

# Probing inflation with precision bispectra

**Philip Clarke and E.P.S. Shellard**

Centre for Theoretical Cosmology, DAMTP, University of Cambridge,  
Cambridge CB3 0WA, United Kingdom

E-mail: [pc559@cam.ac.uk](mailto:pc559@cam.ac.uk), [E.P.S.Shellard@damtp.cam.ac.uk](mailto:E.P.S.Shellard@damtp.cam.ac.uk)

Received February 18, 2021

Accepted June 8, 2021

Published August 2, 2021

**Abstract.** Calculating the 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. Exploiting the inherent separability of the tree level in-in formalism using expansions in separable basis functions provides a means by which to obviate some of these difficulties. Here, we develop this approach further into a practical and efficient numerical methodology which can be applied to a much wider and more complicated range of bispectrum phenomenology, making an important step forward towards observational pipelines which can directly confront specific models of inflation. We describe a simple augmented Legendre polynomial basis and its advantages, then test the method on single-field inflation models with non-trivial phenomenology, showing that our calculation of these coefficients is fast and accurate to high orders.

**Keywords:** inflation, non-gaussianity

**ArXiv ePrint:** [2012.08546](https://arxiv.org/abs/2012.08546)



---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Inflationary bispectra and observations</b>	<b>3</b>
2.1	Bispectrum estimation	3
2.2	The tree level in-in calculation	4
2.3	$\mathcal{P}(\mathbf{X}, \phi)$ theories and approximate bispectrum templates	5
2.4	Configuration-by-configuration codes	9
2.5	In-in separability	9
2.6	Precision tests	10
<b>3</b>	<b>Methodology</b>	<b>11</b>
3.1	Choice of basis	11
3.2	Exploiting the separability of the in-in formalism	18
3.3	The interaction Hamiltonian	21
3.4	Numerical considerations	23
<b>4</b>	<b>Validation</b>	<b>26</b>
4.1	Validation methods	26
4.2	Quadratic slow-roll	28
4.3	DBI inflation	28
4.4	Step features	31
4.5	Resonance features	31
<b>5</b>	<b>Discussion</b>	<b>33</b>

---

## 1 Introduction

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 [1], 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 [2, 3], interactions between multiple fields [4], sharp features [5] and periodic features [6]. 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 [7, 8].

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 [9, 10]. With this high-precision *Planck* data, and data from forthcoming experiments such as the Simons Observatory (SO) [11] and CMB-S4 [12], 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 [13–16] 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 [17–19] 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 [20], and the details of the bispectrum estimation part will be detailed in [21]. We go beyond the work of [20] 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.

The paper is organised as follows. In section 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 [20], and we discuss methods of testing numerical bispectrum results, defining our relative difference measurement. In section 3 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 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 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 section 5.

## 2 Inflationary bispectra and observations

### 2.1 Bispectrum estimation

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 to which we have access to estimate the amplitude of particular bispectrum templates, and use these estimates to constrain inflationary physics; see [22, 23] 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 [7, 8]. 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 [17–19] 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  $f_{NL}^{\text{local}}$ . 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 [17–19] 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.

The primordial bispectrum is usually written as:

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_1, k_2, k_3) \quad (2.1)$$

where  $\zeta_{\mathbf{k}}$  is a Fourier mode of the standard gauge invariant curvature perturbation. The delta function comes from demanding statistical homogeneity; demanding statistical isotropy restricts the remaining dependence to the magnitudes of the vectors. We denote the magnitude of  $\mathbf{k}_i$  as  $k_i$ . This leaves us with a function of three parameters,  $k_1, k_2, k_3$ . It is useful to define the dimensionless shape function:

$$S(k_1, k_2, k_3) = (k_1 k_2 k_3)^2 B(k_1, k_2, k_3). \quad (2.2)$$

The bispectrum is only defined where the triangle condition

$$\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0, \quad (2.3)$$

is satisfied, which implies that the triangle inequality must hold

$$k_1 + k_2 \geq k_3 \text{ and cyclic perms.} \quad (2.4)$$

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 6), the intersection of that cube with the tetrahedron that satisfies (2.4). This has important implications that we will explore in section 3.1.

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

$$B^F(k_1, k_2, k_3) = f_{NL}^F \times F(k_1, k_2, k_3) \quad (2.5)$$

where  $F$  contains the dependence on the  $k$ -configuration. This definition coincides with the definitions of  $f_{NL}^{\text{local}}$ ,  $f_{NL}^{\text{equil}}$  and  $f_{NL}^{\text{ortho}}$  when  $F$  is (respectively) the local (see (2.26)), equilateral (see (2.29)) and orthogonal templates, as defined in [9].

If the shape function (2.2) has the form:

$$S(k_1, k_2, k_3) = X(k_1)Y(k_2)Z(k_3), \quad (2.6)$$

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:

$$b_{l_1 l_2 l_3}^{X_1 X_2 X_3} = \left(\frac{2}{\pi}\right)^3 \int_0^\infty dr r^2 \int_{\mathcal{V}_k} d^3k (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3) \prod_{i=1}^3 [j_{l_i}(k_i r) \Delta_{l_i}^{X_i}(k_i)], \quad (2.7)$$

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  $\Delta_{l_i}$  already appear in a separable way. This property can also be used to efficiently generate non-Gaussian initial conditions for simulations [16].

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

## 2.2 The tree level in-in calculation

The standard starting point for calculating higher-order correlators for models of inflation is the in-in formalism [1, 24]. 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:

$$\langle \zeta_{\mathbf{k}_1}(\tau)\zeta_{\mathbf{k}_2}(\tau)\zeta_{\mathbf{k}_3}(\tau) \rangle = -i \int_{-\infty(1-i\epsilon)}^{\tau} d\tau' a(\tau') \langle 0 | \zeta_{\mathbf{k}_1}(\tau)\zeta_{\mathbf{k}_2}(\tau)\zeta_{\mathbf{k}_3}(\tau) H_{\text{int}}(\tau') | 0 \rangle + c.c \quad (2.8)$$

where all the operators on the right-hand side are in the interaction picture and  $H_{\text{int}}$  is the interaction Hamiltonian, containing terms cubic in  $\zeta$ . From this calculation we obtain the dimensionless shape function  $S(k_1, k_2, k_3)$ , defined in (2.2), which is then used as input into (2.7). As an example, if one takes  $H_{\text{int}} \propto \dot{\zeta}^3$ , 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 [20], 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  $H_{\text{int}}$ . The first part is used to evolve the interaction picture fields,  $\zeta_I$ , which we will simply refer to as  $\zeta$ , as in (2.8). The perturbations see an interaction Hamiltonian  $H_{\text{int}}$ , 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 [25]. This can be used, as pointed out in [20], 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 [2], 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 [1]. 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.

