the remnants of an epochal transition in the early universe. When the baryons decoupled from the radiation (due to the falling temperature) they were dropped, no longer supported by the photon pressure. They began to collapse back into the gravitational wells the dark matter had been forming. The imprint of this can be seen in the statistics of the positions of galaxies in the sky Did I use the zooming out thing in BlueSci? Can I use it here?? [ref two-point], we will discuss two-point statistics in section ref.

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.

1.2 Initial conditions for ΛCDM

1.2.1 Motivations for inflation

The ΛCDM model does not posit a "Big Bang singularity", though this is often how it is described in popular science communication. 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 of number. These components are distributed incredibly uniformly, with inhomogeneities of less than number. This simple scenario then evolves under gravitational collapse (and other interactions) and forms the universe we know.

Is this story of ΛCDM enough to explain all the features of the universe that we observe? The answer is mostly less, 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, these problems are the statements 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? Why is the universe so flat?

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 conformal distance a photon could have traveled between the singularity at $a_i = 0$ and recombination at $a_{rec} = 1/1100$. We find

$$\chi_{rec} = \int_0^{\tau_{rec}} d\tau = \int_0^{a_{rec}} \frac{d\tau}{da} da \propto a_{rec}. \tag{1.15}$$

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

$$\tan^{-1}\left(2\frac{\chi_{rec} - 0}{\chi_0 - \chi_{rec}}\right) \approx 2^\circ \tag{1.16}$$

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

$$\frac{4\pi(\chi_0 - \chi_{rec})^2}{\pi(\chi_{rec} - 0)^2} \approx 10^5. \tag{1.17}$$

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 and explanation.

Photons travelling on $45'$ lines on a $\tau - x$ plot. There not being enough time for opposite sides of the CMB to come into thermal equilibrium. But if you change the relation between t and τ then you extend the plot, and you do make contact. PLOT THIS!!

The flatness problem is the statement that our universe is much flatter that one would expect. The reason we would expect otherwise, is that deviations from flatness are unstable. 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

$$\frac{\delta\rho_i}{\bar{\rho}_i'} = \frac{\delta\rho_j}{\bar{\rho}_j'}. \tag{1.18}$$

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

$$\bar{\rho}_i(t) + \delta\rho_i(t,x) = \bar{\rho}_i(t) + \delta t(x)\bar{\rho}_i' = \bar{\rho}_i(t + \delta t(x)) \tag{1.19}$$

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. Scale invariance is the statement that

$$\langle f(\mathbf{x})f(\mathbf{x}')\rangle = \langle f(\lambda\mathbf{x})f(\lambda\mathbf{x}')\rangle . \tag{1.20}$$

From this, we can see that scale invariance needs $P(k) \propto k^{-3}$. It turns out that observations show that this is not precisely true. In fact, it was found that $P(k) \propto k^{n_s - 4}$, with $n_s = 0.9649 \pm 0.0042$.

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

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{1.21}$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi). \tag{1.22}$$

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

$$H^2 \approx \frac{8\pi G}{3} V(\phi), \tag{1.23}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(-2V(\phi)\right). \tag{1.24}$$

1.2.2 Criteria for successful inflation

Sufficient e-folds, matching the statistics of the primordial perturbations.

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

$$\frac{d}{dt}(aH)^{-1} = -\frac{\ddot{a}}{\dot{a}^2} < 0 \tag{1.25}$$

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

$$\frac{d}{dt}(aH)^{-1} = -\frac{1}{(aH)^2}\left(\dot{a}H + a\dot{H}\right) \tag{1.26}$$

$$= -\frac{1}{a}(1-\varepsilon) \tag{1.27}$$

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 $\varepsilon \sim O(10^{-2})$.

1.3 Statistical observables

1.3.1 Checking dice for fairness

The prediction of the fundamental quantum theory is a statistical one, i.e. a prediction of the distribution from which our observation will be drawn. As such, we need to talk in terms of estimators, estimating how likely it would be to see the sky we do, assuming some fundamental theory.

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 out hypothesis of fairness was incorrect. One can calculate that, for a fair coin, 95% of the time the events will fall within the range $[41, 59]$. 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,

$$\sigma = \sqrt{E\left[(X-\mu)^2\right]}. \tag{1.28}$$

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 point 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

$$\langle f(x) \rangle = \sum_i x_i P(x_i) \tag{1.29}$$

$$\langle f(x) \rangle = \int dx \; x \rho(x) \tag{1.30}$$

$$\langle F[f(x)] \rangle = \int \mathcal{D}f \; F[f(x)] \, P[f(x)] \tag{1.31}$$

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 (1.29) 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 (1.30). In this thesis, we will be working with the expectation value of field configurations, so we will use (1.31). Talk about quantum to classical transition!

$$\begin{aligned} \langle \hat{v}_{\mathbf{k}}, \hat{v}_{\mathbf{k}'} \rangle &= \langle 0 | \hat{v}_{\mathbf{k}}, \hat{v}_{\mathbf{k}'} | 0 \rangle & (1.32) \\ &= |v_k|^2 \left\langle 0 \middle| \left[\hat{a}_{\mathbf{k}}, \hat{a}^\dagger_{-\mathbf{k}} \right] \middle| 0 \right\rangle & (1.33) \\ &= |v_k|^2 \delta(\mathbf{k} + \mathbf{k}') & (1.34) \\ &= P_v(k) \delta(\mathbf{k} + \mathbf{k}') & (1.35) \end{aligned}$$

1.3.2 Power spectra

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 on the sky, $\langle \phi(x) \phi(x') \rangle$. When there is a large angular separation between x and x' talk about cosmic variance of different modes.

1.4 Observational data

1.4.1 *Planck*, Simons

High-level descriptions. What fraction of the sky did they look at? Did they include polarisation? What was their angular resolution?

1.4.2 Future missions

High-level descriptions. What fraction of the sky will they look at? Will they include polarisation? What will their angular resolution be?

1.5 Outline of thesis

Should some of this be moved to the end of chapter 2? Maybe a high-level outline here and then a technical outline at the end of chapter 2.

1.5.1 Goals

1. Connecting inflation models directly to observations, through the bispectrum.
2. Constraining the parameters of inflation models, not phenomenological templates and f_{NL}.
3. To obtain the full shape information, not point samples or a limit.
4. Efficient numerics gives access to more accurate, and in some cases new, feature shapes.

1.5.2 Methods

1. Building separability into the tree-level in-in formalism.
2. The CMB calculation [1]: expensive, but need only be done once per primordial basis.
3. So, we want a basis expansion that converges quickly for a broad range of inflation models.
4. Convergence on the cube is different to the tetrapyd.
5. Turns out to be much faster at primordial level than previous numerical methods (as it in a sense converges way faster, and as it enables us to use faster numerical methods than otherwise).

1.5.3 Results

1. First development/implementation of the formalism for calculating the expansion to high orders.
2. We recognised and described the central issue of the cube vs tetra problem.
3. Found a basis with broad descriptive power (and other less powerful basis sets).
4. This allowed the first validation of these methods on features.
5. Connect to CMB, get constraints on DBI 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 [2], and we discuss methods of testing numerical bispectrum results, defining our relative difference measurement. In chapter 4 we present our methods. 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 begin with this distinct and separate, but nonetheless vital, discussion. We discuss the effects of the non-physical k-configurations on the convergence of our expansion on the tetrapyd, and present an efficient basis. Then, we recast the usual in-in calculation into an explicitly separable form, in terms of an expansion in an arbitrary basis, and detail our methods for carefully calculating the coefficients to high order. In section 4.6 we validate our methods and implementation on inflation scenarios with varied features from the literature, and we finish with a discussion of future work in chapter 6.

Chapter 2

Probing inflation

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 [3], 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 [4, 5], interactions between multiple fields [6], sharp features [7] and periodic features [8]. 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 course-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 [9, 10].

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 [11, 12]. With this high-precision *Planck* data, and data from forthcoming experiments such as the Simons Observatory (SO) [13] and CMB-S4 [14], 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 [15, 16, 17, 18] 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 [19, 20, 21] 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 [2], and the details of the bispectrum estimation part will be detailed in [1]. We go beyond the work of [2] 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...

The sound speed c_s. What is the intuition for c_s? What does a photon path look like on a $\tau_s - x$ plot?

2.1.1 Initial conditions for the perturbations

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

2.1.2 Evolution of the perturbations

Derivation of the Mukhanov-Sasaki equation, behaviour at various times. Second order action, in comoving gauge. Expanding the operator in Fourier modes. Quantization. In conformal time, we obtain

$$\frac{\partial^2 u_k}{\partial \eta^2} + \left(c_s^2 k^2 - \frac{1}{z} \frac{d^2 z}{d\eta^2} \right) u_k = 0 \tag{2.1}$$

where $z^2 = 2a^2\varepsilon / c_s^2$ and $u = z\mathcal{R}$. We can see that at early timesℓ.

The Mukhanov-Sasaki equation in N, τ_s—no numerics here, though, just ref forward.

2.1.3 Connecting the primordial power spectrum to observables

Define the CMB power spectrum, discuss constraints.

2.2 The primordial bispectrum

2.2.1 The shape function

Shape versus scale dependence. Consequences of isotropy, homogeneity. Define the shape function. The primordial bispectrum is usually written as:

$$\langle \zeta_{\mathbf{k_1}} \zeta_{\mathbf{k_2}} \zeta_{\mathbf{k_3}} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}) B(k_1, k_2, k_3) \tag{2.2}$$

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:

$$S(k_1, k_2, k_3) = (k_1 k_2 k_3)^2 B(k_1, k_2, k_3). \tag{2.3}$$

The bispectrum is only defined where the triangle condition

$$\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3} = 0, \tag{2.4}$$

is satisfied, which implies that the triangle inequality must hold

$$k_1 + k_2 \geq k_3 \text{ and cyclic perms.} \tag{2.5}$$

The space of configurations we are interested in is therefore reduced from the full cube $[k_{\min}, k_{\max}]^3$ to a tetrapyd (illustrated in figure 4.3), the intersection of that cube with the tetrahedron that satisfies (2.5). 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:

$$B^F(k_1, k_2, k_3) = f_{NL}^F \times F(k_1, k_2, k_3) \tag{2.6}$$

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 (2.31)), equilateral (see (2.34)) and orthogonal templates, as defined in [11].

2.2.2 The definitions of f_{NL}

Describe the various definitions of f_{NL} in different parts of the literature.

2.2.3 Non-linearity during inflation

Connect (self-)interactions during inflation to non-Gaussianity.

2.3 Connecting the primordial bispectrum to observables

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

$$C_l \approx 4\pi \int d\ln k\ \mathcal{P}_\mathcal{R}(k)\left[\frac{(\Theta_0+\psi)(\eta_*,\mathbf{k})}{\mathcal{R}(\mathbf{k})}j_l(k\chi_*) - \frac{v_\gamma(\eta_*,\mathbf{k})}{\mathcal{R}(\mathbf{k})}j_l'(k\chi_*)\right]^2. \tag{2.7}$$

2.4 Calculating the primordial bispectrum from an inflation scenario

(Coming from part iii level, a derivation of the tree-level in-in formalism.) The standard starting point for calculating higher-order correlators for models of inflation is the in-in formalism [3, 22]. 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:

$$\langle\zeta_{\mathbf{k_1}}(\tau)\zeta_{\mathbf{k_2}}(\tau)\zeta_{\mathbf{k_3}}(\tau)\rangle = -i\int_{-\infty(1-i\varepsilon)}^{\tau} d\tau' a(\tau')\,\langle 0|\zeta_{\mathbf{k_1}}(\tau)\zeta_{\mathbf{k_2}}(\tau)\zeta_{\mathbf{k_3}}(\tau)H_{int}(\tau')|0\rangle + c.c \quad (2.8)$$

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 $S(k_1, k_2, k_2)$, defined in (2.3), which is then used as input into (2.56). As an example, if one takes $H_{int} \propto \dot{\zeta}^3$, this set-up can produce the standard EFT shape

$$S(k_1, k_2, k_3) = \frac{k_1 k_2 k_3}{(k_1 + k_2 + k_3)^3}. \quad (2.9)$$

The central point, as noticed in [2], is that the integrand of (2.8) 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 (2.9), 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 (2.8). 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 [23]. This can be used, as pointed out in [2], 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 [4], 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 [3]. However, by breaking these assumptions large signals can arise. These signals are usually calculated using (2.8) within tailored approximations. The results are not always separable, so further approximations must then be made to allow comparison with the CMB.

2.5 Calculating the interaction Hamiltonian

General overview of how the interaction Hamiltonian is obtained.

2.6 Self-interactions

Simpler derivation neglecting the metric perturbations to begin with.

Discuss DBI, then $P(X, \phi)$.

2.7 The Maldacena calculation

Review, with metric perturbations.

2.8 The field-redefinition

Discuss field-redefinition being unnecessary, as per [4].

2.9 Templates

Using K_{pq} notation. Link to Enrico's symmetric polynomials, reference discussion under equation (57) in [24].

2.9.1 Basic templates, f_{NL}

The three basic phenomenological templates, *Planck*. f_{NL} from the data analysis perspective, and from the primordial perspective.

2.10 Shapes

2.10.1 Basic shapes

Maldacena, DBI.

2.10.2 $P(X, \phi)$, EFT

There is an extensive literature on the calculation of bispectra from models of inflation [25, 26, 27, 28, 4, 7, 8, 29, 30]. 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 [31, 5, 32, 33, 34, 35, 36]; effectively single-field models with imaginary sound speeds can generate a bispectrum mostly orthogonal to the usual equilateral and local templates [37]. 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. We will work with an inflaton action of the form

$$S = \int d^4x \sqrt{-g} P(X, \phi) \tag{2.10}$$

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: $x' = \frac{dx}{dN} = a\frac{dx}{da}$. We define the Hubble parameter and the standard "slow-roll" parameters:

$$\begin{aligned} H &= \frac{d\ln a}{dt}\,, \qquad & \epsilon &= -\frac{d\ln H}{dN} \\ \eta &= \frac{d\ln \varepsilon}{dN}\,, \qquad & \epsilon_s &= +\frac{d\ln c_s}{dN}\,. \end{aligned} \tag{2.11}$$

though we make no assumption that these are actually small. c_s is the sound speed of the theory, which can vary with time:

$$c_s = \frac{P_{,X}}{P_{,X} + 2XP_{,XX}}. \tag{2.12}$$

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

$$\rho = 2XP_{,X} - P, \tag{2.13}$$

so then

$$\rho' = -6XP_{,X}\,. \tag{2.14}$$

Then, using the Friedmann equation we obtain

$$\varepsilon = -\frac{1}{2}\phi'^2 P_{,X}\,. \tag{2.15}$$

The equation of motion for ϕ is [38]

$$\phi'' + (3c_s^2 - \varepsilon)\phi' + H^{-2}\frac{\rho_\phi}{\rho_X} = 0. \tag{2.16}$$

We can now evolve τ_s, ϕ and H numerically using (2.19), (2.16) and (2.15) 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}, obtaining consistent initial conditions to start out numerical evolution.

The equation of motion for the perturbations is:

$$\zeta_k'' + (3 - \varepsilon + \eta - 2\varepsilon_s)\zeta_k' + \frac{c_s^2 k^2}{a^2 H^2}\zeta_k = 0 \tag{2.17}$$

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:

$$\zeta_k = \frac{i}{a}\sqrt{\frac{c_s}{4\varepsilon k}}e^{-ik\tau_s} \tag{2.18}$$

where we define τ_s through $\tau_s' = \frac{c_s}{aH}$ in analogy with the usual τ with

$$\tau' = \frac{1}{aH}. \tag{2.19}$$

The solution in slow-roll (without features) is then approximately

$$\zeta_k \propto (1 + ik\tau_s)e^{-ik\tau_s}. \tag{2.20}$$

At leading order in slow-roll the power spectrum is [39, 27]:

$$P^{\zeta}(k) = \frac{1}{8\pi^2}\frac{H^2}{c_s\varepsilon}, \tag{2.21}$$

where the right hand side is evaluated at $c_s k = aH$. The spectral index is (also to leading order):

$$n_s - 1 = -2\varepsilon - \eta - \varepsilon_s. \tag{2.22}$$

Similarly to [2], at early times we extract the factor of $e^{-ik\tau_s}$ from the mode functions and numerically evolve $\zeta_k e^{ik\tau_s}$[1]. Unless interrupted, this prefactor decays exponentially. Eventually 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 gives robust results without needing to loosen the tolerance.

Initially we consider the same basic models as in [2]; a quadratic potential

$$V_{\phi^2}(\phi) = \frac{1}{2}m^2\phi^2. \tag{2.23}$$

with a canonical kinetic term, and a non-canonical model, with a DBI kinetic term

$$S_{DBI} = \int d^4x\sqrt{-g}\left(-\frac{1}{f(\phi)}\left((1+f(\phi)\partial_\mu\phi\partial^\mu\phi)^{\frac{1}{2}} - 1\right) - V(\phi)\right), \tag{2.24}$$

with

$$f(\phi) = \frac{\lambda_{DBI}}{\phi^4}, \qquad V(\phi) = V_0 - \frac{1}{2}m^2\phi^2, \qquad m = \sqrt{\beta_{IR}}H. \tag{2.25}$$

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

$$V(\phi) = V_{\phi^2}(\phi)\left(1 - c\tanh\left(\frac{\phi_f - \phi}{d}\right)\right). \tag{2.26}$$

To explore non-Gaussianity from deeper in the horizon we imprint extended resonant

[1]In fact [2] extracts a factor of $e^{-ikc_s(\tau)\tau}$, losing efficiency due to slow-roll corrections.

features on the basic potential

$$V(\phi) = V_{\phi^2}(\phi)\left(1 + bf\sin\left(\frac{\phi}{f}\right)\right). \tag{2.27}$$

For more details on these models, see [26]. To express the bispectrum results more compactly we use the symmetric polynomial notation employed in [20]:

$$\begin{aligned} K_p &= \sum_{i=1,2,3} k_i^p, \\ K_{pq} &= \frac{1}{\Delta_{pq}} \sum_{i\neq j} k_i^p k_j^q, \\ K_{prs} &= \frac{1}{\Delta_{prs}} \sum_{i\neq j\neq l} k_i^p k_j^r k_l^s, \end{aligned} \tag{2.28}$$

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

$$S^{Malda}(k_1, k_2, k_3) = A^{Malda}\left((3\varepsilon - 2\eta)\frac{K_3}{K_{111}} + \varepsilon\left(K_{12} + 8\frac{K_{22}}{K}\right)\right), \tag{2.29}$$

$$A^{Malda} = -\frac{1}{32}\frac{H^4}{12\varepsilon^2}. \tag{2.30}$$

with $\eta = 2\varepsilon$ for (2.23). At the primordial level, this is well approximated by the separable local template

$$S^{local}(k_1, k_2, k_3) = \frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_3 k_1} + \frac{k_3^2}{k_1 k_2} = \frac{K_3}{6\,K_{111}}. \tag{2.31}$$

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 [40]. The local template is in fact used to test for multi-field effects [11]. For the featureless DBI scenario, the shape function is [5]:

$$S^{DBI}(k_1, k_2, k_3) = A^{DBI}\frac{K_5 + 2K_{14} - 3K_{23} + 2K_{113} - 8K_{122}}{K_{111}K^2}, \tag{2.32}$$

$$A^{DBI} = -\frac{1}{32}\frac{H^4}{12\varepsilon^2}\left(\frac{1}{c_s^2} - 1\right), \tag{2.33}$$

to leading order in slow-roll. Any constraint on the magnitude A^{DBI} 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 [11]. The shape (2.32) can be approximated by the separable equilateral template

$$S^{equil}(k_1,k_2,k_3) = \frac{(k_2+k_3-k_1)(k_3+k_1-k_2)(k_1+k_2-k_3)}{k_1k_2k_3}. \tag{2.34}$$

These templates can be modified to be more physically realistic by including scaling consistent with the spectral index n_s [11]. For example, we can add some scale dependence to the DBI model in a reasonable first approximation by including a prefactor. We define the product scaling template

$$S^{DBI-n_s}(k_1,k_2,k_3) = \left(\frac{k_1k_2k_3}{k_\star^3}\right)^{\frac{n_{NG}}{3}} S^{DBI}(k_1,k_2,k_3) \tag{2.35}$$

and the sum scaling template

$$S^{DBI-n_s}(k_1,k_2,k_3) = \left(\frac{k_1+k_2+k_3}{3k_\star}\right)^{n_{NG}} S^{DBI}(k_1,k_2,k_3) \tag{2.36}$$

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

We now turn to feature templates. The result of adding a feature of the form (2.26) is to add oscillatory features of the form

$$S^{\cos}(k_1,k_2,k_3) = \cos(w(k_1+k_2+k_3)) \tag{2.37}$$

though more realistically there is some phase, shape dependence and a modulating envelope, as detailed in [7]. The result of adding a resonant feature of the form (2.27) is to generate logarithmic oscillatory features in the shape function of the form

$$S^{\ln-\cos}(k_1,k_2,k_3) = \cos(w\ln(k_1+k_2+k_3)). \tag{2.38}$$

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 [35].

Much success has been had in constraining non-Gaussianity in the CMB using separable approximations to these approximate templates. Other methods target oscillations [41], 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.

2.10.3 Shapes from features during inflation

Explicit details of how resonance and features generate large NG.

2.11 Previous work on in-in separability.

In [2] 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, [2] 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 [19, 20, 21] 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 [2]) are constraints on the parameters of certain inflation models through approximate phenomenological templates. These constraints can be found in [11, 12]. The idea of [2] is an extension of that philosophy to the primordial level, and our work is in implementing that idea. In [19, 20, 21] 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 [2]. We restructure the methods, improving on the work of [2] 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 [2] 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 [2] allow us to validate on non-trivial bispectra for the first time, including sharp deviations from slow-roll, which we present in section 4.6. We quote our results in terms of a measure that is easier to interpret than the correlation defined in [2], and that includes the magnitude as well as the shape information on the full tetrapyd. This is discussed in section 3.2.

2.11.1 Comparison to the present work

Summary of the achievements and limitations of [2], how I went beyond them.

2.12 Configuration-by-configuration codes

Previous work on the numerical calculations of inflationary non-Gaussianity include the BINGO code [42], Chen et al [25, 26], the work of Horner et al [43, 44, 45] and the

Transport Method [46, 47, 48, 49]. 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 formalism.[2] 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 [37].

2.12.1 Usage in recent works

Sidetracked inflation, etc… [50] using the transport method [48].

2.12.2 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 of points which must be processed further to be usefully compared to observation. Secondly, they must carefully implement some variation of the $i\epsilon$ prescription without affecting the numerical results. In [46] this is achieved in the initial conditions for the bispectra; other methods impose some non-trivial cutoff at early times.

[2] See A.3 for details on applying this strategy to our own modal coefficients.

2.13 The squeezed limit consistency condition

Could this be better placed somewhere else? 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 [40]. 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 [51, 52]. With $\mathbf{k_S} \equiv (\mathbf{k_2} - \mathbf{k_3})/2$:

$$S(k_1, k_2, k_3) = -\left[(n_s - 1)|_{k_S} + \mathcal{O}\left(\frac{k_1^2}{k_S^2}\right)\right] P_\zeta(k_1) P_\zeta(k_S), \quad k_1 \ll k_S \tag{2.39}$$

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 [52].

2.14 Review forecasts for future surveys

E.g. [53]. Talk about Astro2020 papers here? [54, 55]

2.15 Review estimators, KSW, separability

E.g. [56, 9].

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

$$\frac{\Delta T}{T}(\hat{\mathbf{n}}) = \sum_{lm} a_{lm} Y_{lm}(\hat{\mathbf{n}}) \tag{2.40}$$

and the power spectrum is

$$C_l = \sum_m a_{lm} a_{l-m}. \tag{2.41}$$

The bispectrum derived from theory is denoted

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle \tag{2.42}$$

whereas the bispectrum measured from the CMB is

$$B^{l_1 l_2 l_3}_{m_1 m_2 m_3} = a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}. \tag{2.43}$$

We also have

$$B_{l_1 l_2 l_3} = \sum_{m_i} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B^{l_1 l_2 l_3}_{m_1 m_2 m_3}. \tag{2.44}$$

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:

$$\sum_{l_i, m_i} \left(\lambda \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} - \frac{a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right)^2. \tag{2.45}$$

The solution to this is simply

$$\lambda = \frac{\frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \cdot \frac{a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}}}{\left| \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right|^2} \tag{2.46}$$

where $\cdot$ denotes summation over l_i and m_i. Defining

$$N = \left| \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right|^2 \tag{2.47}$$

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

$$\mathcal{E} = \frac{1}{N} \sum_{l_i, m_i} \frac{\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle \, a^{obs}_{l_1 m_1} a^{obs}_{l_2 m_2} a^{obs}_{l_3 m_3}}{C_{l_1} C_{l_2} C_{l_3}}. \tag{2.48}$$

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 to 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:

$$C_{l_1 m_1, l_2 m_2} \approx C_l \delta_{l_1 l_2} \delta m_1 \; - m_2, \tag{2.49}$$
$$\tilde{C}_l = b_l^2 C_l + N_l, \tag{2.50}$$
$$\bar{b}_{l_1 l_2 l_3} = b_{l_1} b_{l_2} b_{l_3} b_{l_1 l_2 l_3}. \tag{2.51}$$

so that the estimator can be written as

$$\mathcal{E} = \frac{1}{\tilde{N}^2} \sum_{l_i m_i} \frac{\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} \bar{b}_{l_1 l_2 l_3}}{\tilde{C}_{l_1} \tilde{C}_{l_2} \tilde{C}_{l_3}} \left(a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} - 6 C^{sim}_{l_1 m_1, l_2 m_2} a_{l_3 m_3} \right). \tag{2.52}$$

What are the beam effects? What is the noise? What approximations go into this? What are the references for those? We have here used the Gaunt integral

$$\mathcal{G}^{l_1 l_2 l_3}_{m_1 m_2 m_3} = \int d\Omega Y_{l_1 m_1} \hat{\mathbf{n}} Y_{l_2 m_2} \hat{\mathbf{n}} Y_{l_3 m_3} \hat{\mathbf{n}} \qquad = h_{l_1 l_2 l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \tag{2.53}$$

and we can then define

$$b_{l_1 l_2 l_3} = h^{-1}_{l_1 l_2 l_3} B_{l_1 l_2 l_3}. \tag{2.54}$$

2.16 Separable approximations to non-separable templates

Separability is required, so usually one approximates non-separable templates by separable ones.

Focus on equilateral to DBI. Quantitative comparison of equilateral template and DBI shape, with and without scaling.

2.17 Estimating the bispectrum, complexity

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 [55, 54] 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 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 [9, 10]. 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 [19, 20, 21] 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 [19, 20, 21] 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 (2.3) has the form:

$$S(k_1, k_2, k_3) = X(k_1)Y(k_2)Z(k_3), \tag{2.55}$$

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 the following expression:

$$b_{l_1 l_2 l_3}^{X_1 X_2 X_3} = \left(\frac{2}{\pi}\right)^3 \int_0^\infty dr r^2 \int_{\mathcal{V}_k} d^3k (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3) \prod_{i=1}^{3} \left[j_{l_i}(k_i r) \Delta_{l_i}^{X_i}(k_i) \right] , \quad (2.56)$$

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 [18].

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

2.18 Modal methods, constraints from Planck

Leveraging the separable benefits in a broader set of models through expansion.

2.19 Wuhyun's work as development of Modal methods

Contrast Wuhyun's work (our pipeline) to previous modal methods. What Wuhyun does is purely frequentist, not Bayesian. It is essentially linear regression, so can do Fisher forecast to determine the best possible variance for an estimator.

Chapter 3

Decomposing primordial shapes

3.1 Setting up the formalism

3.1.1 Pulling out the k-dependence

Given its separable form, the tree-level in-in formalism is amenable to more efficient calculation using separable modes, as first mentioned in [2]. That work extended the separable methodology previously implemented for the CMB bispectrum [21]. Our goal in this work is the efficient calculation of more general bispectra which may have significant (possibly oscillatory) features, requiring searches across free parameter dependencies. To achieve this, we represent the shape function (2.3) using a set of basis functions as

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

where the basis functions $Q_n(k_1, k_2, k_3)$ are explicitly separable functions of their arguments. Translating this result into a constraint from the CMB will require a large once-off computational cost, paid once per set of basis functions Q_n, not per scenario (encoded in α_n). The details of this once-per-basis calculation will be presented in [1]. 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; we do this in chapter 3. In section 3.1 we set the notation we will use to recast the standard numerical in-in calculation into a calculation of α_n, and sketch the steps involved. In section 4.4 we outline the details of the interaction Hamiltonian, including accounting for the spatial derivatives in our final result. In section 3.1.1 we make precise the numerical considerations of the calculation, especially our methods of dealing with the high-frequency oscillations at early times.

3.1.2 Relationship to decomposition on the cube

3.1.3 The core of the calculation

(Defining the notation for the two main time integrals.) In this section we set up the notation, and sketch the steps required to calculate the coefficients α_n in (3.1). The values of these coefficients will depend on the choice of basis, but the description of the methods below will remain mostly basis agnostic. Our aim will be to separate out the dependence on k and τ_s, without losing information, 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 (2.8) is inherently separable given the form of the cubic interaction Hamiltonian H_{int}. Consider indexing with $i = 1, 2, 3...$ the interaction vertices in H_{int}, so then the bispectrum (2.8) can be expressed as a sum over separable contributions of the form:

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

where $w^{(i)}(\tau)$ is a function of the scale factor and the other background parameters (2.11) for the i-th interaction vertex, while the terms $F^{(i)}, G^{(i)}, J^{(i)}$ are given by the Fourier mode functions $k^2\zeta_{\mathbf{k}}(0)\zeta^*_{\mathbf{k}}(\tau)$ or their time derivatives $k^2\zeta_{\mathbf{k}}(0)\zeta^{*\prime}_{\mathbf{k}}(\tau)$. Spatial derivative terms, such as $\partial_i\zeta\,\partial_i\zeta \longrightarrow (\mathbf{k_2}\cdot\mathbf{k_3})\zeta^*_{\mathbf{k}_2}(\tau)\,\zeta^*_{\mathbf{k}_3}(\tau)$ also separate because of the triangle condition (2.4) giving $\mathbf{k_2}\cdot\mathbf{k_3} = (k_1^2 - k_2^2 - k_3^2)/2$, yielding 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.

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; for details, see section 4.4. 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. These terms look like $F^{(i)}(k,\tau) = k^2\zeta_{\mathbf{k}}(0)\zeta^*_{\mathbf{k}}(\tau)$ (or $k^2\zeta_{\mathbf{k}}(0)\zeta^{*\prime}_{\mathbf{k}}(\tau)$), where k^2 comes from using the weighting of the scale-invariant shape function (2.3). This 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 (see chapter 3) 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 spatial derivatives in section 4.4 we will find it useful to change to a different basis than the one used to perform the time integrals of the decompositions.

Using the approximate mode functions (2.20), an explicit example for the first interaction term in (4.9), i.e. $H^{(1)}_{\mathrm{int}} = \zeta'^2\zeta$, takes the form

$$F^{(1)}(\tau,k) = G^{(1)}(\tau,k) \;=\; c_s k^2\tau\frac{H^2}{4\varepsilon c_s}e^{ic_s k\tau}\,, \qquad J^{(1)}(\tau,k) \;=\; (1-ikc_s\tau)\frac{H^2}{4\varepsilon c_s}e^{ic_s k\tau}\,. \tag{3.3}$$

In the simple mode approximation (2.20), 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 [3]. However, for high precision bispectrum predictions we must incorporate the full time-dependence, while solving (2.17) 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.

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 HERE real part only?????

$$S(k_1,k_2,k_3) = \sum_i I^{(i)}(k_1,k_2,k_3) = \sum_i\sum_n \alpha^{(i)}_n\, Q_n(k_1,k_2,k_3)\,, \tag{3.4}$$

where $Q_n(k_1,k_2,k_3)$ is separable, built out of some orthonormal set $q_p(k)$ as in (3.18). Armed with this set of modes, we can expand any of the interaction terms $F^{(i)}(\tau,k)$,

[1] At early times the modes are highly oscillatory in both k and τ_s, which certainly requires special attention. We discuss this in section 3.1.1.

$G^{(i)}(\tau,k)$, $J^{(i)}(\tau,k)$ in (3.2) as:

$$F^{(i)}(\tau,k)=\sum_p f_p^{(i)}(\tau)\,q_p(k)\,, \tag{3.5}$$

$$\text{where}\quad f_p^{(i)}(\tau)=\int_{k_{\min}}^{k_{\max}} dk\,F^{(i)}(\tau,k)\,q_p(k)\,. \tag{3.6}$$

Note that in the simple mode approximation, as in (3.3), we must expand $e^{ic_s k\tau}$ in the terms of the $q_p(k)$. At early times τ 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, reducing the number of needed k-samples; see section (3.1.1) for details. Secondly, it forces us to calculate $f_p^{(i)}(\tau)$ up to very high p if we want to accurately converge to $F^{(i)}(\tau,k)$, for sets of basis functions such as the Legendre polynomials. In fact, obtaining a convergent final bispectrum result does not require calculating the full convergent sum for $F^{(i)}(\tau,k)$ in (3.5), as the highly oscillatory parts will cancel in the time integrals for any sufficiently smooth $S(k_1,k_2,k_3)$.

Substituting these expansions into (3.2), we obtain the following decomposition for the i-th vertex contribution,

$$\begin{aligned}
&I^{(i)}(k_1,k_2,k_3)\\
=\;& v^{(i)}(k_1,k_2,k_3)\int d\tau\, w^{(i)}(\tau)\,\sum_p f_p^{(i)}(\tau)q_p(k_1)\sum_r g_r^{(i)}(\tau)q_r(k_2)\sum_s h_s^{(i)}(\tau)q_s(k_3)+\text{cyclic perms}\\
=\;& v^{(i)}(k_1,k_2,k_3)\sum_{prs}\left(\int d\tau\, w^{(i)}(\tau)\, f_p^{(i)}(\tau)\, g_r^{(i)}(\tau)\, h_s^{(i)}(\tau)\right) q_p(k_1)\,q_r(k_2)\,q_s(k_3)+\text{cyclic perms}\,.
\end{aligned}$$

For the sake of compactness we use P to stand for the triplet p,r,s and $\tilde{P}$ to stand for the triplet $\tilde{p},\tilde{r},\tilde{s}$. Writing $q_P(k_1,k_2,k_3)=q_p(k_1)\,q_r(k_2)\,q_s(k_3)$, we continue,

$$\begin{aligned}
I^{(i)}(k_1,k_2,k_3) &= v^{(i)}(k_1,k_2,k_3)\sum_P \tilde{\alpha}_P^{(i)}\,q_P(k_1,k_2,k_3)+\text{cyclic perms}\\
&= \sum_P \tilde{\alpha}_P^{(i)}\sum_{\tilde{P}} V_{P\tilde{P}}^{(i)}\,q_{\tilde{P}}(k_1,k_2,k_3)+\text{cyclic perms}\\
&= \sum_{\tilde{P}} \alpha_{\tilde{P}}^{(i)}\,q_{\tilde{P}}(k_1,k_2,k_3)+\text{cyclic perms}\,,
\end{aligned} \tag{3.7}$$

where we have written

$$\tilde{\alpha}_P^{(i)}=\tilde{\alpha}_{prs}^{(i)}=\int d\tau\, w^{(i)}(\tau)\, f_p^{(i)}(\tau)\, g_r^{(i)}(\tau)\, h_s^{(i)}(\tau)\,, \tag{3.8}$$

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

$$v^{(i)}(k_1,k_2,k_3)q_P(k_1,k_2,k_3) \;=\; \sum_{\tilde{P}} V^{(i)}_{P\tilde{P}} q_{\tilde{P}}(k_1,k_2,k_3)\,, \tag{3.9}$$

and

$$\alpha^{(i)}_P = \sum_{\tilde{P}} \tilde{\alpha}^{(i)}_{\tilde{P}} V^{(i)}_{\tilde{P}P}. \tag{3.10}$$

We connect to the coefficients of the ordered, symmetrised basis in (3.4) by taking the symmetry factor into account,

$$\alpha^{(i)}_n = \frac{3!}{\Xi_P}\alpha^{(i)}_P. \tag{3.11}$$

The numerical calculation of $V^{(i)}_{P\tilde{P}}$ (as defined by (3.9)) 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 (3.5)) 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 (3.4) 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 (3.8) we need to calculate.

Note the basis sets on the left and right hand side of (3.9) 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 (3.9) 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 $p_{\text{max}}+3$.

We can see from (3.8) that the number of time integrals needed is controlled by $N_V \times {p^3_{\text{max}}}^2$, where N_v is the number of interaction vertices and p_{max} is the size of the final 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

[2] In fact the number is not quite p^3_{max}. Since we have extracted the spatial derivatives, the only remaining possible source of asymmetric k-dependence comes from ζ^3, $\zeta^2\zeta'$, $\zeta'^2\zeta$ or ζ'^3 so the time integral in (3.8) will always be (at least) symmetric in p and r.

gained from reducing the number of terms needed in the final basis.

We will calculate the contribution of each H_{int} vertex separately, indexing the vertices as above by (i), so the overall shape function (3.2), (3.4) is then simply

$$S(k_1, k_2, k_3) = \sum_n \left(\sum_i \alpha_n^{(i)} \right) Q_n(k_1, k_2, k_3) = \sum_n \alpha_n \, Q_n(k_1, k_2, k_3) \,. \tag{3.12}$$

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.

The raison d'etre for this approach is that all time integrals (3.8) are now independent of the k-configuration[3]. In a configuration-by-configuration method one improves the precision by decreasing the spacing which defines the density of the grid of points within the tetrapyd. Instead, in the modal approach, we increase precision by adding more modes to the shape function expansion (3.12) 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) plus 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 (2.29) and (2.32), but is also true of bispectra with complicated features, as seen in chapter 3.

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 while the primordial basis is chosen for computational speed and convenience, it can be independent of the final bispectrum basis employed for observational tests; a change of basis $Q_n \to \tilde{Q}_n$ can be achieved through a linear transformation Γ with the new expansion coefficients given by $\tilde{\alpha}_m = \Gamma_{mn}\alpha_n$.

Discussion of convergence in this work 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

[3]They 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.

discuss the actual numerical implementation of these methods.

3.2 Inner product choice

The inner product of two bispectrum shape functions is given by

$$S_1 \cdot S_2 = \langle S_1 \, , S_2 \rangle = \int_{T_k} d^3k \; S_1(k_1, k_2, k_3) \; S_2(k_1, k_2, k_3) \, , \tag{3.13}$$

where T_k refers to the tetrapyd, the region of the cube $[k_{\rm min}, k_{\rm max}]^3$ that obeys the triangle inequality. Following [57] we define the two correlators:

$$\mathcal{S}(S_1, S_2) = \frac{S_1 \cdot S_2}{\sqrt{(S_1 \cdot S_1)(S_2 \cdot S_2)}} \, , \qquad \mathcal{A}(S_1, S_2) = \sqrt{\frac{S_1 \cdot S_1}{S_2 \cdot S_2}} \, . \tag{3.14}$$

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 [19, 20, 21], but in this work we restrict ourselves to accurately calculating the full primordial bispectra, weighting each configuration equally.