### 2.3 $P(X, \phi)$ theories and approximate bispectrum templates

There is an extensive literature on the calculation of bispectra from models of inflation [2, 5, 6, 26–31]. 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 [3, 32–37];

effectively single-field models with imaginary sound speeds can generate a bispectrum mostly orthogonal to the usual equilateral and local templates [38]. The methods outlined in this paper 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

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

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:

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

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

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

The background quantities are evolved according to the Friedmann equations, which are set with consistent initial conditions. The equation of motion for the perturbations is:

$$\zeta_k'' + (3 - \epsilon + \eta - 2\epsilon_s)\zeta_k' + \frac{c_s^2 k^2}{a^2 H^2} \zeta_k = 0 \quad (2.13)$$

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:

$$\zeta_k = \frac{i}{a} \sqrt{\frac{c_s}{4\epsilon k}} e^{-ik\tau_s} \quad (2.14)$$

where we define  $\tau_s$  through  $\tau_s' = \frac{c_s}{aH}$  in analogy with the usual  $\tau$  with  $\tau' = \frac{1}{aH}$ . The solution in slow-roll (without features) is then approximately

$$\zeta_k \propto (1 + ik\tau_s) e^{-ik\tau_s}. \quad (2.15)$$

At leading order in slow-roll the power spectrum is [28, 39]:

$$P^\zeta(k) = \frac{1}{8\pi^2} \frac{H^2}{c_s \epsilon}, \quad (2.16)$$

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

$$n_s - 1 = -2\epsilon - \eta - \epsilon_s. \quad (2.17)$$

Similarly to [20], at early times we extract the factor of  $e^{-ik\tau_s}$  from the mode functions and numerically evolve  $\zeta_k e^{ik\tau_s}$ .<sup>1</sup> Unless interrupted, this prefactor decays exponentially.

<sup>1</sup>In fact [20] extracts a factor of  $e^{-ikc_s(\tau)\tau}$ , losing efficiency due to slow-roll corrections.

Eventually we switch to evolving  $\zeta_k$  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  $\zeta_k$  before the onset of the feature gives robust results without needing to loosen the tolerance.

Initially we consider the same basic models as in [20]; a quadratic potential

$$V_{\phi^2}(\phi) = \frac{1}{2}m^2\phi^2, \quad (2.18)$$

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

$$S_{\text{DBI}} = \int d^4x \sqrt{-g} \left( -\frac{1}{f(\phi)} \left( (1 + f(\phi) \partial_\mu \phi \partial^\mu \phi)^{\frac{1}{2}} - 1 \right) - V(\phi) \right), \quad (2.19)$$

with

$$f(\phi) = \frac{\lambda_{\text{DBI}}}{\phi^4}, \quad V(\phi) = V_0 - \frac{1}{2}m^2\phi^2, \quad m = \sqrt{\beta_{\text{IR}}}H. \quad (2.20)$$

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

$$V(\phi) = V_{\phi^2}(\phi) \left( 1 - c \tanh \left( \frac{\phi f - \phi}{d} \right) \right). \quad (2.21)$$

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

$$V(\phi) = V_{\phi^2}(\phi) \left( 1 + bf \sin \left( \frac{\phi}{f} \right) \right). \quad (2.22)$$

For more details on these models, see [27]. To express the bispectrum results more compactly we use the symmetric polynomial notation employed in [18]:

$$\begin{aligned} K_p &= \sum_{i=1,2,3} k_i^p, \\ K_{pq} &= \frac{1}{\Delta_{pq}} \sum_{i \neq j} k_i^p k_j^q, \\ K_{prs} &= \frac{1}{\Delta_{prs}} \sum_{i \neq j \neq l} k_i^p k_j^r k_l^s, \end{aligned} \quad (2.23)$$

where  $\Delta_{pq}$  is 2 if  $p = q$ , 1 otherwise and  $\Delta_{prs}$  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:

$$S^{\text{Malda}}(k_1, k_2, k_3) = A^{\text{Malda}} \left( (3\varepsilon - 2\eta) \frac{K_3}{K_{111}} + \varepsilon \left( K_{12} + 8 \frac{K_{22}}{K} \right) \right), \quad (2.24)$$

$$A^{\text{Malda}} = -\frac{1}{32} \frac{H^4}{12\varepsilon^2}, \quad (2.25)$$



with  $\eta = 2\varepsilon$  for (2.18). At the primordial level, this is well approximated by the separable local template

$$S^{\text{local}}(k_1, k_2, k_3) = \frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_3 k_1} + \frac{k_3^2}{k_1 k_2} = \frac{K_3}{6 K_{111}}. \quad (2.26)$$

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 [9]. For the featureless DBI scenario, the shape function is [3]:

$$S^{\text{DBI}}(k_1, k_2, k_3) = A^{\text{DBI}} \frac{K_5 + 2K_{14} - 3K_{23} + 2K_{113} - 8K_{122}}{K_{111} K^2}, \quad (2.27)$$

$$A^{\text{DBI}} = -\frac{1}{32} \frac{H^4}{12\varepsilon^2} \left( \frac{1}{c_s^2} - 1 \right), \quad (2.28)$$

to leading order in slow-roll. Any constraint on the magnitude  $A^{\text{DBI}}$  can be translated into one on the effective sound speed which from *Planck* has a lower limit  $c_s^{\text{DBI}} \geq 0.087$  at 95% significance [9]. The shape (2.27) can be approximated by the separable equilateral template

$$S^{\text{equil}}(k_1, k_2, k_3) = \frac{(k_2 + k_3 - k_1)(k_3 + k_1 - k_2)(k_1 + k_2 - k_3)}{k_1 k_2 k_3}. \quad (2.29)$$