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

$$\begin{aligned} \mathcal{E}(S_1, S_2) &= \sqrt{\frac{|S_1 - S_2|^2}{|S_2|^2}} = \sqrt{\frac{|S_1|^2 - 2S_1 \cdot S_2 + |S_2|^2}{|S_2|^2}} \\ &= \sqrt{\mathcal{A}(S_1, S_2)^2 - 2\mathcal{A}(S_1, S_2)\mathcal{S}(S_1, S_2) + 1}. \end{aligned} \tag{3.15}$$

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

$$\mathcal{E}(S_1, S_2) = \sqrt{2(1 - \mathcal{S}(S_1, S_2))}. \tag{3.16}$$

3.2.1 Interpretation

With this measure of relative difference, a shape correlation of 0.9 corresponds to an error of 45%, a shape correlation of 0.99 corresponds to an relative difference of 14%, a shape correlation of 0.999 corresponds to an relative difference of 4%. Thus this more exacting measure $\mathcal{E}$ from [57] is a far better representation of actual convergence between two shape functions than the correlation used in [2], as it is easier to interpret and more stringent.

The full relative difference (3.15) 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 4.6. In that section we also plot residuals on slices through the tetrapyd, relative to the representative value

$$\sqrt{\frac{S \cdot S}{\int_{T_k} d^3k}}. \tag{3.17}$$

3.2.2 Weighting

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.

3.3 Testing on templates

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 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.

3.4 The cube and the tetrapyd

We begin our methods discussion by exploring 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 symmetrised triplet products of normalized one-dimensional modes $q_p(k)$ as

$$Q_n(k_1, k_2, k_3) \;\equiv\; \Xi_{prs}\, q_{(p}(k_1)\, q_r(k_2)\, q_{s)}(k_3)\,. \tag{3.18}$$

Here, n labels the ordered integer triplet $n \leftrightarrow \{prs\}$ in an appropriate manner (see some ordering alternatives in [21]), while the symmetrised average of all $\{prs\}$ permutations is

$$q_{(p}q_r q_{s)} \equiv (1/3!) \sum_{\text{perms}} q_p q_r q_s \qquad \text{and} \qquad \Xi_{prs} = \begin{cases} 1, & p = r = s \quad \text{all equal,} \\ \sqrt{3} & \{prs\} \;\; \text{any two equal,} \\ \sqrt{6} & \{prs\} \;\; \text{all different.} \end{cases} \tag{3.19}$$

Unless stated otherwise, the $\{prs\}$ triples for each permutation set n in (3.18) are represented by the coefficient with $0 \le p \le r \le s$, that is, $\alpha_n = \alpha_{prs} \equiv \alpha_{(prs)}$. This modal expansion is terminated at some $p_{\max}$ for which $\max(p, r, s) < p_{\max}$. Given the basis-agnostic methods we shall outline in the following sections 3.1, 4.4 and 3.1.1, 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 (3.18), thus determining the nature of the numerical errors in the calculated bispectrum. However, since we are going beyond the featureless examples of [2] 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 $[k_{\min}, k_{\max}]^3$. This is because the only decomposition we actually perform is a one-dimensional integral over $[k_{\min}, k_{\max}]$ (as we will see in (3.6)). 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 (2.29) 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 (2.32), this effect can be disastrous if not handled properly. This can be easily seen from (2.34), in the limit of small k_3. The triangle condition in that limit enforces $(k_2 - k_1)^2 \leq k_3^2$. This implies that $0 \leq k_3^2 - (k_2 - k_1)^2 \leq k_3^2$, forcing the shape to go to zero in that limit despite the k_3 in the denominator. On the non-physical part of the face, $k_2 - k_1$ is not small, and so the shape is boosted by $1/k_3$ relative to the equilateral configurations. These regions then dominate any attempted basis expansion. To overcome this problem, as we shall discuss, we will extend our basis to explicitly include this $1/k$ behaviour[4].

To date the most useful starting choice for modal bispectrum expansions has been shifted Legendre polynomials $P_r(x)$:

$$q_r(k) = \left(\frac{2r+1}{k_{\rm max} - k_{\rm min}}\right)^{1/2} P_r(\bar{k})\,, \tag{3.20}$$

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

$$\bar{k} = \frac{2k - k_{\rm max} - k_{\rm min}}{k_{\rm max} - k_{\rm min}}\,. \tag{3.21}$$

There is freedom to vary this mapping, which we shall exploit in section 3.8. We shall label as $\mathcal{P}_0$ the basis function set of pure Legendre polynomials in (3.20), with $r = 0, 1, \ldots, p_{\rm max}-1$. These were considered also in [2], 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

[4]There are results in the literature that describe generic $K = k_1 + k_2 + k_3$ poles in correlators—see for example [58]. A simple example can be understood by recalling that in standard calculations using the in-in formalism, the $i\varepsilon$ prescription is used to damp out contributions in the infinite past. This does not work for $K = 0$. While the resulting divergence (in K) is clearly outside the physical region of the tetrapyd, we will see its effects in the physical configurations. Given that this three-dimensional behaviour is generic, one might worry that we should take more care in building it into our one-dimensional basis. However, the excellent convergence in section 4.6 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.

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. It 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. We can mitigate against this by including a basis function to capture this $1/k$ behaviour,

$$q_{p_{\max}}(k) = \mathrm{Orth}\left[1/k\right] , \qquad \text{with} \quad \mathcal{P}_1 = \{q_r(k) \mid r = 0, 1, 2, \ldots, p_{\max}\} , \tag{3.22}$$

where Orth represents the projection orthogonal to the original Legendre polynomial basis $\mathcal{P}_0$. As we see in figure 3.1, the convergence properties for the augmented basis $\mathcal{P}_1$ are dramatically improved.

The two basis function sets actually used in [2] 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

$$q_r(k) = \left(\frac{2r+1}{\ln k_{\max} - \ln k_{\min}}\right)^{1/2} P_r\left(\overline{\ln k}\right) \qquad \text{with} \qquad \overline{\ln k} = \frac{2\ln k - \ln k_{\max} - \ln k_{\min}}{\ln k_{\max} - \ln k_{\min}} . \tag{3.23}$$

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

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

where the rescaled $\bar{k}$ is defined in (3.21). This can also be thought of as expanding the bispectrum $k_1k_2k_3\, S(k_1, k_2, k_3)$ in $\mathcal{P}_0$, instead of the shape function $S(k_1, k_2, k_3)$ itself. The consequence is that neither (3.23) nor (3.24) are orthogonal with respect to the flat weighting of the inner product (3.13). However, as shown in [2], these two basis sets (3.23) and (3.24) are able to approximate the three canonical bispectrum shapes. Nevertheless, our aim is to go beyond the featureless examples investigated in [2], 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 (3.23) and (3.24). 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 (2.56), 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 (3.23) and (3.24) 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. In the following we will perform a more general exploration of orthogonal sets of basis functions that can efficiently describe the necessary $1/k$ behaviour. In addition to using the Legendre polynomials as building blocks, we will also consider a Fourier basis for the purposes of comparison.

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. If we want to use some function f to augment a given set of orthogonal functions q_r, with $r = 0, \ldots, p_{\max}-1$, then we define

$$\tilde{f}(k) \;=\; \mathrm{Orth}\,[f(k)] \;\equiv\; f(k) - \sum_{r=0}^{p_{\max}-1} \frac{\langle f, q_r\rangle}{\langle q_r, q_r\rangle} q_r(k) \tag{3.25}$$

and add $\tilde{f}$ to our basis set, now of size $p_{\max}+1$. 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.

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

$$q_0(k) = 1\,, \quad q_{2r-1}(k) = \sin(\pi r \bar{k})\,, \quad q_{2r}(k) = \cos(\pi r \bar{k})\,, \quad 1 \le r \le (p_{\max}-3)/2 \tag{3.26}$$

$$q_{p_{\max}-2}(k) = \bar{k}\,, \quad q_{p_{\max}-1}(k) = \bar{k}^2\,. \tag{3.27}$$

Here, even the basic Fourier series have to be augmented by the linear k and quadratic k^2 terms (for a total size of $p_{\max}$), in order to satisfactorily approximate equilateral shapes (reflecting in part the preference for periodic functions). As with $\mathcal{P}_1$ defined in (3.22), we will similarly create an augmented Fourier basis $\mathcal{F}_1$ by adding the inverse $1/k$ term to the $\mathcal{F}_0$ basis, i.e. using $q_{p_{\max}}(k) = \mathrm{Orth}\,[1/k]$ with (3.25). 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 these four different basis function sets ($\mathcal{P}_0$, $\mathcal{F}_0$, $\mathcal{P}_1$ and $\mathcal{F}_1$), we have investigated their convergence on Maldacena's shape (2.29) and the DBI shape (2.32). To mimic the in-in calculation, we expand the shape on the cube, but test the result on the tetrapyd using (3.15). The results are shown in figure 3.1 where 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

Notation	Building Blocks	Augmented by	Definition
$\mathcal{P}_0$	Legendre polynomials		(3.20)
$\mathcal{F}_0$	Fourier Series	k, k^2	(3.26)
$\mathcal{P}_1$	Legendre polynomials	k^{-1}	
$\mathcal{F}_1$	Fourier Series	k, k^2, k^{-1}	
$\mathcal{P}_1^{n_s}$	Legendre polynomials	$k^{-1+(n_s^*-1)}$	(3.28)
$\mathcal{P}_{01}^{n_s}$	Legendre polynomials	$k^{n_s^*-1}$, $k^{-1+(n_s^*-1)}$	(3.29)
scaling	Legendre polynomials	k^{-1}, $\ln(k)k^{-1}$	
resonant	$\frac{\mathcal{P}_n(\bar{k})}{\sqrt{k}}$, with $\bar{k} = \frac{2\ln(k)-\ln(k_{\min}k_{\max})}{\ln(k_{\max})-\ln(k_{\min})}$		

Table 3.1: Basis summary—the augmentation of the basis is done using (3.25). The size of each basis is referred to as $p_{\max}$. Some of these basis sets are plotted in figure 3.7.

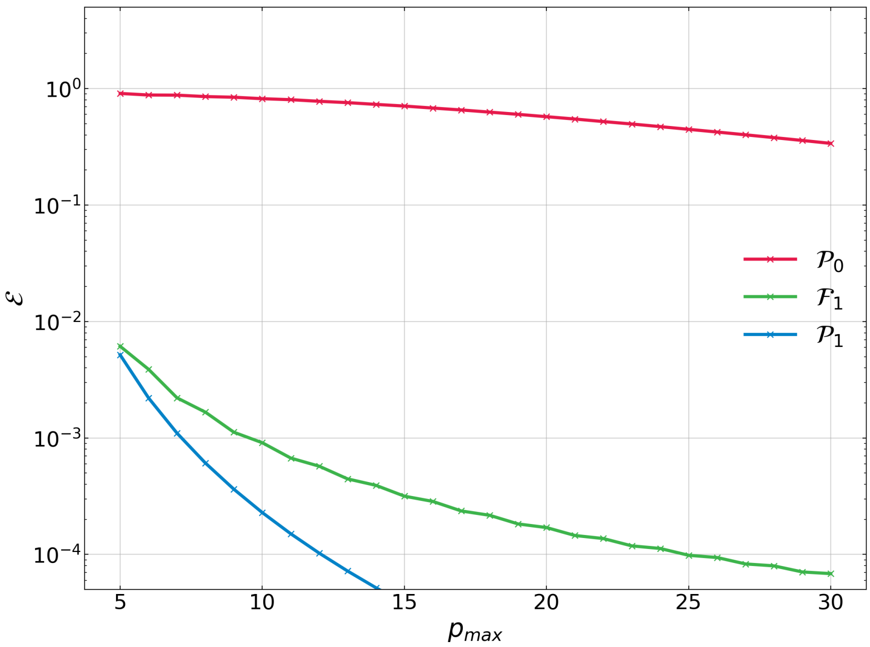
magnitude at $p_{\max} = 15$. The augmented Fourier basis $\mathcal{F}_1$ also converges quickly relative to $\mathcal{P}_0$, but is outdone by $\mathcal{P}_1$. Though we do not show the convergence on the cube, we find that for Maldacena's template this is of the same order of magnitude as the error 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 (2.37) and the resonance model (2.38) 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 (2.32) 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 (2.32) is superimposed, even the augmented Fourier modes $\mathcal{F}_1$ converge poorly. Instead, the basis sets built from Legendre modes offer a better more robust option, with the *scaling* basis performing best. In figure 3.3 we see that for a logarithmic oscillation, the *resonant* basis converges in the fewest modes.

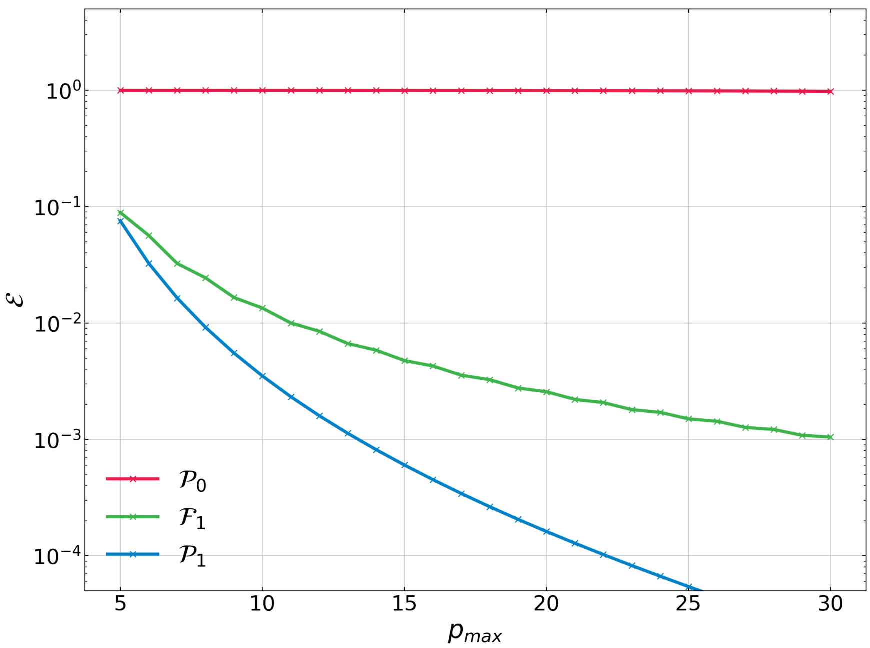
Finally, we consider convergence in the light of the more subtle scale-dependence due to the spectral index n_s of the power spectrum. The simple canonical examples in figure 3.1 had shape dependence and no scale dependence, but this would only be expected of scenarios unrealistically deep in the slow-roll limit. When we include this scale dependence, using (2.35) with n_s, it proves very useful to include these deviations from integer power laws in the basis functions. We consider two cases, first augmenting $\mathcal{P}_0$ by a scale-dependent $1/k$ term using the orthogonalisation procedure (3.25),

$$q_{p_{\max}}(k) = \mathrm{Orth}\left[k^{-1+(n_s^*-1)}\right], \tag{3.28}$$

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

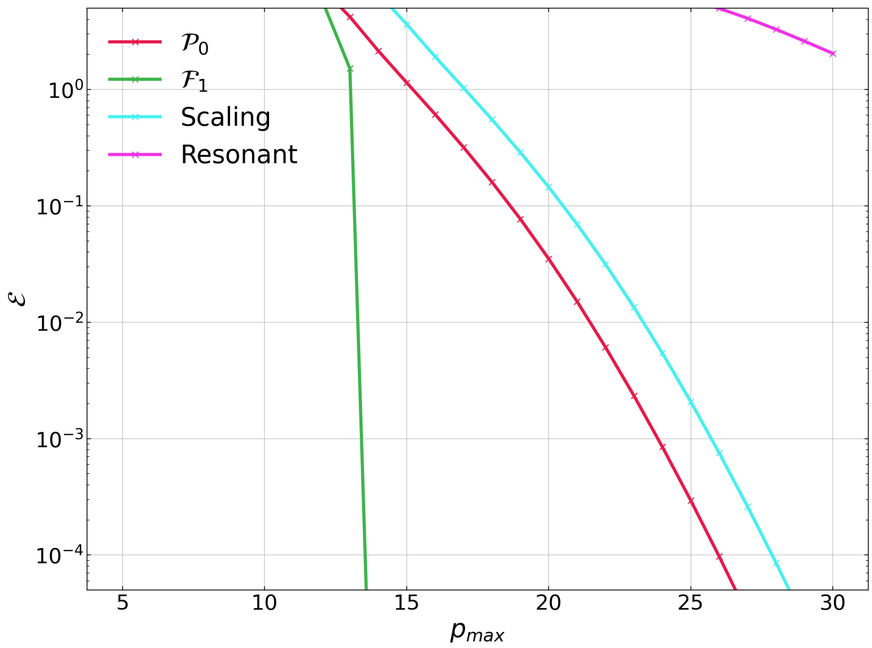


(a) Reconstructing the Maldacena Template

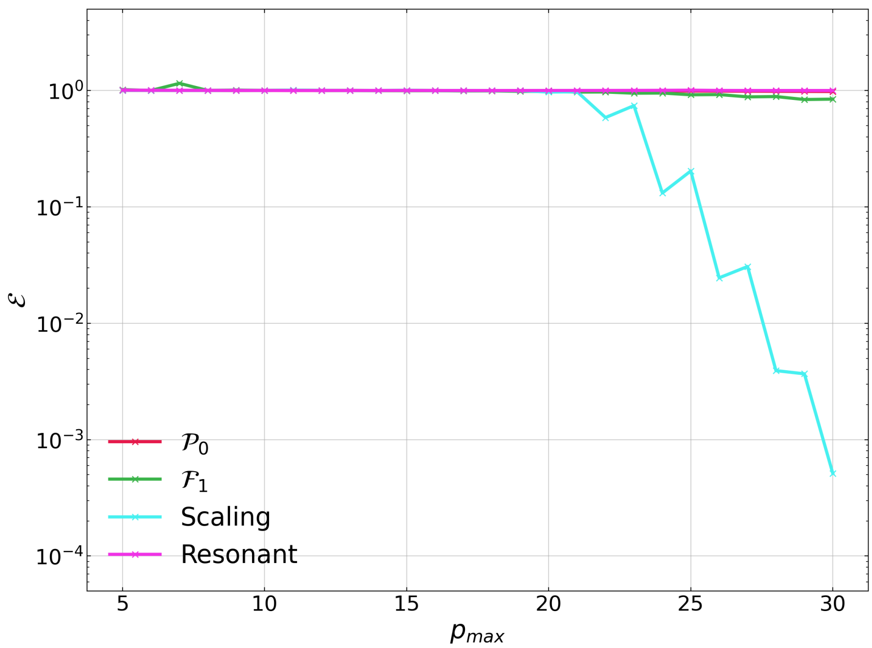


(b) Reconstructing the DBI Template

Figure 3.1: Convergence comparisons for the Legendre and Fourier basis functions for (a) the Maldacena template (2.29) and (b) the DBI template (2.32). The pure Legendre $\mathcal{P}_0$ basis requires many terms to fit the $1/k$ behaviour in both Maldacena's template (2.29) and the DBI template (2.32). 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 (2.32) are larger than (2.29) because of the larger contributions outside the tetrapyd dominating the fit.