These templates can be modified to be more physically realistic by including scaling consistent with the spectral index  $n_s$  [9]. For example, we can add some scale dependence to the DBI model in a reasonable first approximation by including a prefactor

$$S^{\text{DBI}-n_s}(k_1, k_2, k_3) = \left( \frac{k_1 k_2 k_3}{k_*^3} \right)^{n_s-1} S^{\text{DBI}}(k_1, k_2, k_3). \quad (2.30)$$

We now turn to feature templates. The result of adding a feature of the form (2.21) is to add oscillatory features of the form

$$S^{\text{cos}}(k_1, k_2, k_3) = \cos(w(k_1 + k_2 + k_3)) \quad (2.31)$$

though more realistically there is some phase, shape dependence and a modulating envelope, as detailed in [5]. The result of adding a resonant feature of the form (2.22) is to generate logarithmic oscillatory features in the shape function of the form

$$S^{\text{ln-cos}}(k_1, k_2, k_3) = \cos(w \ln(k_1 + k_2 + k_3)). \quad (2.32)$$

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 [36].

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.

## 2.4 Configuration-by-configuration codes

Previous work on the numerical calculations of inflationary non-Gaussianity include the BINGO code [42], Chen et al. [26, 27], the work of Horner et al. [43–45] and the Transport Method [46–49]. 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 (3.14). 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.<sup>2</sup> 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 [38].

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.

## 2.5 In-in separability

In [20] 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, [20] 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 [17–19] 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 [20]) are constraints on the parameters of certain inflation models through approximate phenomenological templates. These constraints can be found in [9, 10]. The idea of [20] is an extension of that philosophy to the primordial level, and our work is in implementing that idea. In [17–19] 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 [20]. We restructure the methods, improving on the work of [20] 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 [20] in its rate of convergence, its transparency, and its flexibility. We do this without sacrificing orthogonality. This is detailed

---

<sup>2</sup>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 (2.13) are coupled in an infinite hierarchy.

in section 3.1. Our improvements over the methods sketched in [20] allow us to validate on non-trivial bispectra for the first time, including sharp deviations from slow-roll, which we present in section 4. We quote our results in terms of a measure that is easier to interpret than the correlation defined in [20], and that includes the magnitude as well as the shape information on the full tetrapyd. This is discussed in section 2.6.

## 2.6 Precision tests

The inner product of two bispectrum shape functions is given by

$$S_1 \cdot S_2 = \langle S_1, S_2 \rangle = \int_{T_k} d^3k S_1(k_1, k_2, k_3) S_2(k_1, k_2, k_3), \quad (2.33)$$

where  $T_k$  refers to the tetrapyd, the region of the cube  $[k_{\min}, k_{\max}]^3$  that obeys the triangle inequality. Following [50] we define the two correlators:

$$\mathcal{S}(S_1, S_2) = \frac{S_1 \cdot S_2}{\sqrt{(S_1 \cdot S_1)(S_2 \cdot S_2)}}, \quad \mathcal{A}(S_1, S_2) = \sqrt{\frac{S_1 \cdot S_1}{S_2 \cdot S_2}}. \quad (2.34)$$

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 [17–19], but in this work we restrict ourselves to accurately calculating the full primordial bispectra, weighting each configuration equally.

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

$$\begin{aligned} \mathcal{E}(S_1, S_2) &= \sqrt{\frac{|S_1 - S_2|^2}{|S_2|^2}} = \sqrt{\frac{|S_1|^2 - 2S_1 \cdot S_2 + |S_2|^2}{|S_2|^2}} \\ &= \sqrt{\mathcal{A}(S_1, S_2)^2 - 2\mathcal{A}(S_1, S_2)\mathcal{S}(S_1, S_2) + 1}. \end{aligned} \quad (2.35)$$

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

$$\mathcal{E}(S_1, S_2) = \sqrt{2(1 - \mathcal{S}(S_1, S_2))}. \quad (2.36)$$

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 [50] is a far better representation of actual convergence between two shape functions than the correlation used in [20]. 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. In that section we also plot residuals on slices through the tetrapyd, relative to the representative value

$$\sqrt{\frac{S \cdot S}{\int_{T_k} d^3k}}. \quad (2.37)$$

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  $\mathbf{k}_S \equiv (\mathbf{k}_2 - \mathbf{k}_3)/2$ :

$$S(k_1, k_2, k_3) = - \left[ (n_s - 1)|_{k_S} + \mathcal{O} \left( \frac{k_1^2}{k_S^2} \right) \right] P_\zeta(k_1) P_\zeta(k_S), \quad k_1 \ll k_S \quad (2.38)$$

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].

### 3 Methodology

Given its separable form, the tree-level in-in formalism is amenable to more efficient calculation using separable modes, as first mentioned in [20]. That work extended the separable methodology previously implemented for the CMB bispectrum [19]. 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 (2.2) using a set of basis functions as

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

where the basis functions  $Q_n(k_1, k_2, k_3)$  are explicitly separable functions of their arguments. 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  $\alpha_n$ ). The details of this once-per-basis calculation will be presented in [21]. 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 section 3.1. In section 3.2 we set the notation we will use to recast the standard numerical in-in calculation into a calculation of  $\alpha_n$ , and sketch the steps involved. In section 3.3 we outline the details of the interaction Hamiltonian, including accounting for the spatial derivatives in our final result. In section 3.4 we make precise the numerical considerations of the calculation, especially our methods of dealing with the high-frequency oscillations at early times.