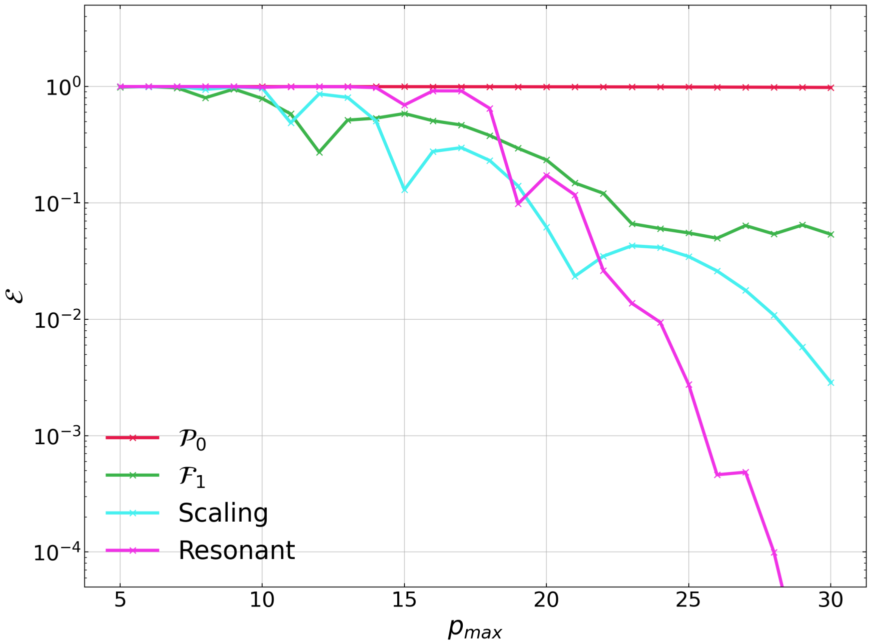


(a) $\cos(f(k_1+k_2+k_3))$

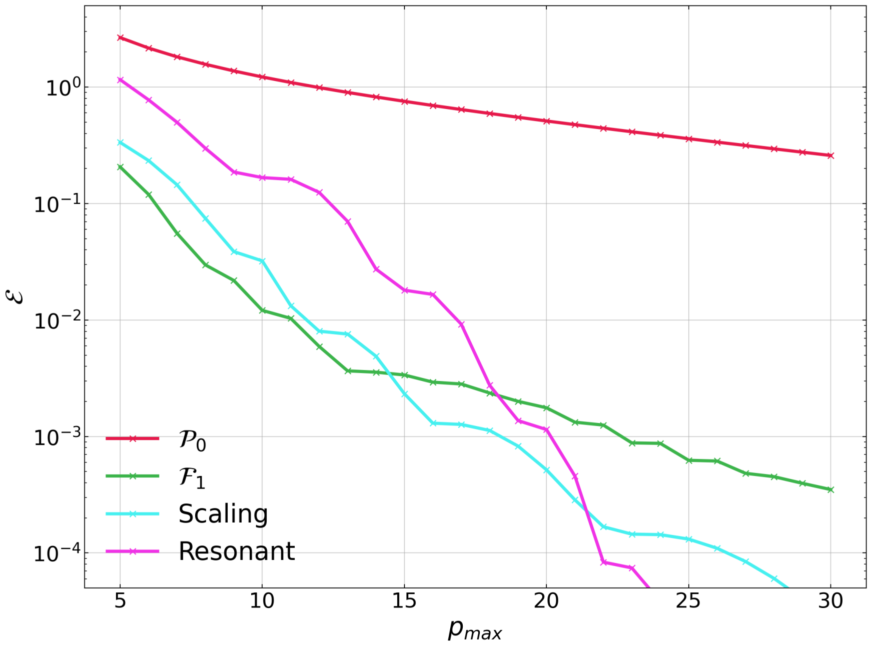


(b) $\cos(f(k_1+k_2+k_3))S^{DBI}$

Figure 3.2: Convergence comparison for oscillatory models. For linear oscillations we choose $f = 150.64$ (so we obtain 15 whole oscillations in the k-range). (a) As expected, the $\mathcal{F}_1$ basis fits an oscillation with no shape dependence (2.37) (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 (2.32) and (2.37)), we see that again *scaling* 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 (3.15) to the physical configurations on the tetrapyd; when considered over the entire cube the fit improves monotonically. The *resonant* basis, naturally, does not converge well to a linear oscillation.

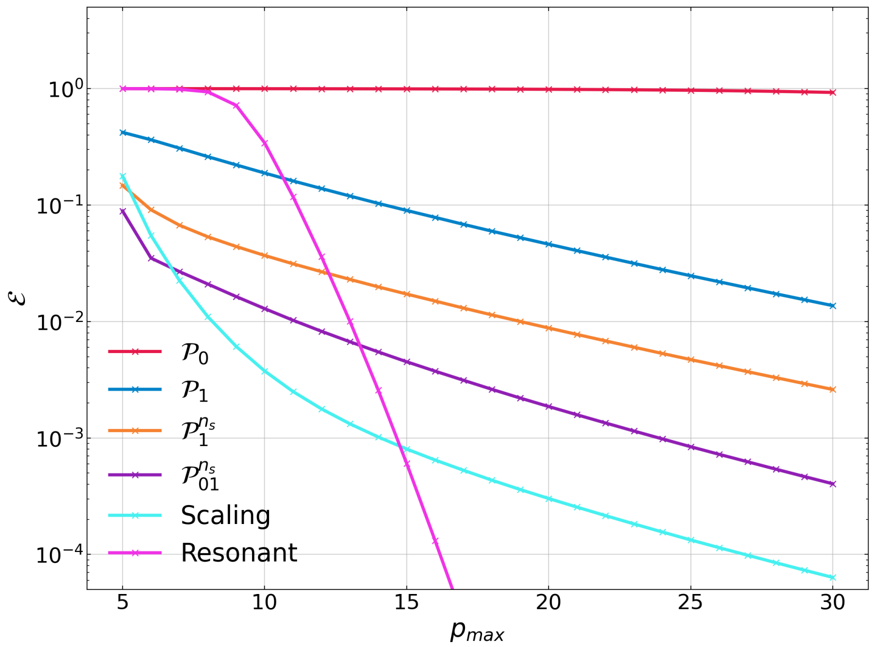


(a) $\cos(f \log(k_1 + k_2 + k_3))S^{DBI}$ on the tetrapyd

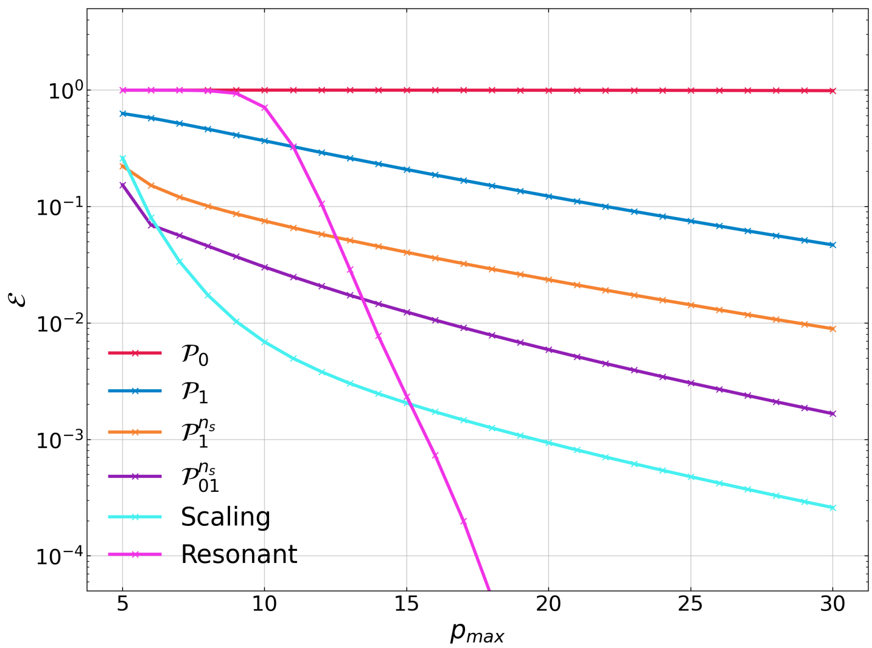


(b) $\cos(f \log(k_1 + k_2 + k_3))S^{DBI}$ on the cube

Figure 3.3: We plot the convergence of various basis sets for the more complex case of a resonant feature with scale and shape dependence, in this case the product of (2.32) and (2.38). 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 (3.8) 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 3.4.1.

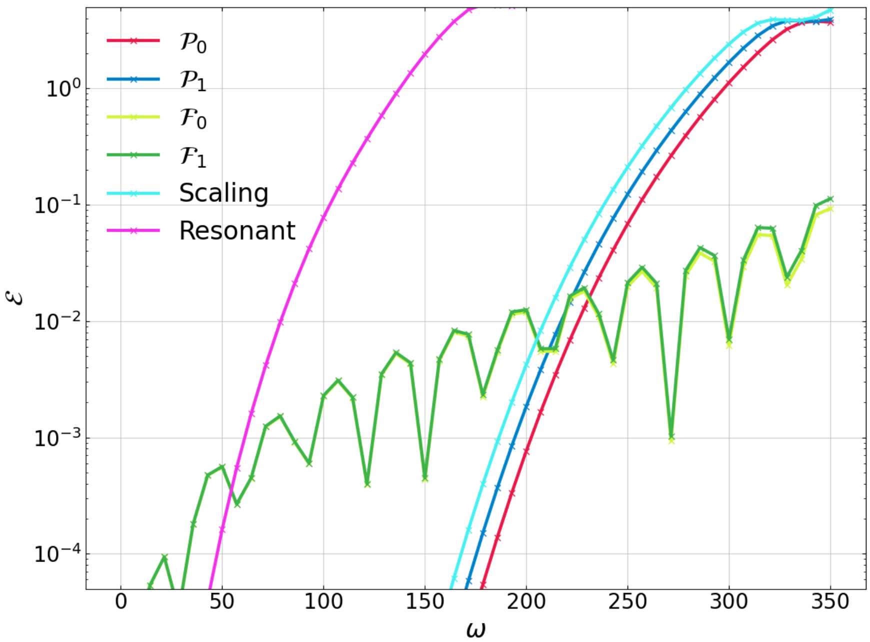


(a) $k_{\max}/k_{\min} = 550$

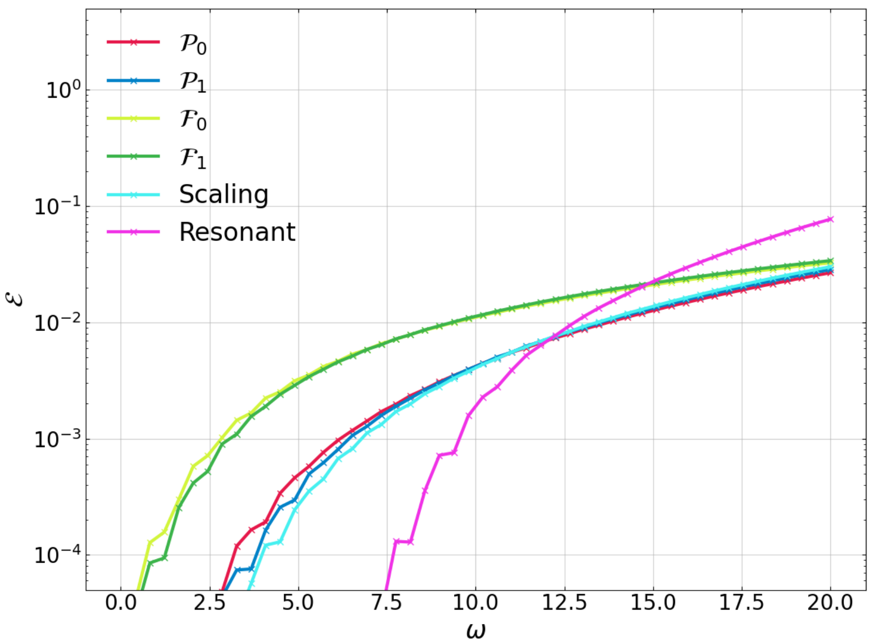


(b) $k_{\max}/k_{\min} = 1000$

Figure 3.4: For the scale-dependent DBI template (2.35), by including a minimal amount of power spectrum information using (3.28) and (3.29) ($\mathcal{P}_1^{n_s}$ 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.

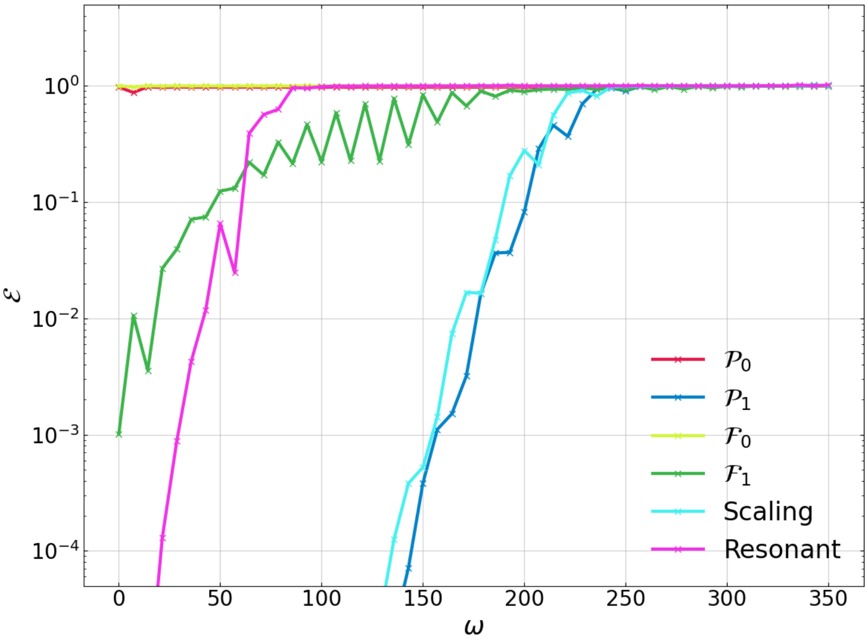


(a) $\cos(\omega(k_1 + k_2 + k_3))$

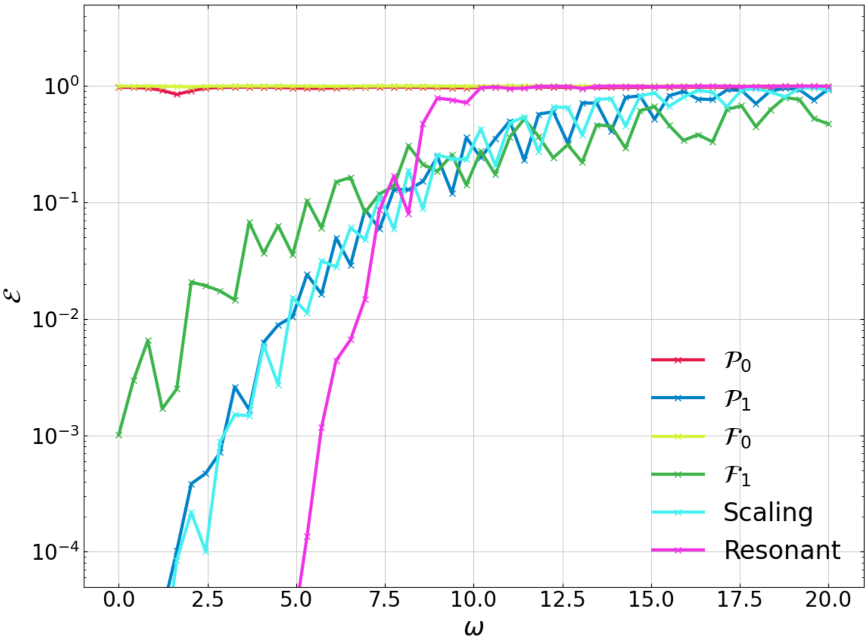


(b) $\cos(\omega \ln(k_1 + k_2 + k_3))$

Figure 3.5: Here we test our basis sets on oscillatory templates, which are a good proxy for feature models. These plots are useful to aid in deciding which basis to run BEstModal for, figuring out which covers the widest range of features PLANCK COMPARISON. P1ns isn't included in all the plots, but it would always lie in between $\mathcal{P}_1^{n_s}$ and *scaling* basis. The basis sets are defined in table 3.1, and some are plotted in figure 3.7. Basis sets $\mathcal{P}_1^{n_s}$ and *scaling* work well for the linear oscillations, converging up to around $\omega \approx 200$. For logarithmic oscillations the resonant basis works best, as expected. However, the improvement is less dramatic.



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



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

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

constant term as

$$q_{p_{\max}}(k) = \text{Orth}\left[k^{(n_s^*-1)}\right], \qquad q_{p_{\max}+1}(k) = \text{Orth}\left[k^{-1+(n_s^*-1)}\right], \tag{3.29}$$

which we refer to as $\mathcal{P}_{01}^{n_s}$. 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 be accurately approximated with a limited number of modes. We conclude that augmenting the basis functions with terms incorporating the expected dependence on the spectral index enables the efficient approximation of high precision primordial bispectra.

3.4.1 Large non-physical contributions

When a shape is dominated by its non-physical configurations (those that do not obey the triangle inequality (2.5)) then we can see slow convergence on the tetrapyd, despite possibly fast convergence on the cube as a whole. We will refer to convergence on the tetrapyd as $\mathcal{E}^{tetra}$ and the convergence on the cube as $\mathcal{E}^{cube}$. For a pure oscillation we see no significant difference between the convergence on the cube and on the tetrapyd, as expected for a shape where the physical and non-physical configurations are of the same order of magnitude. For an oscillation with a DBI shape however, we see a large difference, as shown in figure 3.3. On the cube we see fast, monotonic convergence—the *resonant* basis and *scaling* basis perform well, quickly bring $\mathcal{E}^{cube}$ below 0.1%. On the tetrapyd however, we see that $\mathcal{E}^{tetra}$ only drops below 0.1% for $p_{\max} < 30$ for the *resonant* basis.

We also see that the improvement in $\mathcal{E}^{tetra}$ is no longer monotonic, there are regions where increasing $p_{\max}$ results in a larger $\mathcal{E}^{tetra}$. HERE Explain this better!! 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. CHECK THIS!!

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 (3.15), $S_2 \gg S_1$ everywhere on the cube.

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

3.5 Setting up a basis

3.5.1 Legendre and Fourier as 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, and augment this set with other 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 as we describe below.

3.5.2 Augmentation

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. This means that while $\mathcal{P}_0$ with $p_{\max} = 20$ is a subset of $\mathcal{P}_0$ with $p_{\max} = 30$, $\mathcal{P}_1^{n_s}$ with $p_{\max} = 20$ is not a subset of $\mathcal{P}_1^{n_s}$ with $p_{\max} = 30$, as the orthogonalised term in the former will include a part non-orthogonal to (e.g.) $\mathcal{P}_{25}$ but it will not include that part in the latter.

Define Orth notation, etc.

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 (2.35), 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 (2.32) 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.

3.7 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 (3.28) 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 (3.25)) with

$$q_{p_{\max}-2}(k) = \mathrm{Orth}\left[k^{-1}\right], \tag{3.30}$$
$$q_{p_{\max}-1}(k) = \mathrm{Orth}\left[\ln(k)k^{-1}\right]. \tag{3.31}$$

This is motivated by

$$k^{\epsilon} = e^{\epsilon\ln\left(\frac{k}{k_{\min}}\right)+\epsilon\ln(k_{\min})} \tag{3.32}$$
$$\approx e^{\epsilon\ln(k_{\min})}\left(1+\epsilon\ln\left(\frac{k}{k_{\min}}\right)\right) \tag{3.33}$$

which is valid when $\epsilon\ln\left(k_{\max}/k_{\min}\right)$ is sufficiently small. Since for the DBI case the scaling exponent ϵ is expected to be of the same order as $n_s - 1$ and ε_s, we see that this approximation is good for our case, where $\ln\left(\frac{k_{\max}}{k_{\min}}\right) \approx 6.91$. 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).

As we can see in figure 3.4 the *scaling* basis converges quickly to a DBI template with a realistic scaling, outperforming the $\mathcal{P}_1^{n_s}$ basis. For the feature templates however, 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.