#### 3.1 Choice of basis

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

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

Here,  $n$  labels the ordered integer triplet  $n \leftrightarrow \{prs\}$  in an appropriate manner (see some ordering alternatives in [19]), while the symmetrised average of all  $\{prs\}$  permutations is

$$q_{(pq_rq_s)} \equiv (1/3!) \sum_{\text{perms}} q_p q_r q_s \quad \text{and} \quad \Xi_{prs} = \begin{cases} 1, & p = r = s \text{ all equal,} \\ \sqrt{3} & \{prs\} \text{ any two equal,} \\ \sqrt{6} & \{prs\} \text{ all different.} \end{cases} \quad (3.3)$$

Unless stated otherwise, the  $\{prs\}$  triples for each permutation set  $n$  in (3.3) are represented by the coefficient with  $0 \leq p \leq r \leq s$ , 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 3.2, 3.3 and 3.4, we are free to choose our set of basis functions to optimise for efficient convergence, ensuring our results are useful for comparison with observations. There are a wide variety of options available, such as polynomial bases or Fourier series, that can be chosen for the  $q_p(k)$ . While not strictly necessary for the method, it is more convenient if the resulting 3D basis functions  $Q_n(k_1, k_2, k_3)$  are orthogonal on the cubic region of selected wavenumbers, making it much more straightforward to obtain controlled convergence. Overall, then, rapid convergence is the key criterion in choosing the basis functions  $q_p(k)$  in (3.2), thus determining the nature of the numerical errors in the calculated bispectrum. However, since we are going beyond the featureless examples of [20] 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 (3.18)). 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 (2.24) 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 (2.27), this effect can be disastrous if not handled properly. This can be easily seen from (2.29), 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.<sup>3</sup>

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

$$q_r(k) = \left( \frac{2r+1}{k_{\max} - k_{\min}} \right)^{1/2} P_r(\bar{k}), \quad (3.4)$$

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

$$\bar{k} = \frac{2k - k_{\max} - k_{\min}}{k_{\max} - k_{\min}}. \quad (3.5)$$

We shall label as  $\mathcal{P}_0$  the basis function set of pure Legendre polynomials in (3.4), with  $r = 0, 1, \dots, p_{\max}-1$ . These were considered also in [20], however, while they prove to be particularly functional building blocks for other modal applications, in the context of the in-in formalism they converge so slowly even for simple shapes as to be inadequate when taken on their own. This poor rate of convergence with  $\mathcal{P}_0$  for two local- and equilateral-type shapes is shown in figure 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,

$$q_{p_{\max}}(k) = \text{Orth}[1/k], \quad \text{with } \mathcal{P}_1 = \{q_r(k) \mid r = 0, 1, 2, \dots, p_{\max}\}, \quad (3.6)$$

where Orth represents the projection orthogonal to the original Legendre polynomial basis  $\mathcal{P}_0$ . As we see in figure 1, the convergence properties for the augmented basis  $\mathcal{P}_1$  are dramatically improved.

The two basis function sets actually used in [20] 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

$$q_r(k) = \left( \frac{2r+1}{\ln k_{\max} - \ln k_{\min}} \right)^{1/2} P_r(\overline{\ln k}) \quad \text{with} \quad \overline{\ln k} = \frac{2 \ln k - \ln k_{\max} - \ln k_{\min}}{\ln k_{\max} - \ln k_{\min}}. \quad (3.7)$$

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  $k$ :

$$q_r(k) = \left( \frac{2r+1}{k_{\max} - k_{\min}} \right)^{1/2} \frac{P_r(\bar{k})}{k}, \quad (3.8)$$

---

<sup>3</sup>There are results in the literature that describe generic  $K = k_1 + k_2 + k_3$  poles in correlators — see for example [53]. 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 shows that  $\mathcal{P}_1$  and  $\mathcal{P}_{01}^{ns}$  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.

where the rescaled  $\bar{k}$  is defined in (3.5). This can also be thought of as expanding the bispectrum  $k_1 k_2 k_3 S(k_1, k_2, k_3)$  in  $\mathcal{P}_0$ , instead of the shape function  $S(k_1, k_2, k_3)$  itself. The consequence is that neither (3.7) nor (3.8) are orthogonal with respect to the flat weighting of the inner product (2.33). However, as shown in [20], these two basis sets (3.7) and (3.8) are able to approximate the three canonical bispectrum shapes. Nevertheless, our aim is to go beyond the featureless examples investigated in [20], 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 (3.7) and (3.8). 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 (2.7), 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 (3.7) and (3.8) 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  $r = 0, \dots, p_{\max}-1$ , then we define

$$\tilde{f}(k) = \text{Orth}[f(k)] \equiv f(k) - \sum_{r=0}^{p_{\max}-1} \frac{\langle f, q_r \rangle}{\langle q_r, q_r \rangle} q_r(k) \quad (3.9)$$

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

$$q_0(k) = 1, \quad q_{2r-1}(k) = \sin(\pi r \bar{k}), \quad q_{2r}(k) = \cos(\pi r \bar{k}), \quad 1 \leq r \leq (p_{\max}-3)/2 \quad (3.10)$$

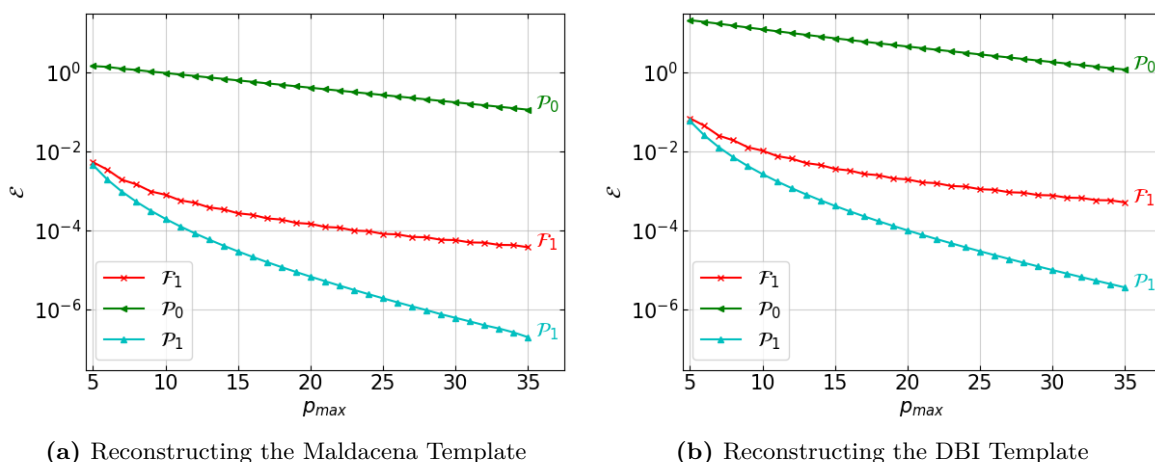
$$q_{p_{\max}-2}(k) = \bar{k}, \quad q_{p_{\max}-1}(k) = \bar{k}^2. \quad (3.11)$$