3.8 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 [12]. 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 [2] in that it also includes a factor of $\frac{1}{\sqrt{k}}$ to retain orthogonality. We define the nth basis element as

$$\frac{\mathcal{P}_n(\bar{k})}{\sqrt{k}}, \text{ with } \bar{k} = \frac{2\ln(k) - \ln(k_{\min}k_{\max})}{\ln(k_{\max}) - \ln(k_{\min})}. \tag{3.34}$$

We then see that

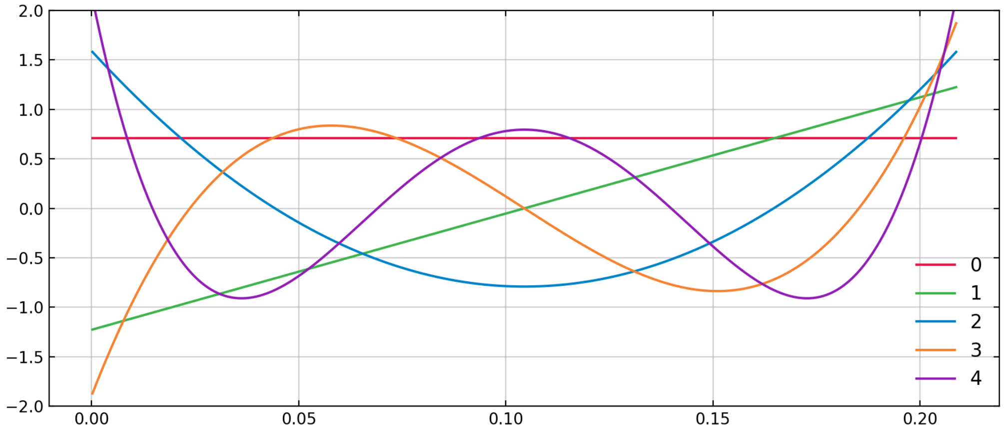
$$\int_{k_{\min}}^{k_{\max}} dk \frac{\mathcal{P}_m(\bar{k})}{\sqrt{k}} \frac{\mathcal{P}_n(\bar{k})}{\sqrt{k}} = \int_{\ln k_{\min}}^{\ln k_{\max}} d\ln k \mathcal{P}_m(\bar{k})\mathcal{P}_n(\bar{k}) \tag{3.35}$$

$$\propto \delta_{mn} \tag{3.36}$$

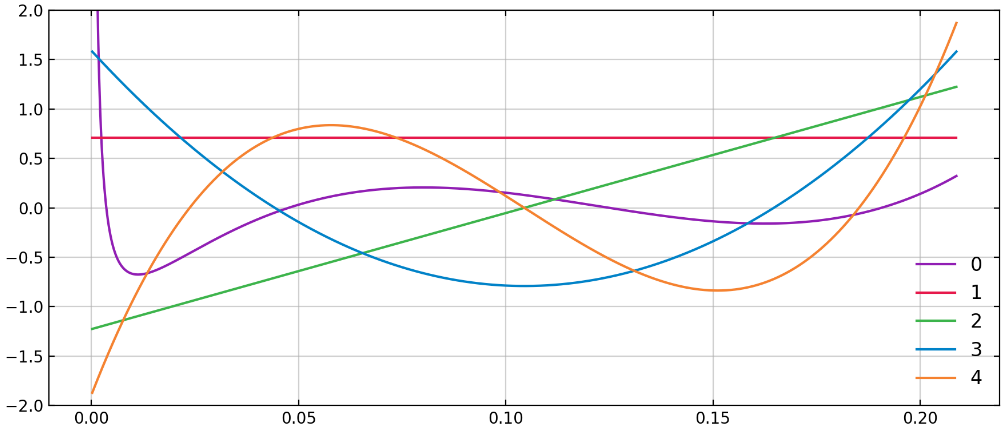
so the *resonant* basis is orthogonal due to the way we defined $\bar{k}$ for this basis set.

In figure 3.5 we see that for linear oscillations, as expected, the *resonant* basis converges for a much smaller range for frequencies (for fixed basis set size $p_{\max} = 30$) than the other basis sets we test. For logarithmic oscillations however, we see that the range of accessible frequencies is nearly doubled. 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 that 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 $p_{\max} = 30$.

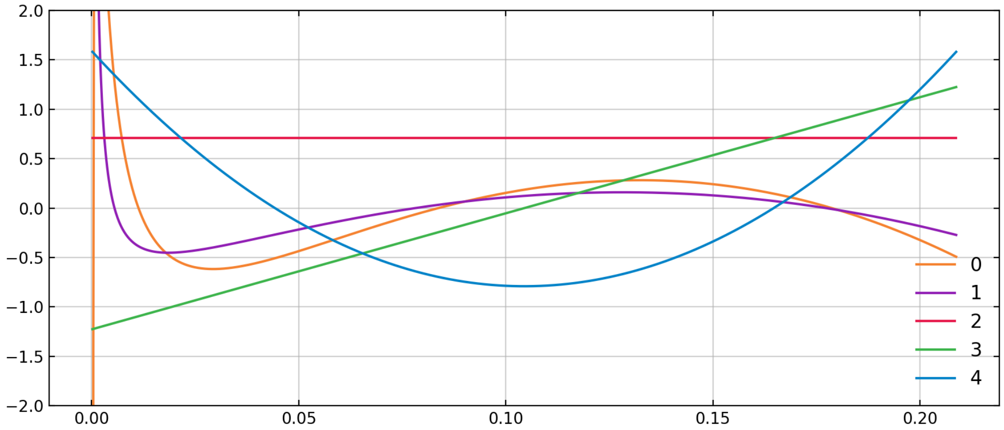
We also note the surprising result in figure 3.4 that the *resonant* basis performs best at converging to the scale-dependent DBI template. However, the convergence of the *scaling* basis is still perfectly adequate in this case.



(a) $\mathcal{P}_0$

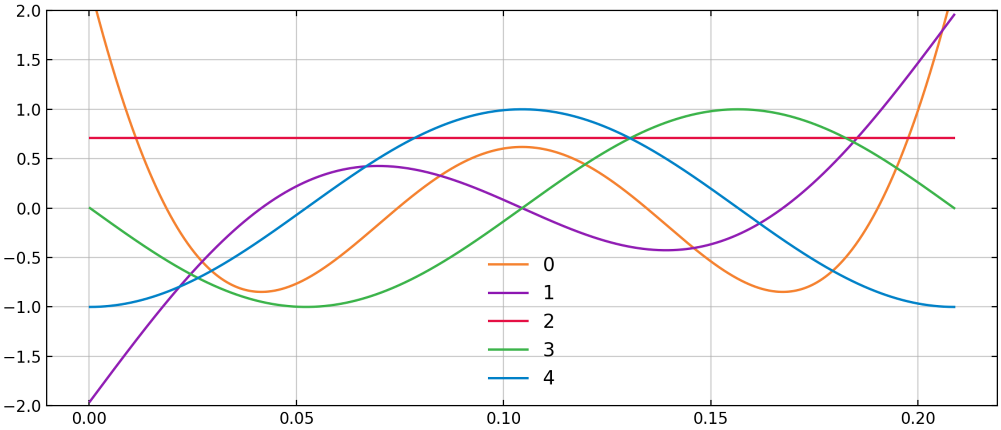


(b) $\mathcal{P}_1$

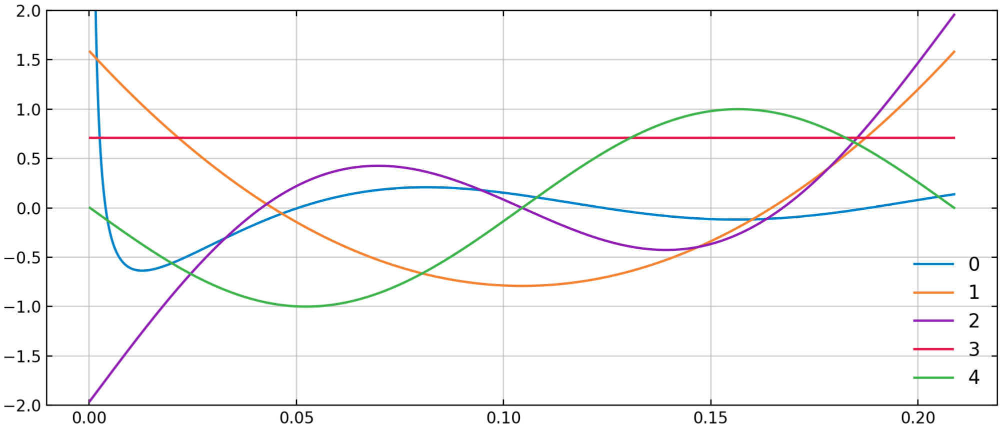


(c) *scaling* basis

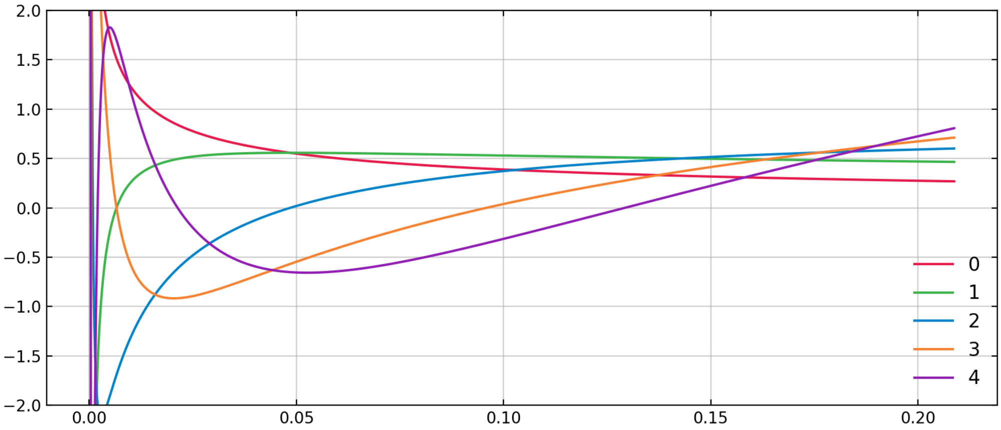
Figure 3.7



(d) $\mathcal{F}_0$



(e) $\mathcal{F}_1$



(f) *resonant* basis

Figure 3.7: We plot various orthogonal basis sets from table 3.1, for $p_{\max} = 5$. Note that $\mathcal{P}_0$, $\mathcal{P}_1$ and the *scaling* basis have Legendre polynomial basis elements in common, but differ in which functions they are augmented by. The *resonant* basis does not have a constant basis element due to the factor of $1/\sqrt{k}$ added to retain orthogonality.

3.9 Factor basis?

Words

3.10 Higher $p_{\mathbf{max}}$

Going beyond $p_{\max} = 30$. How high can I go, in terms of parameters that can't be compared to Planck?

3.11 Conclusion

3.11.1 FIGURES: Coefficient plot for $\mathcal{P}_0$, $\mathcal{P}_1$, *resonant*

Refit mixed basis functions to pure Legendre's subtract to show convergence of rest?

3.11.2 My new basis sets, and their dramatic improvement

Take from above.

3.11.3 Nice table with descriptions and some single-number comparison on examples

Make table. 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 and Fourier series basis sets 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 as ideal solution for a pipeline which requires a single basis that can capture the broadest range of shapes, including different scalings. To improve upon this situation, 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 in the alternative way forward of performing two runs of CMB-BEST, i.e. doing both the scaling basis (for linear oscillations) and the resonant basis (for log oscillations).

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

Chapter 4

Methods and Validation

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 they have converged to the attractor solution before the start of the numerical in-in integration, see for example figure 4.2. When the modes are deep in the horizon they are highly oscillatory, but as they evolve and cross the horizon they 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 (3.8), as in (3.4) we only keep the imaginary part, and the integrand of (3.2) will either be real or vanish. In each of these regimes we must use different numerical set-ups to ensure the efficient calculation of an accurate result.

4.2 Starting the integration with a pinch

While this method has extra suppression at early times compared to the usual in-in calculation (3.8) still converges slowly in its limit at early times. Due to the fact that the integrand is highly oscillatory with a slowly varying amplitude here (so long as we are earlier than any relevant features or resonances), this regime will not significantly contribute to the final result, but if neglected will cause errors that will swamp the true result. To overcome this inefficiency we will discuss in detail how to start this time integration.

The previous two sections detailed the methods required to calculate the coefficients α_n in (3.1), obtaining an explicitly separable expression for the shape function. In this set-up, there are two kinds of integrals we must compute: integrals over time of the form (3.8), and integrals over k of the form (3.6). The first is done once per coefficient for each vertex, the second is a decomposition done once for ζ_k and ζ_k' each, every timestep. In this section we detail how to numerically evaluate these integrals accurately and efficiently.

Since calculating each point in the time integrand requires a decomposition (3.6), which is highly oscillatory at early times, it is worthwhile to consider how to perform the time integral efficiently. 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. A second important consideration comes from how early we sample 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 (3.8) that we must calculate have $\tau = -\infty$ as their lower limit. The highly oscillatory nature of the mode functions in these early times ($|k\tau_s| \gg 1$) suppresses the coefficients of our basis expansion by a factor of $1/\tau_s$. As noted in [2], 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 more precisely by considering the plane wave expansion (e.g. [59]):

$$e^{-i\bar{k}(k_{\max}-k_{\min})\tau/2} = \sum_{n=0}^{\infty}(2n+1)i^n P_n(\bar{k}) j_n(-(k_{\max}-k_{\min})\tau/2) \tag{4.1}$$

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 (2.18) expanded in Legendre polynomials (and similar) give us suppression of $1/\tau^3$ in (3.8).

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 [25] 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 [26] 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 [25, 26], we can safely use the simpler first method.

We understand this situation by taking advantage of asymptotic behaviour of highly oscillatory integrals (for a review see [60]). Since the leading order term depends on the value of the non-oscillatory part only at the endpoints, and the next-to-leading order correction depends on the derivative only at the endpoints, we can approximate the integral $\int_{-\infty}^{T} f(\tau_s)e^{iw\tau_s}d\tau_s$ by replacing the non-oscillatory part $f(\tau_s)$ with a function with matching value and derivative at $\tau_s = T$, but which converges far faster. We use

$$f(\tau_s)e^{-\beta^2(\tau_s-T)^2}, \tag{4.2}$$

for $\tau_s < T$. In this way, for sufficiently large T, we obtain the accuracy of the first two terms of the asymptotic expansion ($O(\beta^2/w^2)$, $w = 3k_{\text{max}}$) without needing to explicitly calculate the derivative at T, or needing any phase information (as one would need to accurately impose a sharp cut on the integrand).

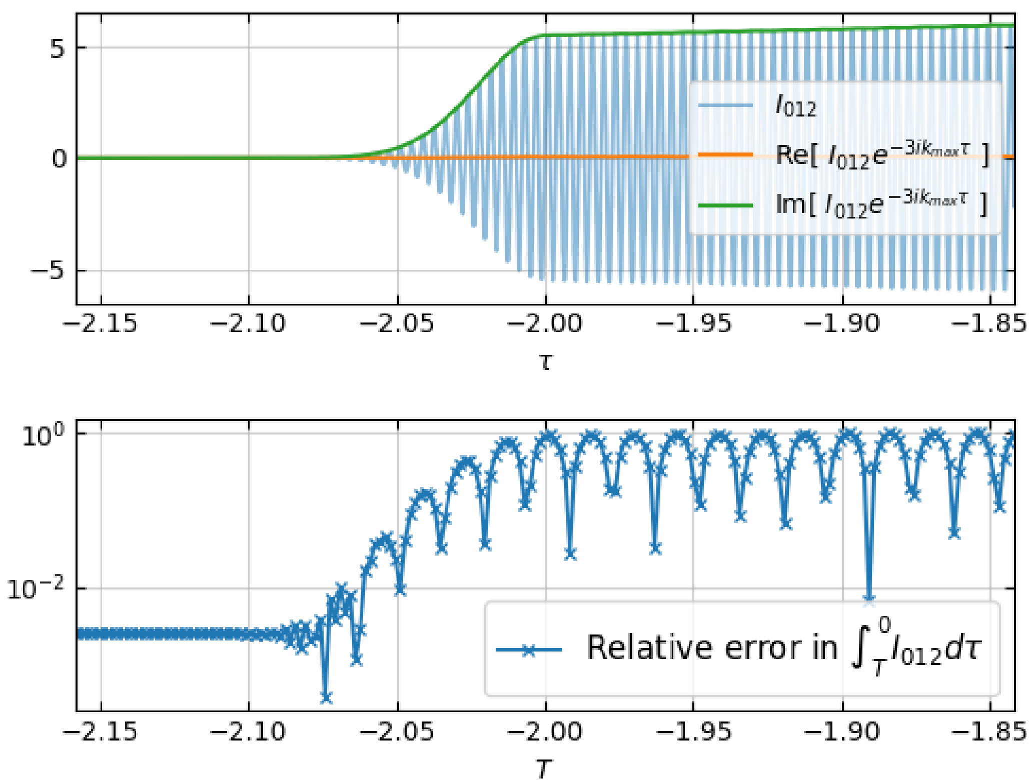
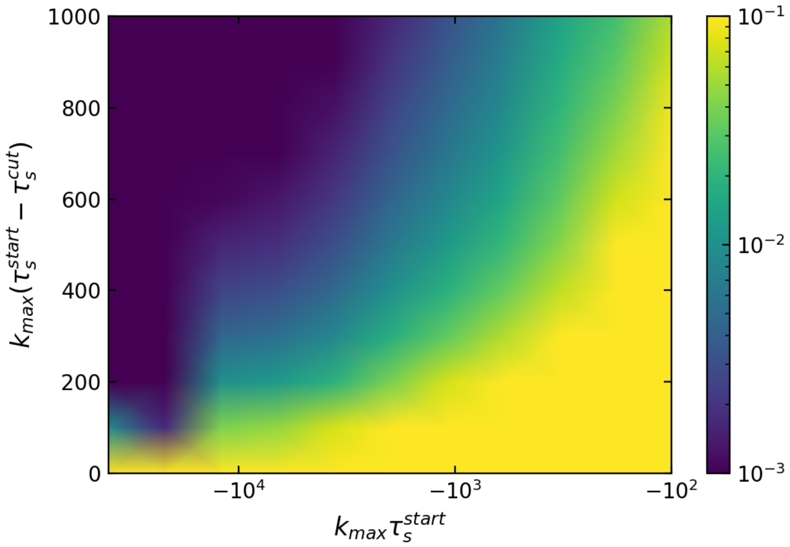
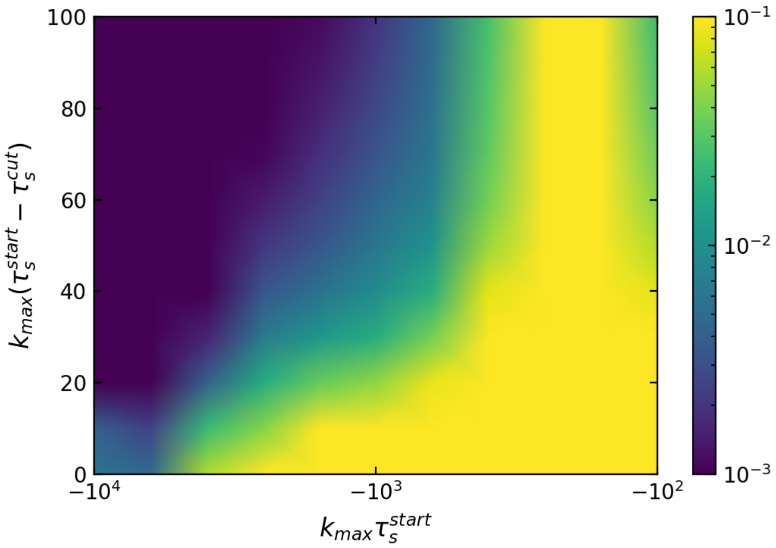


Figure 4.1: A toy example demonstrating the considerations involved in performing the time integrals (3.8). By carefully starting the time integrations, using the form (4.2), we can avoid errors that would otherwise swamp our result. The coefficient being calculated is the α_{012} coefficient of the $\mathcal{P}_0$ expansion of (2.9).

We use a damping of the form $e^{-\beta^2(\tau_s-T)^2}$ for $\tau_s < T$ to smoothly set the integrand to zero before a certain initial time, T. As long as T is sufficiently early and β is not too



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



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

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

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 $H_{int} = (-1/\tau)\dot{\zeta}^3$, as in (2.9).