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 (3.6), we will similarly create an augmented Fourier basis  $\mathcal{F}_1$  by adding the inverse  $1/k$  term to the  $\mathcal{F}_0$  basis, i.e. using  $q_{p_{\max}}(k) = \text{Orth}[1/k]$  with (3.9). 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 (2.24) and the DBI shape (2.27). To mimic the in-in calculation, we expand the shape on the cube, but test the result on the tetrapyd using (2.35). The results are shown in figure 1 where we find that the

Notation	Building Blocks	Augmented by	Definition
$\mathcal{P}_0$	Legendre polynomials		(3.4)
$\mathcal{F}_0$	Fourier Series	$k, k^2$	(3.10)
$\mathcal{P}_1$	Legendre polynomials	$k^{-1}$	
$\mathcal{F}_1$	Fourier Series	$k, k^2, k^{-1}$	
$\mathcal{P}_1^{n_s}$	Legendre polynomials	$k^{-1+(n_s^*-1)}$	(3.12)
$\mathcal{P}_{01}^{n_s}$	Legendre polynomials	$k^{n_s^*-1}, k^{-1+(n_s^*-1)}$	(3.13)

**Table 1.** Basis summary — the augmentation of the basis is done using (3.9). The size of each basis is referred to as  $p_{\max}$ .

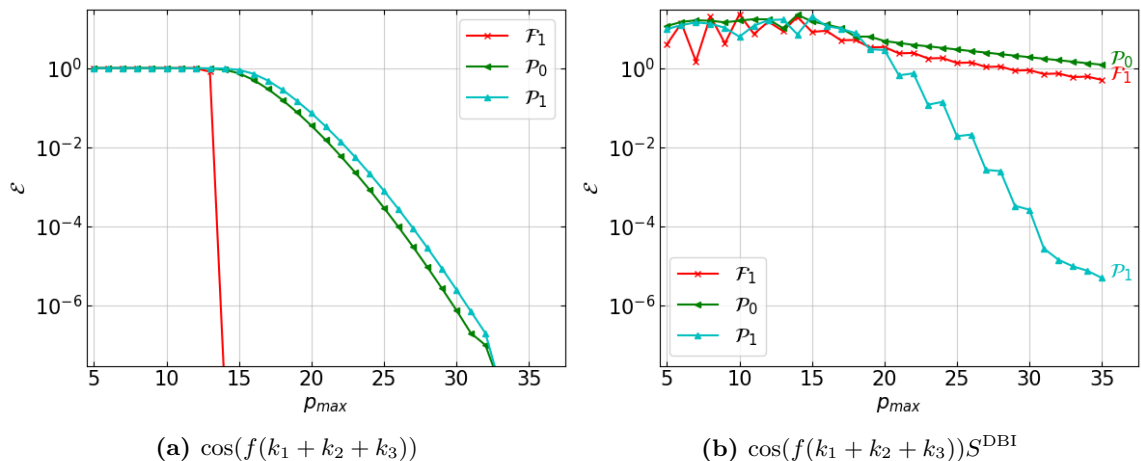


**Figure 1.** Convergence comparisons for the Legendre and Fourier basis functions for (a) the Maldacena template (2.24) and (b) the DBI template (2.27). The pure Legendre  $\mathcal{P}_0$  basis requires many terms to fit the  $1/k$  behaviour in both Maldacena’s template (2.24) and the DBI template (2.27). 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 (2.27) are larger than (2.24) because of the larger contributions outside the tetrapyd dominating the fit.

Legendre polynomials basis set  $\mathcal{P}_0$  converges so slowly as to be unusable (with the Fourier modes  $\mathcal{F}_0$  worse and not plotted). However, the augmented Legendre basis  $\mathcal{P}_1$  (including  $1/k$ ) leads to rapid convergence with an improvement of four orders of magnitude at  $p_{\max} = 15$ . The augmented Fourier basis  $\mathcal{F}_1$  also converges quickly relative to  $\mathcal{P}_0$ , but is outdone by  $\mathcal{P}_1$ . Though we do not show the convergence on the cube, we find that for Maldacena’s template this is of the same order of magnitude as the error on the tetrapyd. For the DBI shape, however, the fit on the tetrapyd lags significantly behind, due to the effect of the large non-physical configurations. This explains the order of magnitude difference between the convergence at each  $p_{\max}$  for the two shapes in figure 1.

Next, we investigate oscillatory model templates. The simple feature model (2.31) and the resonance model (2.32) 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 (2.27) to obtain a





**Figure 2.** Convergence comparison for oscillatory models. (a) As expected, the  $\mathcal{F}_1$  basis fits an oscillation with no shape dependence (2.31) (that is periodic in the  $k$ -range) perfectly. For this special case, the  $\mathcal{P}_0$  and  $\mathcal{P}_1$  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 (2.27) and (2.31)), we see that again  $\mathcal{P}_1$  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 (2.35) to the physical configurations on the tetrapyd; when considered over the entire cube the fit improves monotonically.

feature template with both shape and scale dependence. As shown in figure 2,  $\mathcal{F}_0$  naturally outperforms  $\mathcal{P}_0$  for a pure oscillation, but when the equilateral-type DBI template (2.27) is superimposed, even the augmented Fourier modes  $\mathcal{F}_1$  converge poorly and the Legendre modes  $\mathcal{P}_1$  clearly offer a better more robust option. In figure 3 we see that for a logarithmic oscillation,  $\mathcal{P}_1$  always converges in the fewest modes.

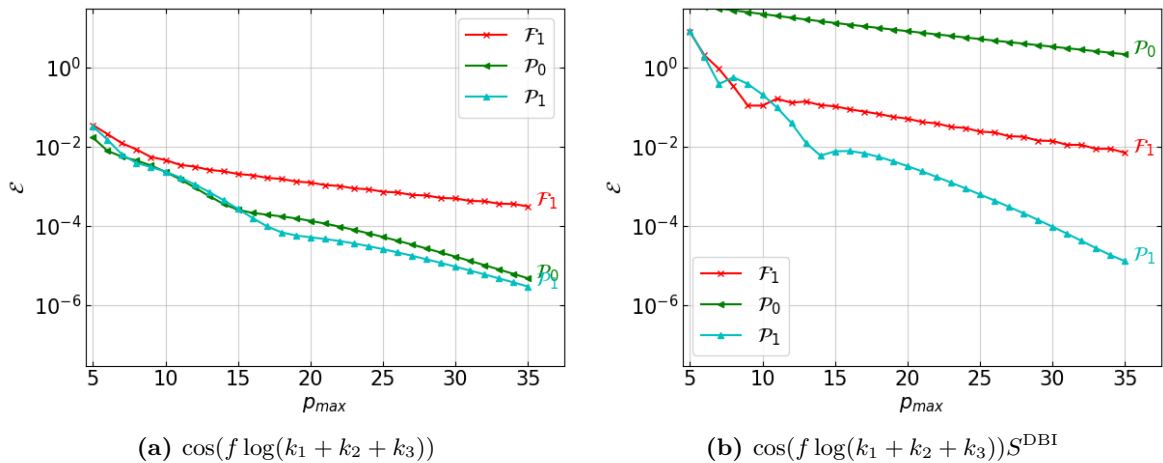
Finally, we consider convergence in the light of the more subtle scale-dependence due to the spectral index  $n_s$  of the power spectrum. The simple canonical examples in figure 1 had shape dependence and no scale dependence, but this would only be expected of scenarios unrealistically deep in the slow-roll limit. When we include this scale dependence, using (2.30) 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 (3.9),

$$q_{p_{\max}}(k) = \text{Orth} \left[ k^{-1+(n_s^*-1)} \right], \quad (3.12)$$

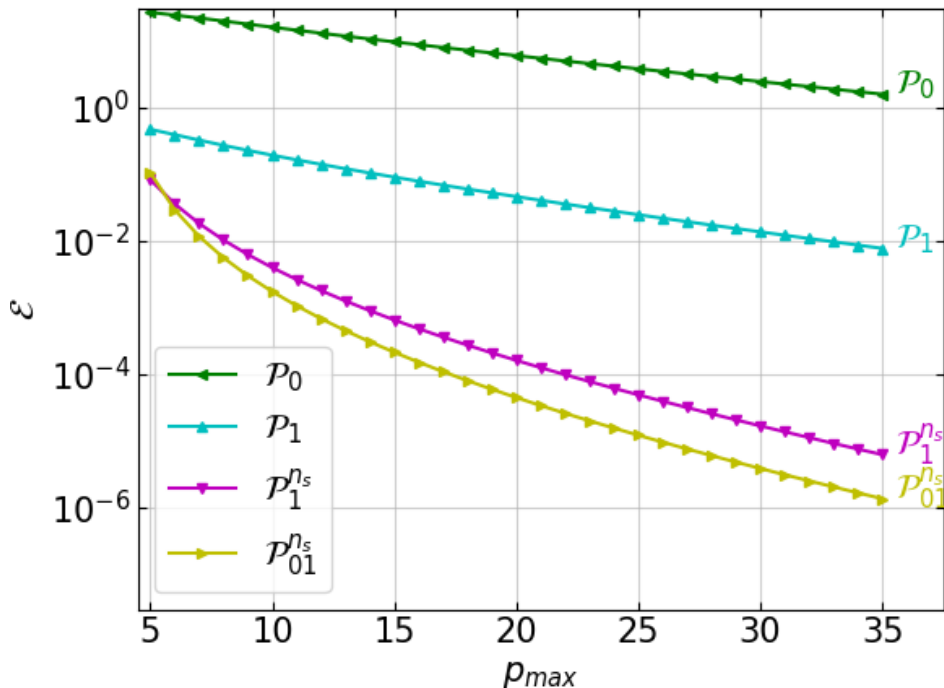
which we refer to as  $\mathcal{P}_1^{n_s}$ . Secondly, we also augment  $\mathcal{P}_0$  with an additional scale-dependent constant term as

$$q_{p_{\max}}(k) = \text{Orth} \left[ k^{(n_s^*-1)} \right], \quad q_{p_{\max}+1}(k) = \text{Orth} \left[ k^{-1+(n_s^*-1)} \right], \quad (3.13)$$

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



**Figure 3.** (a) The convergence for a log oscillation model (2.32) with no shape dependence. For this type of feature, the  $\mathcal{P}_0$  and  $\mathcal{P}_1$  sets of basis functions require fewer modes to accurately describe the shape than the  $\mathcal{F}_1$  basis. (b) For the more complex and realistic case of a feature with scale and shape dependence (in this case the product of (2.27) and (2.32)), we see that again  $\mathcal{P}_1$  converges with the fewest modes, though in this case  $\mathcal{F}_1$  outperforms  $\mathcal{P}_0$ .



**Figure 4.** For the scale-dependent DBI template (2.30), by including a minimal amount of power spectrum information using (3.12) and (3.13) (here with  $n_s^* = 0.97$ ), we can push the errors to less than 0.1% at  $\rho_{\text{max}} = 15$ , allowing us to work with a basis that can efficiently capture the expected scale dependence of the shape function.

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. See table 1 for a summary of the basis sets described in this paper.

### 3.2 Exploiting the separability of the in-in formalism

In this section we set up the notation, and sketch the steps required to calculate the coefficients  $\alpha_n$  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  $\tau_s$ , 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  $H_{\text{int}}$ . Consider indexing with  $i = 1, 2, 3 \dots$  the interaction vertices in  $H_{\text{int}}$ , so then the bispectrum (2.8) can be expressed as a sum over separable contributions of the form:

$$\begin{aligned} S(k_1, k_2, k_3) &= \sum_i I^{(i)}(k_1, k_2, k_3) \\ &= \sum_i \left[ v^{(i)}(k_1, k_2, k_3) \int_{-\infty(1-i\varepsilon)}^0 d\tau w^{(i)}(\tau) F^{(i)}(\tau, k_1) G^{(i)}(\tau, k_2) J^{(i)}(\tau, k_3) \right. \\ &\quad \left. + \text{cyclic perms} \right] \end{aligned} \quad (3.14)$$

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  $k^2 \zeta_{\mathbf{k}}(0) \zeta_{\mathbf{k}}^{*\prime}(\tau)$ . Spatial derivative terms, such as  $\partial_i \zeta \partial_i \zeta \rightarrow (\mathbf{k}_2 \cdot \mathbf{k}_3) \zeta_{\mathbf{k}_2}^*(\tau) \zeta_{\mathbf{k}_3}^*(\tau)$  also separate because of the triangle condition (2.3) 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 3.3. 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  $k^2 \zeta_{\mathbf{k}}(0) \zeta_{\mathbf{k}}^{*\prime}(\tau)$ ), where  $k^2$  comes from using the weighting of the scale-invariant shape function (2.2). 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 section 3.1) or the behaviour can be absorbed into the analytic prefactor,  $v^{(i)}(k_1, k_2, k_3)$ .<sup>4</sup> We use the former, as the numerics of the latter are less transparent and less physically motivated.