We also demonstrate this for a realistic DBI model and resonance model in figure 4.2. Here we reparametrise the damping in a form that with a more obvious physical interpretation. We use the form $e^{-100\left(\frac{\tau_s^{start}-\tau_s}{\tau_s^{start}-\tau_s^{cut}}\right)^2}$. The point $\tau_s = \tau_s^{start}$ is the earliest point we calculate the integrand for, i.e. before this point we set the integrand to exactly zero—we see that at this point the integrand is suppressed by a factor of e^{-100}. In figure 4.2 we see that $\tau_s^{start} - \tau_s^{cut} = 0$ gives unacceptably large errors across the range of the scan, but for sufficiently large τ_s^{start} this is ameliorated by a positive $\tau_s^{start} - \tau_s^{cut}$. 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 is more robust. For the resonant scenario (2.27) we have $c_s = 1$, so we expect (from [8]) 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 physically relevant time to be $\tau_s \approx -\frac{0.14}{2(0.02)(2.088\times10^{-4})} \approx -1.7 \times 10^4$. Indeed, in figure 4.2 we see that $k_{\max}\tau_s^{start}$ must be set earlier that -10^3 to correctly capture the feature. CHECK THIS.

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 $k_{\max}$. The k-integrals themselves can be computed quite efficiently since at every timestep the integral is over the same sample points. One can therefore calculate and store the values of the basis functions at each of these points, along with the integration weights which will depend on the distribution of k-samples. The actual integration at each timestep, the calculation of the coefficients in (3.6), then becomes nothing more than a dot product of a time-independent array with the numerically evolved mode functions, for each order up to $p_{\max}$. We have found the best convergence results from distributing the k-samples according to the prescription of Gauss-Legendre quadrature.

To calculate the basis expansion of the bispectrum using the in-in formalism we must first calculate the basis expansion of the mode functions at each timestep (3.6). At early times the mode functions are highly oscillatory, 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. We want an expansion of the form

$$z_k e^{-ik\tau_s} = \sum_{n=0}^{\infty} \alpha_n q_n(k). \tag{4.3}$$

We can obtain this by using standard oscillatory quadrature, if the τ_s dependence of the weights does not add too much overhead. We can also use an expansion of $e^{-ik\tau_s}$ with a known explicit time dependence, for example the expansion (4.1).

To use this second method, the first (smooth) factor z_k can be expanded in whatever basis we are working in, $q_n(k)$, and the second factor (highly oscillatory in k) is expanded in some convenient basis $\tilde{q}_n(k)$ (e.g. $\mathcal{F}_0$, or $\mathcal{P}_0$ using the analytic form (4.1)). Then by precomputing $q_a(k)\tilde{q}_b(k)$ as a linear combination of the set of basis functions $q_c(k)$ all we need calculate at each timestep is the coefficients of the smoother $z_k(\tau_s)$, which we then convert to the coefficients of $F^{(i)}(\tau, k)$. In this way we can retain flexibility in our bispectrum basis, as well as efficiency and precision in the calculation. In the case of using $\mathcal{P}_0$ for the $\tilde{q}_n(k)$, assuming the expansion in (3.12) converges, we need only compute the expansion for $e^{-ik\tau_s}$ to enough terms that the first $p_{\max}$ of the coefficients in the expansion (3.5) of the $F^{(i)}(\tau, k)$ converge, not until the actual sum (3.5) converges, since for high enough orders the integrals in (3.8) will integrate to zero.

Clearly, once τ_s becomes small enough these considerations will no longer be necessary and we can simply decompose the mode function directly. We do this around the horizon crossing of the geometric mean of $k_{\min}$ and $k_{\max}$. If there is an extreme feature which causes a large deviation from the usual slow-roll form this switch will need to be made sooner. Also, this method would need to be adapted for non-Bunch-Davies initial conditions. 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, for example, a parameter scan. Using the methods outlined above, (3.8) and (3.6) can be computed precisely and efficiently in a mostly basis-agnostic context allowing 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.

4.2.1 When to start the integration

Words

4.2.2 Dangers of a sharp start

Show how easy it is to swamp the result with a sharp start.

4.3 Integration weights

There are two central types of numerical integral that we have to perform in this method, (3.8) and (3.6). 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 a decomposition, given the values of the k-modes at these given sample points. These decomposition 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 final coefficients, over the decompositions calculated at each timestep. Thus, it is important that we can efficiently integrate many different functions over the same sample points.

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 $[x_0, x_1, x_2, \ell, x_{N-1}]$. To calculate $\int_{x_{i-1}}^{x_{i+1}} f(x)dx$ take each triplet of sample points $[x_{i-1}, x_i, x_{i+1}]$, and map it to $[\bar{x}_{i-1}, \bar{x}_i, \bar{x}_{i+1}]$, where

$$\bar{x}_{i-1} = -1, \tag{4.4}$$

$$\bar{x}_i = \frac{2x_i - (x_{i-1} + x_{i+1})}{x_{i+1} - x_{i-1}}, \tag{4.5}$$

$$\bar{x}_{i+1} = 1. \tag{4.6}$$

We then calculate

$$A_{pq} = \mathcal{P}_p(\bar{x}_q) \tag{4.7}$$

for $p \in [0, 1, 2]$ 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. $\left(A^{-1}\right)_{0q}$ is the contribution that these segments give to the weight for the sample point x_q. When all the weights w_i have been calculated, the resulting approximation for the integral is simply

$$\int_{x_0}^{x_{N-1}} f(x)dx \approx \sum_{i=0}^{i=N-1} f(x_i)w_i, \tag{4.8}$$

i.e. a simple dot product between the sampled function and the weights, which need only be calculated once for an arbitrary number of functions f.

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 $\int_{x_{i-1}}^{x_{i+1}} \cos(wx)\mathcal{P}_n(\bar{x})dx$ 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 trivial cost of performing an extra dot product.

When to swap, validation. Why negative weights are usually bad.

4.4 The interaction Hamiltonian

(Using integration by parts from RP paper.) 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 [4, 27, 28]

$$\begin{aligned} H_{int}(t) = \int d^3x \Big\{ & -\frac{a^3\varepsilon}{Hc_s^4}\left(1 - c_s^2 - 2c_s^2\frac{\lambda}{\Sigma}\right)\dot{\zeta}^3 + \frac{a^3\varepsilon}{c_s^4}\left(3 - 3c_s^2 - \varepsilon + \eta\right)\zeta\dot{\zeta}^2 \\ & - \frac{a\varepsilon}{c_s^2}\left(1 - c_s^2 + \varepsilon + \eta - 2\varepsilon_s\right)\zeta(\partial\zeta)^2 \\ & - \frac{a^3\varepsilon^2}{2c_s^4}(\varepsilon - 4)\dot{\zeta}\partial\zeta\partial(\partial^{-2}\dot{\zeta}) - \frac{a^3\varepsilon^3}{4c_s^4}\partial^2\zeta(\partial(\partial^{-2}\dot{\zeta}))^2 \Big\} \end{aligned} \tag{4.9}$$

with $\Sigma = \frac{H^2\varepsilon}{c_s^2}$ and $\lambda = X^2 P_{,XX} + \frac{2}{3}X^3 P_{,XXX}$. See [4] for further details.

This is commonly quoted with a term proportional to the equation of motion, but this will never contribute [4, 61, 30, 62]. We do not need to make a slow-roll approximation (the quantities defined in (2.11) 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 [4] (see also [61, 30, 62]) we do not work with any boundary terms.

4.5 Stopping the integration

(A note on boundary terms and difficult (time-dep) cancellations.) Numerically this is preferable to forms with boundary terms, whether they come from undoing a field redefinition or from integration by parts. Since the boundary term contribution will depend on the choice of when to end the integration, its time dependence must cancel with a late-time time-dependent contribution of some vertex, requiring us to track the necessary quantities much longer than otherwise needed to obtain the desired precision.

Schematically, the correction from a field redefinition would look like

$$\langle\zeta_{\mathbf{k_1}}\zeta_{\mathbf{k_2}}\zeta_{\mathbf{k_3}}\rangle = \left\langle\tilde{\zeta}_{\mathbf{k_1}}\tilde{\zeta}_{\mathbf{k_2}}\tilde{\zeta}_{\mathbf{k_3}}\right\rangle + \lambda\left\langle\tilde{\zeta}_{\mathbf{k_1}}\tilde{\zeta}_{\mathbf{k_2}}\right\rangle\left\langle\tilde{\zeta}_{\mathbf{k_1}}\tilde{\zeta}_{\mathbf{k_3}}\right\rangle + cyclic \tag{4.10}$$

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 complication.

By integrating by parts and using the equation of motion, the interaction Hamiltonian can be rewritten without picking up boundary terms [23]. Using (3.7) from [23], with $f = -\varepsilon/(c_s^2 H)$, we obtain the following form:

$$\begin{aligned} H_{int}(t) = \int d^3x \Big\{ & -\frac{a^3\varepsilon}{Hc_s^4}\left(-c_s^2 - 2c_s^2\frac{\lambda}{\Sigma}\right)\dot{\zeta}^3 + \frac{a^3\varepsilon}{c_s^4}\left(-3c_s^2\right)\zeta\dot{\zeta}^2 \\ & - \frac{a\varepsilon}{c_s^2}\left(-c_s^2\right)\zeta(\partial\zeta)^2 - \frac{a\varepsilon}{Hc_s^2}\dot{\zeta}(\partial\zeta)^2 \\ & - \frac{a^3\varepsilon^2}{2c_s^4}(\varepsilon-4)\dot{\zeta}\partial\zeta\partial(\partial^{-2}\dot{\zeta}) - \frac{a^3\varepsilon^3}{4c_s^4}\partial^2\zeta(\partial(\partial^{-2}\dot{\zeta}))^2\Big\}. \end{aligned} \tag{4.11}$$

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 [2] that using (4.9) in a scenario that results in an equilateral shape would require sensitive cancellations in the squeezed limit. Likewise, using (4.11) for a local scenario would require sensitive cancellations in the equilateral limit.

As mentioned in [2], the spatial derivatives can be manipulated into simple prefactors of k_i using the triangle condition ($\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3} = 0$), and so preserve the separability of the result. To absorb these prefactors in our calculation, we precompute $k^p q_a(k)$ as a linear combination of the $q_a(k)$ for the relevant values of p, from which $V^{(i)}_{P\tilde{P}}$ defined in (3.9) 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 H_{int}.

4.6 Validation (on numerical results)

4.6.1 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 (3.15), 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 (3.15), 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 (2.39). The consistency condition should hold most precisely at the configurations with smallest k_L/k_S, the most squeezed being the three corners, $(k_{\max}, k_{\max}, k_{\min})$ and permutations. We want our test to be on an extended region of the tetrapyd however, so we choose the line

$$\frac{k_L}{k_S} = \frac{2k_{\min}}{k_{\max}}, \tag{4.12}$$

which connects $(k_{\max}, k_{\max}, 2k_{\min})$ to $(k_{\max}/2, k_{\max}/2, k_{\min})$. We will take $\frac{k_{\min}}{k_{\max}} = \frac{1}{550}$, so this is still sufficiently squeezed to be a stringent test.

First, we investigate convergence on simple featureless models, both local-type (2.23) and equilateral-type (2.25). We find that in our chosen 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 (3.15). We compare the converged results against analytic templates (2.29) and (2.32), using the full shape information (3.15), 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 (2.26). 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 [7], 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 (2.27). 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 [35], 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 (3.17). We emphasise that while these plots display slices through the tetrapyd, our actual result describes the shape function on the entire three-dimensional volume of the tetrapyd, and we measure our convergence over this whole space.

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

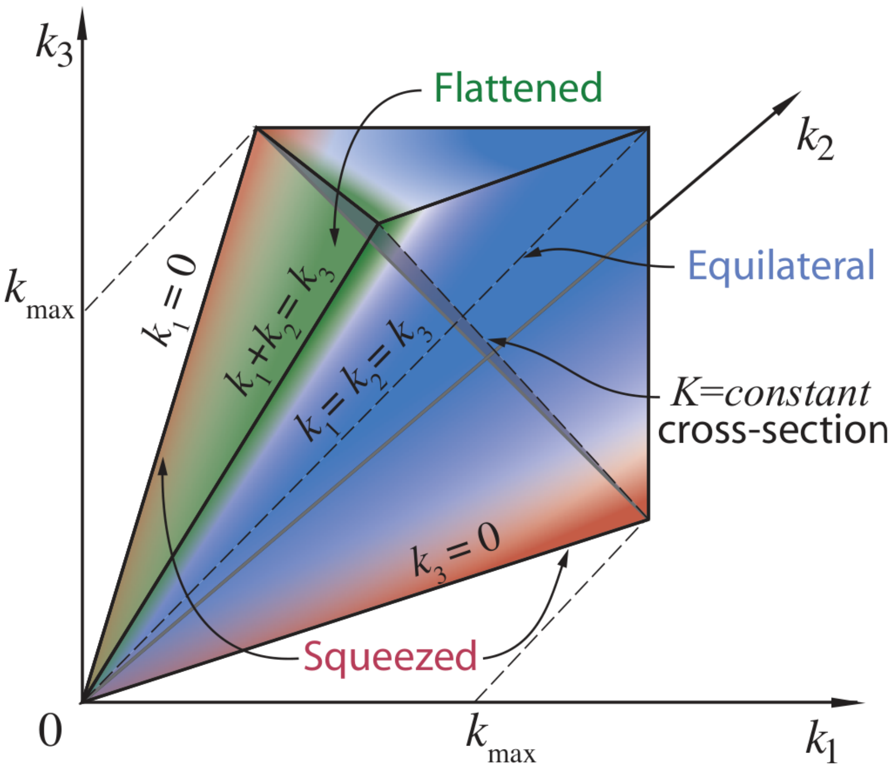
4.6.2 Quadratic slow-roll

The first model we will consider is slow-roll inflation on a quadratic potential (2.23). 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 (2.29) as a precision test. Indeed, we find that it has a scaled relative difference (3.16) of 2.7×10^{-5} with this shape, contrasting a scaled relative difference of 0.077 with the local template (2.31). 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 (2.29) or the local template (2.31), as for $\epsilon \approx 0.8 \times 10^{-2}$ we only expect these templates to match the true result to percent level accuracy. Indeed, we find that our result has a correlation of 0.998 with both (2.29) and (2.31), corresponding (in the sense of (3.16)) 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 (3.8), or in including the spatial derivatives from H_{int}, could have easily swamped the $p_{\max} = 45$ result.



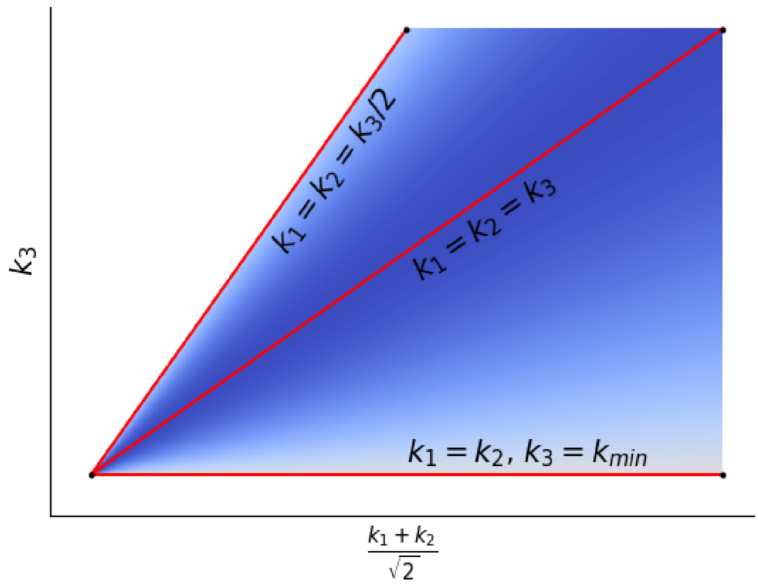


Figure 4.3: For ease of display, we will plot the two-dimensional $k_1 = k_2$ slice of the tetrapyd for each of our validation examples, as shown schematically here. Horizontal lines on this plot have constant k_3. The bottom edge is $k_3 = k_{\rm min}$, the top edge is $k_3 = k_{\rm max}$. The right edge is $k_1 = k_2 = k_{\rm max}$, the left edge is $k_1 = k_2 = k_3/2$, i.e. the limit imposed by the triangle condition. Plotted in red (in the right hand 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?).

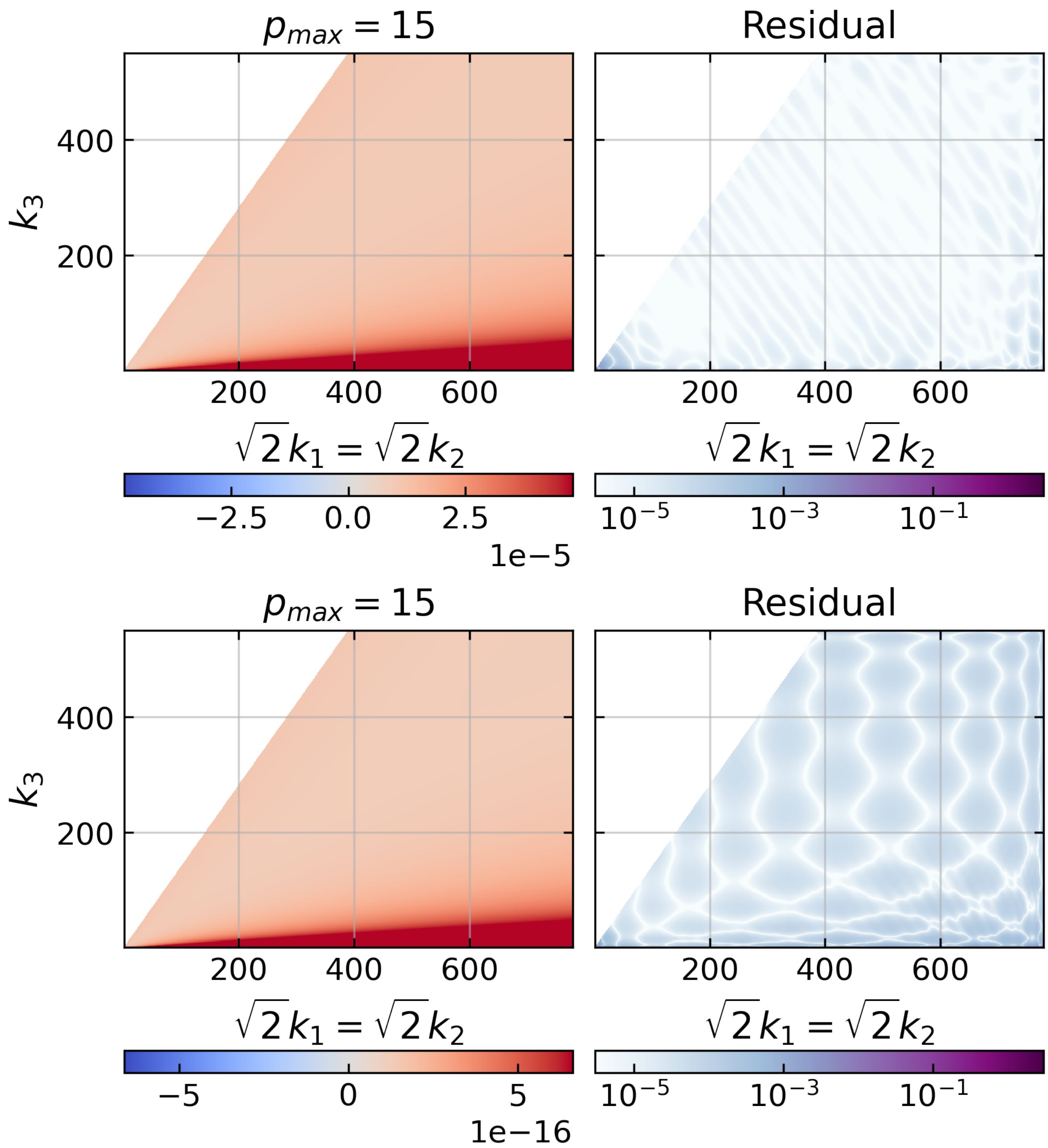


Figure 4.4: A canonical single-field model on a quadratic potential (2.26), slowly-rolling with $\varepsilon \approx 2 \times 10^{-6}$ in the top plot, and $\varepsilon \approx 0.8 \times 10^{-2}$ in the lower plot. This shape is dominated by its squeezed limit, and has a scale dependence determined 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 $p_{\max} = 15$.