<sup>4</sup>At early times the modes are highly oscillatory in both  $k$  and  $\tau_s$ , which certainly requires special attention. We discuss this in section 3.4.

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 3.3 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 (2.15), an explicit example for the first interaction term in (3.25), i.e.  $H_{\text{int}}^{(1)} = \zeta'^2 \zeta$ , takes the form

$$F^{(1)}(\tau, k) = G^{(1)}(\tau, k) = c_s k^2 \tau \frac{H^2}{4\epsilon c_s} e^{i c_s k \tau}, \quad J^{(1)}(\tau, k) = (1 - i k c_s \tau) \frac{H^2}{4\epsilon c_s} e^{i c_s k \tau}. \quad (3.15)$$

In the simple mode approximation (2.15), such terms in (3.14) are straightforward to integrate analytically (using the  $i\epsilon$  prescription), provided the time-dependence of the slow-roll parameters and the sound speed is neglected [1]. However, for high precision bispectrum predictions we must incorporate the full time-dependence, while solving (2.13) 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 (3.14) 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 (3.14) as a mode expansion for each interaction term  $I^{(i)}(k_1, k_2, k_3)$  as

$$S(k_1, k_2, k_3) = \sum_i I^{(i)}(k_1, k_2, k_3) = \sum_i \sum_n \alpha_n^{(i)} Q_n(k_1, k_2, k_3), \quad (3.16)$$

where  $Q_n(k_1, k_2, k_3)$  is separable, built out of some orthonormal set  $q_p(k)$  as in (3.2). Armed with this set of modes, we can expand any of the interaction terms  $F^{(i)}(\tau, k)$ ,  $G^{(i)}(\tau, k)$ ,  $J^{(i)}(\tau, k)$  in (3.14) as:

$$F^{(i)}(\tau, k) = \sum_p f_p^{(i)}(\tau) q_p(k), \quad (3.17)$$

$$\text{where } f_p^{(i)}(\tau) = \int_{k_{\min}}^{k_{\max}} dk F^{(i)}(\tau, k) q_p(k). \quad (3.18)$$

Note that in the simple mode approximation, as in (3.15), we must expand  $e^{i c_s k \tau}$  in the terms of the  $q_p(k)$ . At early times  $\tau$  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 3.4 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 (3.17), 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 (3.14), 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,

$$\begin{aligned}
 I^{(i)}(k_1, k_2, k_3) &= v^{(i)}(k_1, k_2, k_3) \sum_P \tilde{\alpha}_P^{(i)} q_P(k_1, k_2, k_3) + \text{cyclic perms} \\
 &= \sum_P \tilde{\alpha}_P^{(i)} \sum_{\tilde{P}} V_{P\tilde{P}}^{(i)} q_{\tilde{P}}(k_1, k_2, k_3) + \text{cyclic perms} \\
 &= \sum_{\tilde{P}} \alpha_{\tilde{P}}^{(i)} q_{\tilde{P}}(k_1, k_2, k_3) + \text{cyclic perms}, \tag{3.19}
 \end{aligned}$$

where we have written

$$\tilde{\alpha}_P^{(i)} = \tilde{\alpha}_{prs}^{(i)} = \int d\tau w^{(i)}(\tau) f_p^{(i)}(\tau) g_r^{(i)}(\tau) h_s^{(i)}(\tau), \tag{3.20}$$

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

$$v^{(i)}(k_1, k_2, k_3) q_P(k_1, k_2, k_3) = \sum_{\tilde{P}} V_{P\tilde{P}}^{(i)} q_{\tilde{P}}(k_1, k_2, k_3), \tag{3.21}$$

and

$$\alpha_P^{(i)} = \sum_{\tilde{P}} \tilde{\alpha}_{\tilde{P}}^{(i)} V_{\tilde{P}P}^{(i)}. \tag{3.22}$$

We connect to the coefficients of the ordered, symmetrised basis in (3.16) by taking the symmetry factor into account,

$$\alpha_n^{(i)} = \frac{3!}{\Xi_P} \alpha_P^{(i)}. \tag{3.23}$$

The numerical calculation of  $V_{P\tilde{P}}^{(i)}$  (as defined by (3.21)) 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 (3.17)) 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 (3.16) 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 (3.20) we need to calculate.

Note the basis sets on the left and right hand side of (3.21) 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 (3.21) to be an expanded version of that on the left. For example, if the left hand basis was  $\mathcal{P}_0$  of size  $p_{\max}$ , the right hand basis would be  $\mathcal{P}_1$  of size  $p_{\max} + 3$ .

We can see from (3.20) that the number of time integrals needed is controlled by  $N_V \times p_{\max}^3$ ,<sup>5</sup> where  $N_V$  is the number of interaction vertices and  $p_{\max}$  is the size of the final basis.

<sup>5</sup>In fact the number is not quite  $p_{\max}^3$ . Since we have extracted the spatial derivatives, the only remaining possible source of asymmetric  $k$ -dependence comes from  $\zeta^3, \zeta^2 \zeta', \zeta'^2 \zeta$  or  $\zeta'^3$  so the time integral in (3.20) will always be (at least) symmetric in  $p$  and  $r$ .

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

We will calculate the contribution of each  $H_{\text{int}}$  vertex separately, indexing the vertices as above by  $(i)$ , so the overall shape function (3.14), (3.16) is then simply

$$S(k_1, k_2, k_3) = \sum_n \left( \sum_i \alpha_n^{(i)} \right) Q_n(k_1, k_2, k_3) = \sum_n \alpha_n Q_n(k_1, k_2, k_3). \quad (3.24)$$

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

The raison d'être for this approach is that all time integrals (3.20) are now independent of the  $k$ -configuration.<sup>6</sup> 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 (3.24) 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.  $p_{\text{max}}^3$  integrals (again ignoring symmetry) plus the decomposition. However for every model we have explored from the literature, our expansion in  $p_{\text{max}}$  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 (2.24) and (2.27), but is also true of bispectra with complicated features, as seen in section 3.1.