4.6.3 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 (2.24) and (2.25). 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 (2.32), and 13% relative to the equilateral template (2.34). 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 (2.35), we get a relative error of 0.27%. On the line defined by (4.12), 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.

4.6.4 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 (2.26) 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 [7], presented there in equations (48), (54), (55). We plot a comparison of the result of [7] 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 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 [2]. They validate

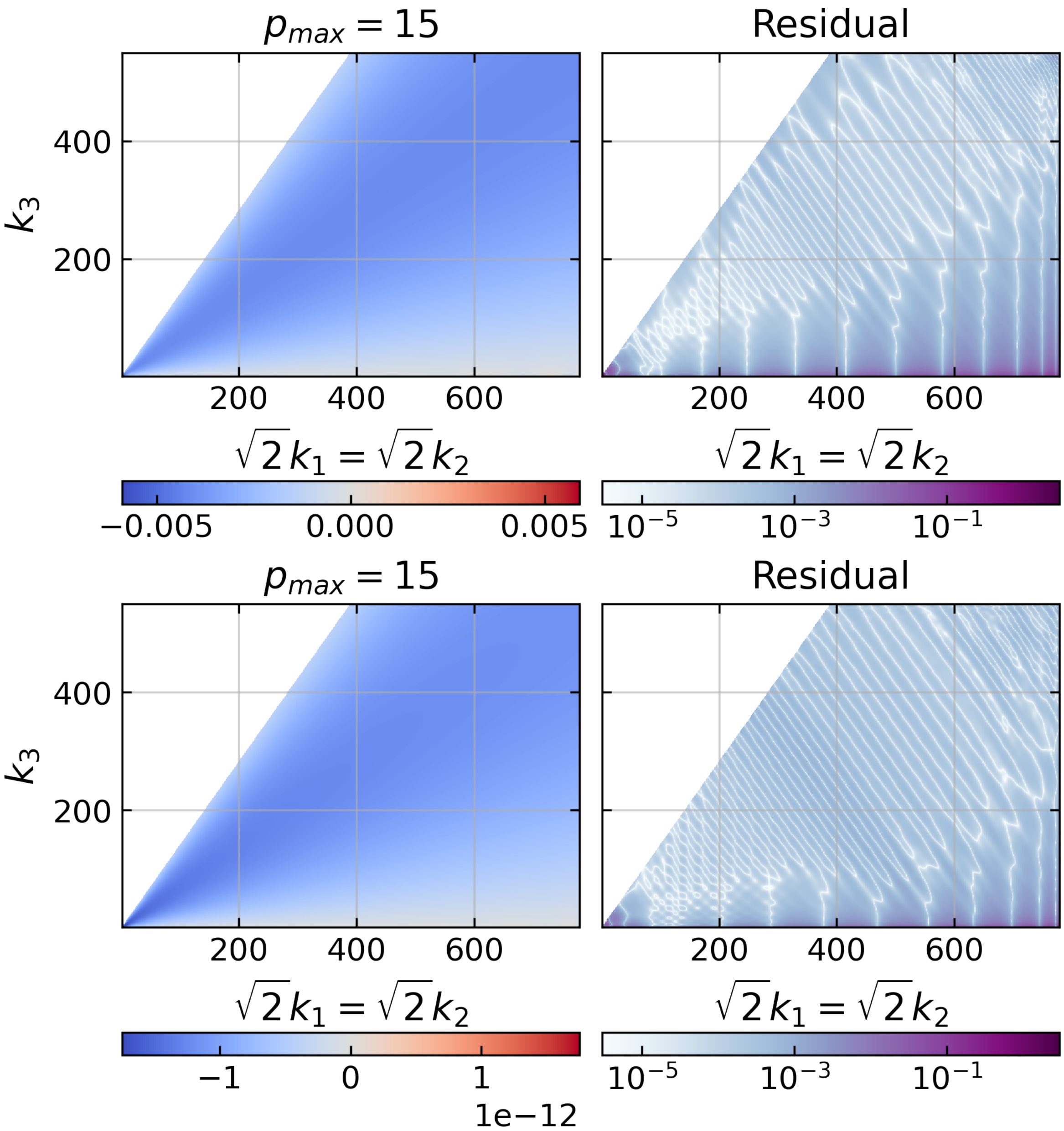


Figure 4.5: The upper plot shows the shape function for a DBI model deep in slow-roll. We set λ_{DBI} in (2.25) to 1.9×10^{18}, obtaining a scenario with $\varepsilon \approx 1.9 \times 10^{-6}$ and $c_s = 2.3 \times 10^{-3}$. This shape is dominated by its equilateral configurations, and has only a 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 (2.25) 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 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.

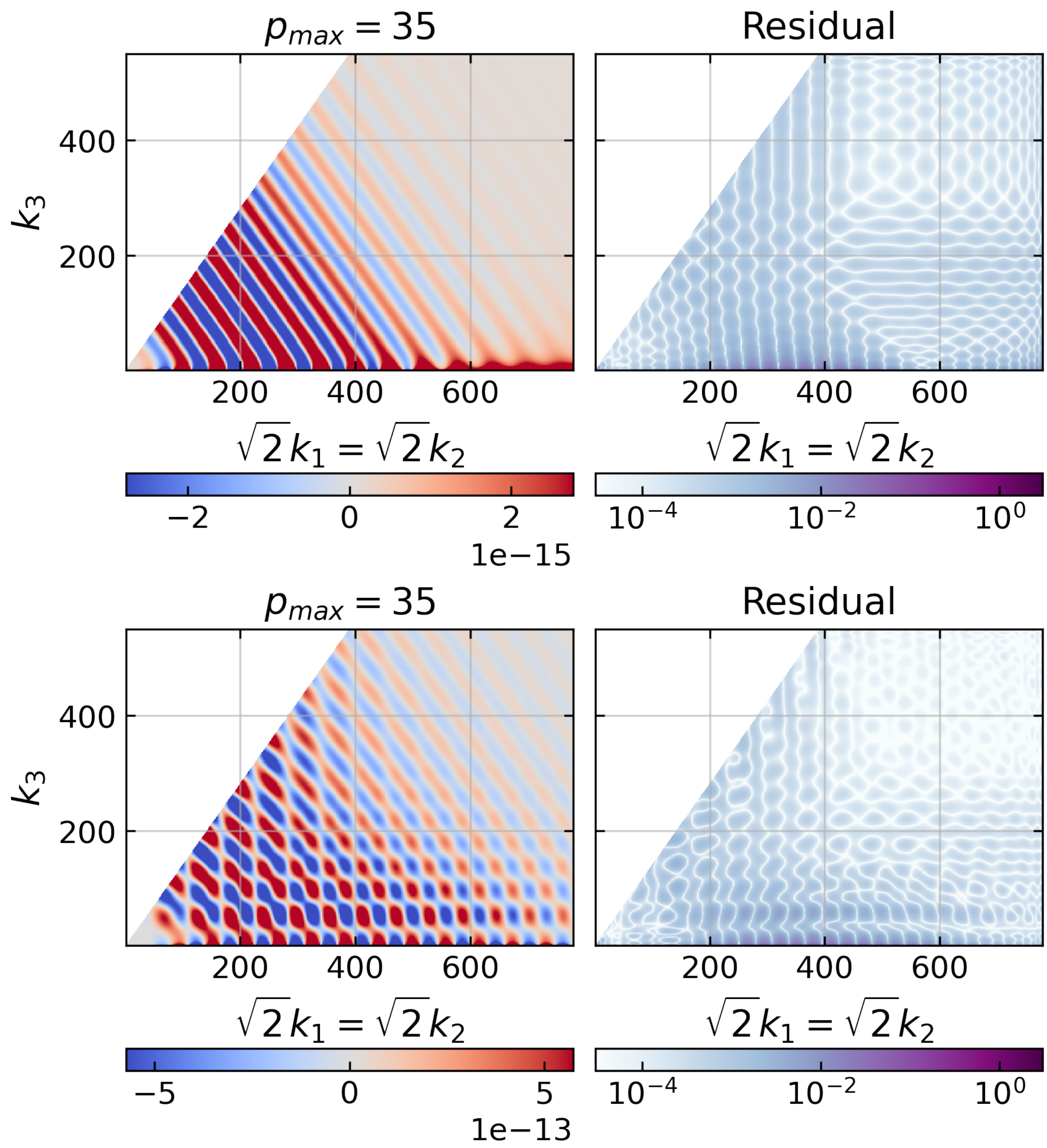


Figure 4.6: The tree-level shape function of a feature model (2.26), shown for step sizes of $c = 5 \times 10^{-5}$ (upper plot) and $c = 5 \times 10^{-3}$ (lower plot). The corresponding expansion parameter values of [7], $\mathcal{C} = 6c/(\varepsilon + 3c)$, are 0.035 and 1.3. For the smaller step 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 $p_{\max} = 35$.

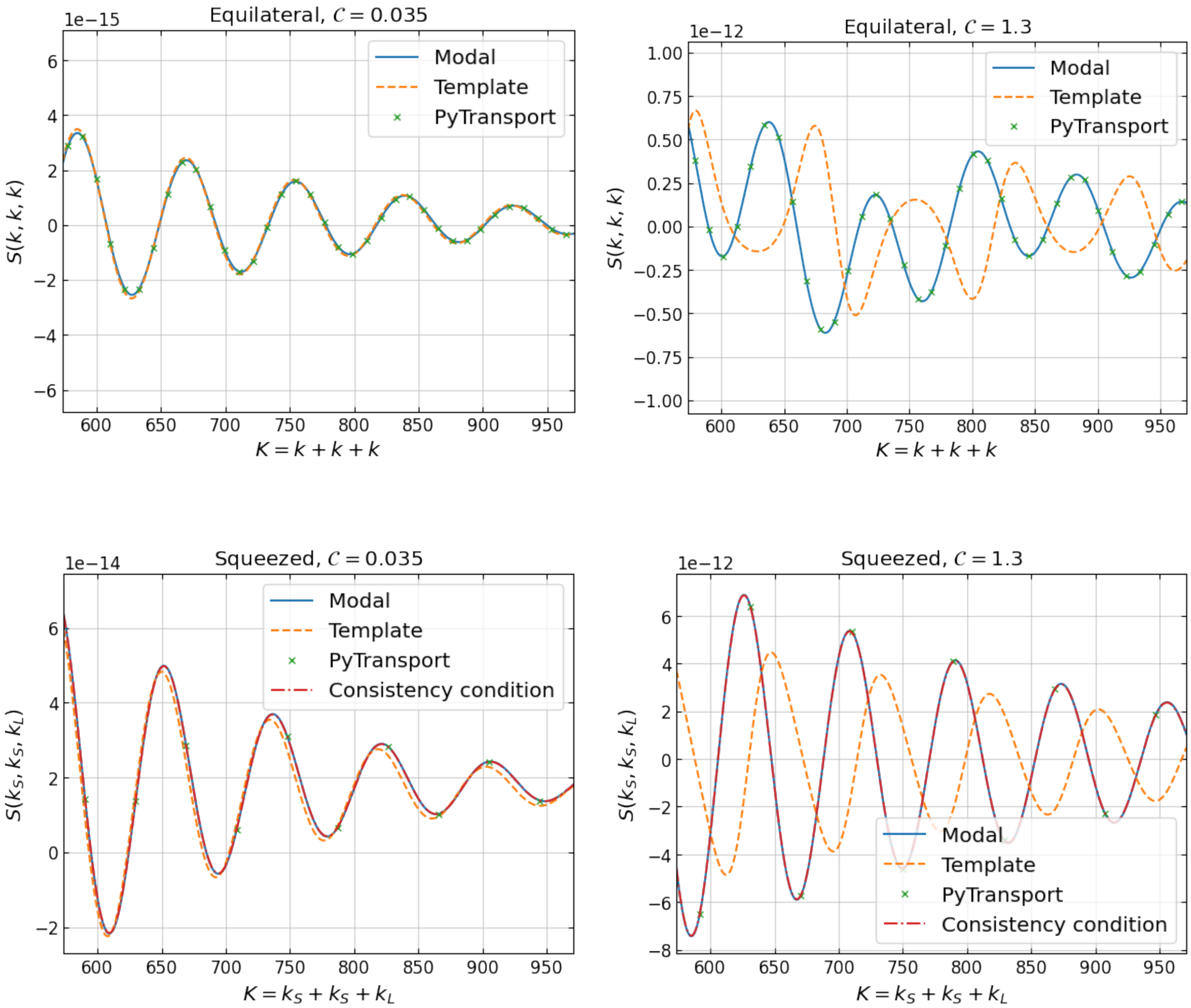


Figure 4.7: In the equilateral limit for the feature models (the top two figures) we validate our modal result against the PyTransport result. In the squeezed limit (the bottom two figures) we validate against PyTransport, and the consistency condition. In both limits, for both step sizes shown, we find excellent agreement. For the small step size (the two plots to the left), we additionally see a good match to the template of [7]. 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.

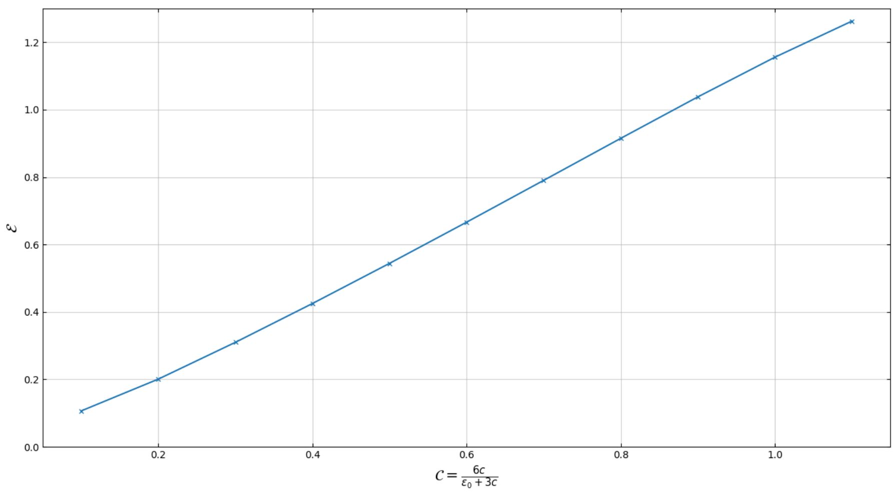


Figure 4.8: We sample more shapes with step sizes between the two feature models shown in figure 4.6. We plot the relative difference, integrated over the full tetrapyd in the sense of (3.15), between the modal result and the analytic template of [7], as a function of the template parameter $\mathcal{C} = \frac{6c}{\varepsilon_0+3c}$ (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 (2.26) occurs, which is kept constant across this scan. The width of the feature was also kept constant.

4.6.5 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 (2.27). 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 $p_{\mathrm{max}} = 65$ and $p_{\mathrm{max}} = 35$) of 0.93%, a squeezed limit test result of 1.1% (along the line defined by (4.12)), 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.

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 [35, 29]. 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%.

4.7 Speed comparison vs PyTransport.

Something BINGO can't do?

4.8 Conclusion

We conclude that our methods have been validated on interesting and realistic example scenarios, for a sufficiently large k-range, $k_{\max}/k_{\min} = 1000$. We have confidence in this due to the varied test we have performed, against templates, other numerical codes, and also against the squeezed limit consistency condition. We have in this work demonstrated separable in-in methods for high orders and features for the first time, going beyond the work done in [2]. 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.

With our methods validated, in the next chapter we will present work performed in collaboration with Wuhyun Sohn, validating the combined Primodal and CMB-BEst pipeline by obtaining a constraint on the sound speed of a DBI inflation scenario.

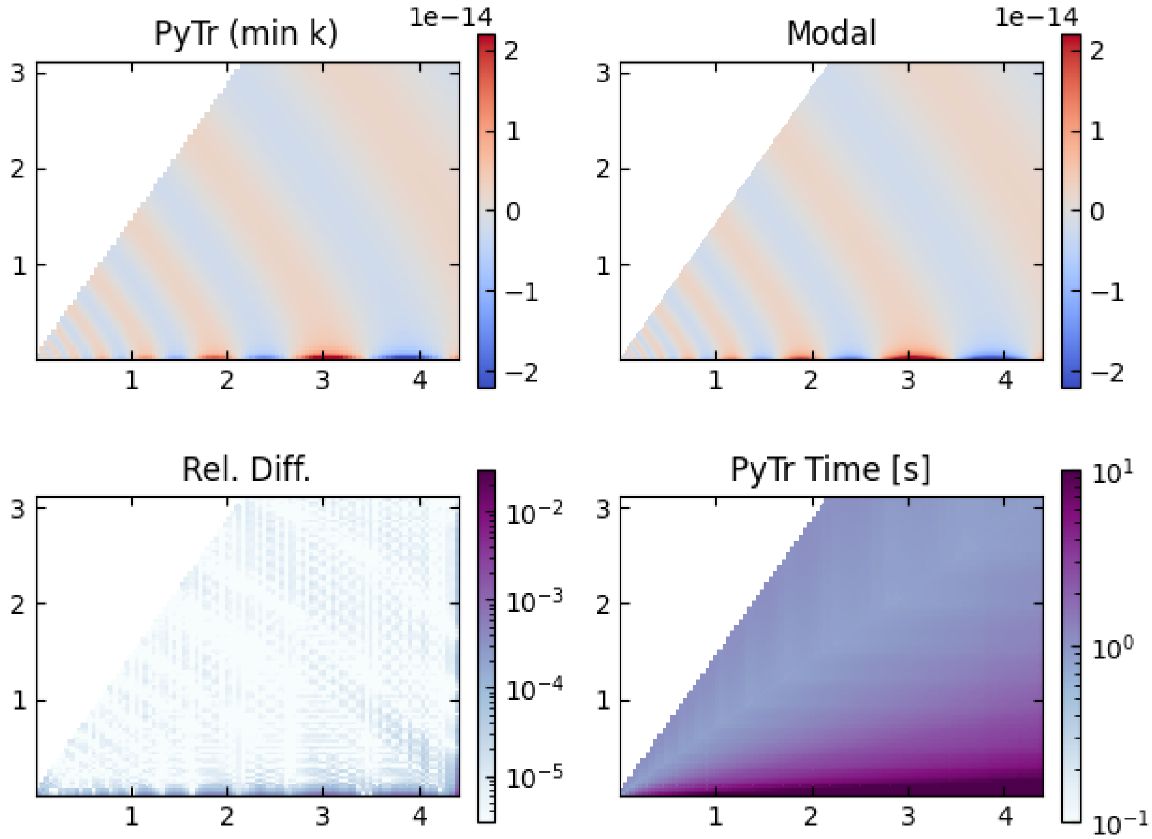


Figure 4.9: Resonance on a quadratic potential (2.27), testing our result using point tests against the PyTransport code. The logarithmic oscillations in the shape function are generated by periodic features deep in the horizon. The differences between our result and the PyTransport result are sufficiently 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 $p_{\max} = 35$.

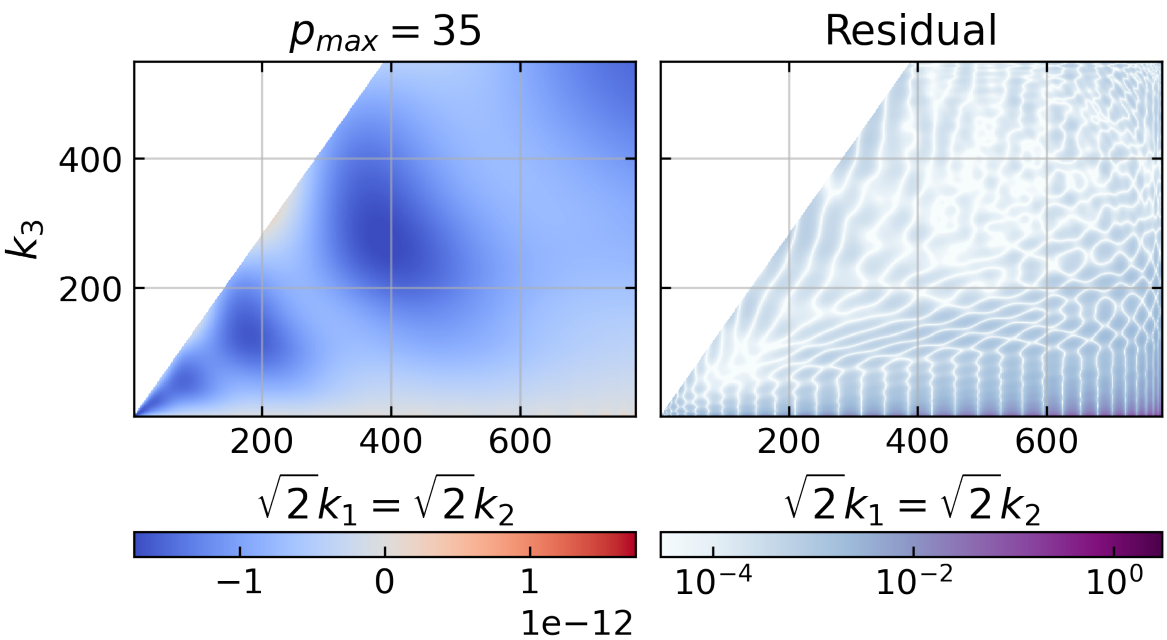


Figure 4.10: Non-Gaussianity generated by periodic features in a DBI model, including a phase difference in the flattened limit as described in [35]. For the purposes of demonstrating the phenomenology, we have placed an envelope on the oscillations in the potential to aid convergence. In the $\mathcal{P}_{01}^{n_s}$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

Chapter 5

Constraints TO DO

5.1 Connecting to CMB-BEst

The goal here is to validate our pipeline by reproducing the *Planck* result from [12], equation (55):

$$c_s^{DBI} \geq 0.079 \qquad (95\%,\ T \text{ only}). \tag{5.1}$$

We would not expect to reproduce this exactly due to Lack of maps? Different approximations? Numerics? Though we can estimate a reasonable disagreement by looking at the expected scatter between the Planck estimators, and by calculating σ??.

5.2 Set-up of the scan

We use equation (2.25), with $\beta_{IR} \in [0.1885, 0.58]$. We find that $\beta_{IR} = 0.331$ produces $c_s^* = 0.0794$, which is close to the above constraint. Thus, we roughly expect to find that $\beta_{IR} < 0.331$ is ruled out (for all the other scenario parameters held fixed). We also have that $\ln(10^{10} A_s)$ for each scenario is within 3.044 ± 0.014 across the scan, and that n_s^* is within 0.9649 ± 0.0042 across the scan. This set-up produces a scenario with ($*$ denotes horizon crossing of the pivot scale):

β_{IR}	c_s^*	ε_s^*	ε^*	n_s^*	n_{NG}^*
1.89×10^{-1}	1.39×10^{-1}	8.57×10^{-3}	7.44×10^{-5}	9.650×10^{-1}	-8.72×10^{-2}
5.80×10^{-1}	4.50×10^{-2}	8.67×10^{-3}	2.31×10^{-4}	9.637×10^{-1}	-8.99×10^{-2}

Table 5.1: Summary of scenario parameters across the scan.

β_{IR}	ϕ^*	H^*	η^*
1.89×10^{-1}	5.19×10^{-1}	1.31×10^{-6}	2.63×10^{-2}
5.80×10^{-1}	5.15×10^{-1}	1.30×10^{-6}	2.71×10^{-2}

Table 5.2: Summary of scenario parameters across the scan.

We now 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}

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 (2.36) and product (2.35) scaling templates, in figure 5.1, 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 (2.29)) which are neglected in the DBI templates.

5.3 Compare convergence at primordial level to convergence at f_{NL} level

We will compare primordial convergence to CMB convergence, by comparing the $p_{\max} = 29$ and $p_{\max} = 30$ results. This will show that the CMB convergence is slightly better, i.e. that the squeezed limit (which is where the primordial shape converges most slowly) is suppressed. For DBI, and also for DBI resonance?

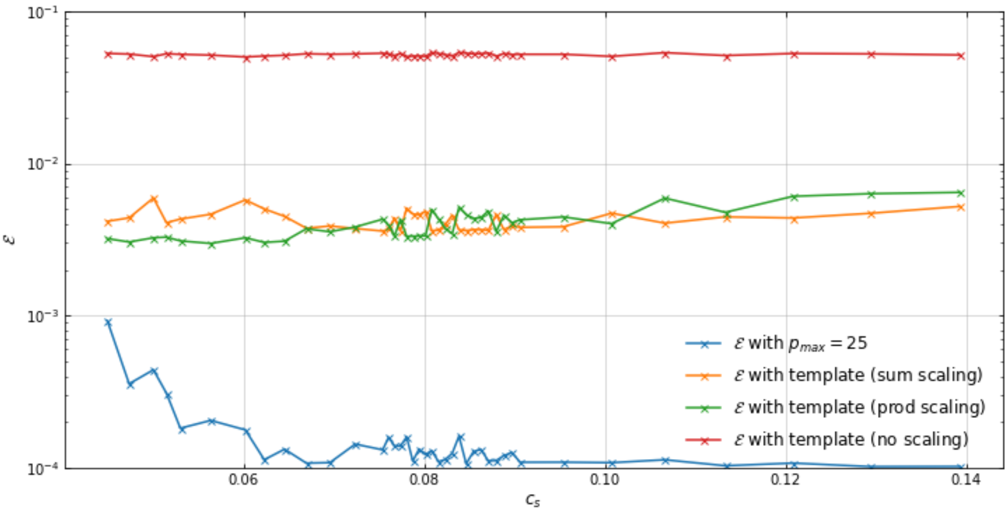


Figure 5.1: The *scaling* basis converges well across the scan range. We see that the bare DBI template is a poor match to the true numerical result. This is mostly due to the error in the overall magnitude. Once this is corrected, we see that the numerical result matches the approximate template to 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 (2.36) and the product scaling (2.35) perform comparably in matching the numerical result. This is mostly due to those templates neglecting the usual slow-roll suppressed contributions (as in (2.29)), which do in fact become relevant to the primordial bispectrum deep enough into the squeezed limit, due to their local-type shape.

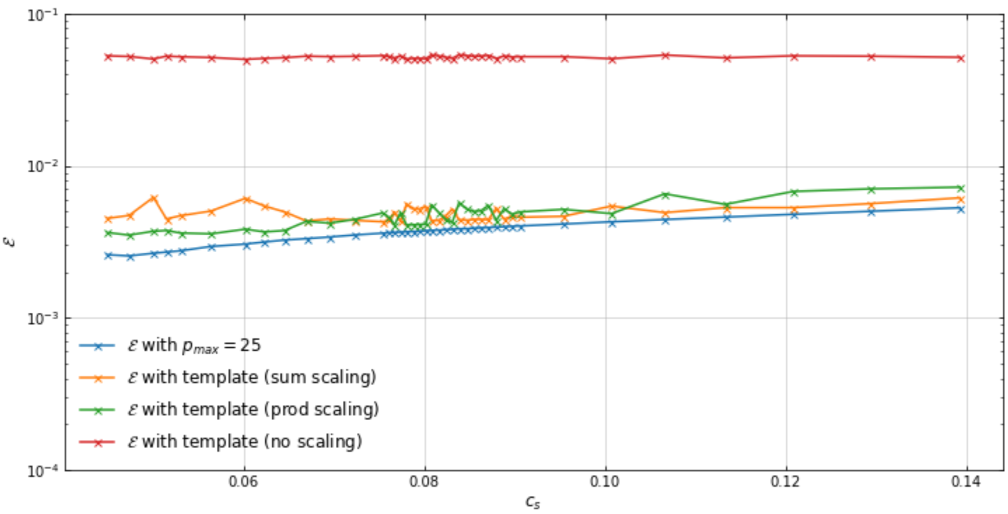


Figure 5.2: The $\mathcal{P}_1^{n_s}$ basis is sufficiently convergent across the scan range to obtain the desired constraint. We see that the convergence error is only slightly better than the error in the slow-roll corrected templates.

5.4 Validate this convergence by reproducing Planck constraint using DBI template decomposition.

Words.

5.5 Then, check how the scaling from the real numerical result affects this.

Here we will compare template decompositions to Primodal results. This will tell us how large the slow-roll corrections are to the final CMB result, but not anything about the primordial convergence. The main slow-roll corrections are a correction to the amplitude, and a deviation from perfect scale-dependence.

5.6 Higher $p_{\mathbf{max}}$

Going beyond $p_{\max} = 30$. How high can I go, in terms of parameters that can't be compared to Planck?

Chapter 6

Future Work/Conclusions

6.1 Further SF constraints

6.1.1 (Depending on what gets done.)

Words

6.2 Multi-field

6.2.1 Specific models, parameter goals.

Words

6.2.2 Notes on implementation.

Words

6.3 Factor basis to assess limits

6.3.1 (Unless this is already done)

Words

6.3.2 Map shape space?

Words

6.4 User guide for Primodal code, public release of Primodal

Not relevant to thesis, but could add some short notes on plans.

6.5 Discussion

In this work we have extended the modal methods of [19, 20, 21] to recast the calculation of the tree-level primordial bispectrum (2.8) 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 (2.3), the integral (2.56) 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 [2] 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 (3.8) 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, $\mathcal{P}_{01}^{n_s}$.

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 [12]. 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 [1]. 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.

6.6 PhD project

The standard cosmological history, the ΛCDM 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 ΛCDM 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 [12], 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 [9].

The formalism used to calculate inflationary non-Gaussianities within an inflation scenario is known as the in-in formalism [22]. 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 [21], and building on the basic idea of [2], 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 [1]) 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) [63], 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 [5] with sharp or oscillatory features, which are reviewed in, for example, [55].

6.7 Research proposal

6.7.1 General Aims

My research goals are to develop methods which will allow us to connect the fundamental physics of the early universe to observations. My most immediate goal is to connect single-field models of inflation with the full three-point correlations of the *CMB*. Beyond that, my goal is to generalise these methods to theories with multiple fields (which are well-motivated by "string swampland" considerations [64]), and later to connect the same theories to other observables coming from large-scale structure.

My PhD has focused on the first of these goals, calculating the bispectrum of the inflationary curvature perturbations, in a way tailored to be linked to observations. I have developed methods to do this in the case of single field inflation. A prime goal I would pursue in this RA position would be to extend these methods (and their implementation in code) to the case of multiple fields. This would open the path to placing constraints on this more general class of theories.

Readying Primodal for a public release is another immediate goal. This would benefit the community by providing it with a powerful tool that can capture the full shape information of a wide class of models far more efficiently that previous similar codes, as described in [63].

These goals support those laid out in the recent community white paper [54], 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.

I will do this work in collaboration with Wuhyun Sohn, Dr James Fergusson and Prof. E.P.S.Shellard.

6.7.2 Specific Objectives

6.7.2.1 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 [65, 66]. Detecting these signatures in the CMB bispectrum requires an understanding of their form [9], 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 [12], include string theory-inspired models such as DBI inflation [33] and axion-monodromy [67, 68] (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.

6.7.2.2 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 [64], 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 [51]. 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 [69, 50]), making the exploration of the case of multiple fields a prime target.

6.7.2.3 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 [63]. 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.

6.7.2.4 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.

6.7.3 Methodology

6.7.3.1 Non-linearities from inflaton self-interactions

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 [12] using a scale-invariant approximate template. We will also explore oscillatory models, obtaining constraints using the full shape dependence, complementing the searches done in [12] 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.

6.7.3.2 Non-linearities from multi-field inflation

While some aspects of multi-field scenarios have been previously explored numerically [50] using the transport method [48] 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.

6.7.3.3 Optimising the separable decomposition of primordial bispectra

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.

6.7.3.4 Public release of Primodal

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.

6.7.4 Work plan

- Obtain single-field constraints using the already-validated codes Primodal and CMB-BEst.
- Exploit tensor rank decomposition for basis choice optimisation, and explore tools from other fields, e.g. signal processing.
- Explore further applications to single-field models, for example time dependence within the EFT of inflation.
- Prepare Primodal for public release in its single-field form.
- Test the basis sets that were used in the single-field case on multi-field templates. If needed, explore new basis sets to cover multi-field shapes.
- Generalise the methods for separable in-in calculation to the multi-field case.
- Prepare single-field constraint results for publication.
- Implement multi-field methods in code, extending Primodal.
- Investigate multi-field phenomenology, with the goal of understanding which scenarios predict observable non-Gaussianity, and thus can be constrained.
- Explore the application of these methods in other contexts, such as preparing non-Gaussian initial conditions for simulations of large scale structure evolution.
- Prepare the multi-field extension of Primodal for release.
- Prepare multi-field results for publication.

References

[1] W. Sohn and J. R. Fergusson, *High-resolution CMB bispectrum estimator*, *(in preparation)* .

[2] H. Funakoshi and S. Renaux-Petel, *A Modal Approach to the Numerical Calculation of Primordial non-Gaussianities*, *JCAP* **1302** (2013) 002 [arxiv/1211.3086].

[3] J. Maldacena, *Non-gaussian features of primordial fluctuations in single field inflationary models*, *Journal of High Energy Physics* **5** (2003) 013 [astro-ph/0210603].

[4] 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].

[5] 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].

[6] C. T. Byrnes and K.-Y. Choi, *Review of local non-gaussianity from multifield inflation*, *Advances in Astronomy* **2010** (2010) 1–18.

[7] 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].

[8] R. Flauger and E. Pajer, *Resonant Non-Gaussianity*, *JCAP* **1101** (2011) 017 [1002.0833].

[9] E. Komatsu, D. N. Spergel and B. D. Wandelt, *Measuring primordial non-gaussianity in the cosmic microwave background*, *The Astrophysical Journal* **634** (2005) 14–19.

[10] M. Münchmeyer, F. Bouchet, M. G. Jackson and B. Wandelt, *The komatsu spergel wandelt estimator for oscillations in the cosmic microwave background bispectrum*, *Astronomy & Astrophysics* **570** (2014) A94.

[11] 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].

[12] Planck collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [1905.05697].

[13] The Simons Observatory Collaboration, P. Ade, J. Aguirre, Z. Ahmed, S. Aiola, A. Ali et al., *The Simons Observatory: Science goals and forecasts*, *ArXiv e-prints* (2018) [1808.07445].

[14] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold et al., *Cmb-s4 science book, first edition*, 2016.

[15] 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].

[16] 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].

[17] 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].

[18] R. Scoccimarro, L. Hui, M. Manera and K. C. Chan, *Large-scale bias and efficient generation of initial conditions for nonlocal primordial non-gaussianity*, *Physical Review D* **85** (2012) .

[19] J. R. Fergusson and E. P. S. Shellard, *Primordial non-Gaussianity and the CMB bispectrum*, *Physical Review D* **76** (2007) 083523 [astro-ph/0612713].

[20] 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].

[21] 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].

[22] S. Weinberg, *Quantum contributions to cosmological correlations*, *Physical Review D* **72** (2005) 043514 [hep-th/0506236].

[23] 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].

[24] J. R. Fergusson, M. Liguori and E. P. S. Shellard, *General cmb and primordial bispectrum estimation: Mode expansion, map making, and measures offnl*, *Physical Review D* **82** (2010) .

[25] 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].

[26] 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].

[27] 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].

[28] 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].

[29] 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].

[30] S. Passaglia and W. Hu, *Scalar bispectrum beyond slow-roll in the unified EFT of inflation*, *Physical Review D* **98** (2018) 023526 [1804.07741].

[31] P. Adshead, W. Hu and V. Miranda, *Bispectrum in single-field inflation beyond slow-roll*, *Physical Review D* **88** (2013) 023507 [1303.7004].

[32] V. Miranda, W. Hu and P. Adshead, *Warp features in DBI inflation*, *Physical Review D* **86** (2012) 063529 [1207.2186].

[33] E. Silverstein and D. Tong, *Scalar speed limits and cosmology: Acceleration from D-cceleration*, *Physical Review D* **70** (2004) 103505 [hep-th/0310221].

[34] V. Miranda and W. Hu, *Inflationary steps in the Planck data*, *Physical Review D* **89** (2014) 083529 [1312.0946].

[35] X. Chen, *Folded resonant non-Gaussianity in general single field inflation*, *Journal of Cosmology and Astroparticle Physics* **12** (2010) 003 [1008.2485].

[36] 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].

[37] S. Garcia-Saenz, S. Renaux-Petel and J. Ronayne, *Primordial fluctuations and non-Gaussianities in sidetracked inflation*, *ArXiv e-prints* (2018) [1804.11279].

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

[39] J. Garriga and V. F. Mukhanov, *Perturbations in k-inflation*, *Physics Letters B* **458** (1999) 219 [hep-th/9904176].

[40] G. Cabass, E. Pajer and F. Schmidt, *How Gaussian can our Universe be?*, *JCAP* **1701** (2017) 003 [1612.00033].

[41] 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].

[42] 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].

[43] J. S. Horner and C. R. Contaldi, *Non-Gaussianity Unleashed*, *ArXiv e-prints* (2013) [1303.2119].

[44] 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].

[45] 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].

[46] 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].

[47] D. J. Mulryne and J. W. Ronayne, *PyTransport: A Python package for the calculation of inflationary correlation functions*, *ArXiv e-prints* (2016) [1609.00381].

[48] 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].

[49] S. Butchers and D. Seery, *Numerical evaluation of inflationary 3-point functions on curved field space*, *ArXiv e-prints* (2018) [1803.10563].

[50] J. Fumagalli, S. Garcia-Saenz, L. Pinol, S. Renaux-Petel and J. Ronayne, *Hyper-non-gaussianities in inflation with strongly nongeodesic motion*, *Physical Review Letters* **123** (2019) .

[51] 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].

[52] 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].

[53] W. Sohn and J. R. Fergusson, *Cmb-s4 forecast on the primordial non-gaussianity parameter of feature models*, *Physical Review D* **100** (2019) .

[54] P. D. Meerburg and et al, *Primordial Non-Gaussianity*, *arXiv e-prints* (2019) arXiv:1903.04409 [1903.04409].

[55] 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].

[56] K. M. Smith and M. Zaldarriaga, *Algorithms for bispectra: forecasting, optimal analysis and simulation*, *Monthly Notices of the Royal Astronomical Society* **417** (2011) 219.

[57] 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].

[58] 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].

[59] A. Dixit, L. Jiu, V. H. Moll and C. Vignat, *The finite Fourier transform of classical polynomials*, *ArXiv e-prints* (2014) [1402.5544].

[60] A. Iserles and S. P. Norsett, *Efficient quadrature of highly efficient integrals using deriatives*, *Royal Society Publishing, Volume 461, Issue 2057* (2005) .

[61] 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].

[62] G. Rigopoulos, *Gauge invariance and non-Gaussianity in inflation*, *Physical Review D* **84** (2011) 021301 [1104.0292].

[63] P. Clarke and E. P. S. Shellard, *Probing inflation with precision bispectra*, 2021.

[64] A. Achúcarro and G. A. Palma, *The string swampland constraints require multi-field inflation*, *ArXiv e-prints* (2018) [1807.04390].

[65] A. J. Tolley and M. Wyman, *Equilateral non-gaussianity from multifield dynamics*, *Physical Review D* **81** (2010) .

[66] 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].

[67] 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].

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

[69] 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].

Appendix A

Appendices

A.1 Efficient numerics

Efficient evaluation and numerical usage of the coefficients (vander, etc). Numeric choices that aren't specific to mode evolution, etc. Gauss-Legendre integration, fixed weights.

A.2 Tools used

Numpy, Scipy, Jupyter Notebooks, QUADPTS...?

A.3 Transport method for the Modal coeffs

Infinite 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 (2.17) are coupled in an infinite hierarchy.