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  $Q_n \rightarrow \tilde{Q}_n$  can be achieved through a linear transformation  $\Gamma$  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 discuss the actual numerical implementation of these methods.

### 3.3 The interaction Hamiltonian

The methods detailed in the previous section depend on the separability of the third-order interaction Hamiltonian,  $H_{\text{int}}$ , 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  $H_{\text{int}}$ , we will take  $P(X, \phi)$  inflation as an example. The full cubic interaction

<sup>6</sup>They are not independent of  $k_{\text{min}}$  and  $k_{\text{max}}$  which define the domain of interest, which is analogous to the coefficients of a Taylor expansion depending on its expansion point.

Hamiltonian, not neglecting boundary terms, can be calculated as [2, 28, 29]

$$\begin{aligned}
 H_{\text{int}}(t) = \int d^3x \left\{ -\frac{a^3\varepsilon}{Hc_s^4} \left(1 - c_s^2 - 2c_s^2 \frac{\lambda}{\Sigma}\right) \dot{\zeta}^3 + \frac{a^3\varepsilon}{c_s^4} \left(3 - 3c_s^2 - \varepsilon + \eta\right) \zeta \dot{\zeta}^2 \right. \\
 - \frac{a\varepsilon}{c_s^2} \left(1 - c_s^2 + \varepsilon + \eta - 2\varepsilon_s\right) \zeta (\partial\zeta)^2 \\
 \left. - \frac{a^3\varepsilon^2}{2c_s^4} (\varepsilon - 4) \dot{\zeta} \partial\zeta \partial(\partial^{-2}\dot{\zeta}) - \frac{a^3\varepsilon^3}{4c_s^4} \partial^2\zeta (\partial(\partial^{-2}\dot{\zeta}))^2 \right\} \quad (3.25)
 \end{aligned}$$

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

This is commonly quoted with a term proportional to the equation of motion, but this will never contribute [2, 31, 54, 55]. 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 [2] (see also [31, 54, 55]) we do not work with any boundary terms. 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

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = \langle \tilde{\zeta}_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_2} \tilde{\zeta}_{\mathbf{k}_3} \rangle + \lambda \langle \tilde{\zeta}_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_2} \rangle \langle \tilde{\zeta}_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_3} \rangle + \text{cyclic} \quad (3.26)$$

where  $\lambda$  is some function of the slow-roll parameters. The correction terms will have a time dependence from  $\lambda$ , so the  $\langle \tilde{\zeta}_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_2} \tilde{\zeta}_{\mathbf{k}_3} \rangle$  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 [25]. Using (3.7) from [25], with  $f = -\varepsilon/(c_s^2 H)$ , we obtain the following form:

$$\begin{aligned}
 H_{\text{int}}(t) = \int d^3x \left\{ -\frac{a^3\varepsilon}{Hc_s^4} \left(-c_s^2 - 2c_s^2 \frac{\lambda}{\Sigma}\right) \dot{\zeta}^3 + \frac{a^3\varepsilon}{c_s^4} \left(-3c_s^2\right) \zeta \dot{\zeta}^2 \right. \\
 - \frac{a\varepsilon}{c_s^2} \left(-c_s^2\right) \zeta (\partial\zeta)^2 - \frac{a\varepsilon}{Hc_s^2} \dot{\zeta} (\partial\zeta)^2 \\
 \left. - \frac{a^3\varepsilon^2}{2c_s^4} (\varepsilon - 4) \dot{\zeta} \partial\zeta \partial(\partial^{-2}\dot{\zeta}) - \frac{a^3\varepsilon^3}{4c_s^4} \partial^2\zeta (\partial(\partial^{-2}\dot{\zeta}))^2 \right\}. \quad (3.27)
 \end{aligned}$$

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 [20] that using (3.25) in a scenario that results in an equilateral shape would require sensitive cancellations in the squeezed limit. Likewise, using (3.27) for a local scenario would require sensitive cancellations in the equilateral limit.

As mentioned in [20], the spatial derivatives can be manipulated into simple prefactors of  $k_i$  using the triangle condition ( $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ ), and so preserve the separability of the result. To absorb these prefactors in our calculation, we precompute  $k^p q_a(k)$  as a linear combination of the  $q_a(k)$  for the relevant values of  $p$ , from which  $V_{P\bar{P}}^{(i)}$  defined in (3.21) 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  $H_{\text{int}}$  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  $H_{\text{int}}$ .

### 3.4 Numerical considerations

The previous two sections detailed the methods required to calculate the coefficients  $\alpha_n$  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 (3.20), and integrals over  $k$  of the form (3.18). The first is done once per coefficient for each vertex, the second is a decomposition done once for  $\zeta_k$  and  $\zeta'_k$  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 (3.18), 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  $\tau_s$ ) to be  $3k_{\text{max}}$ . 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 (3.20) 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 [20], 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. [56]):

$$e^{-i\bar{k}(k_{\text{max}}-k_{\text{min}})\tau/2} = \sum_{n=0}^{\infty} (2n+1) i^n P_n(\bar{k}) j_n(-(k_{\text{max}}-k_{\text{min}})\tau/2) \quad (3.28)$$

for  $\bar{k}$  in  $[-1, 1]$ . When  $(k_{\text{max}}-k_{\text{min}})\tau/2$  is large, the spherical Bessel functions oscillate with an amplitude  $\propto \frac{1}{\tau}$ . Thus, the initial conditions (2.14) expanded in Legendre polynomials (and similar) give us suppression of  $1/\tau^3$  in (3.20).

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 [26] 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 [27] 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 [26, 27], 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 [57]). 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  $\int_{-\infty}^T f(\tau_s) e^{i w \tau_s} d\tau_s$  by replacing the non-oscillatory part  $f(\tau_s)$  with a function with matching value and derivative at  $\tau_s = T$ , but which converges far faster. We use

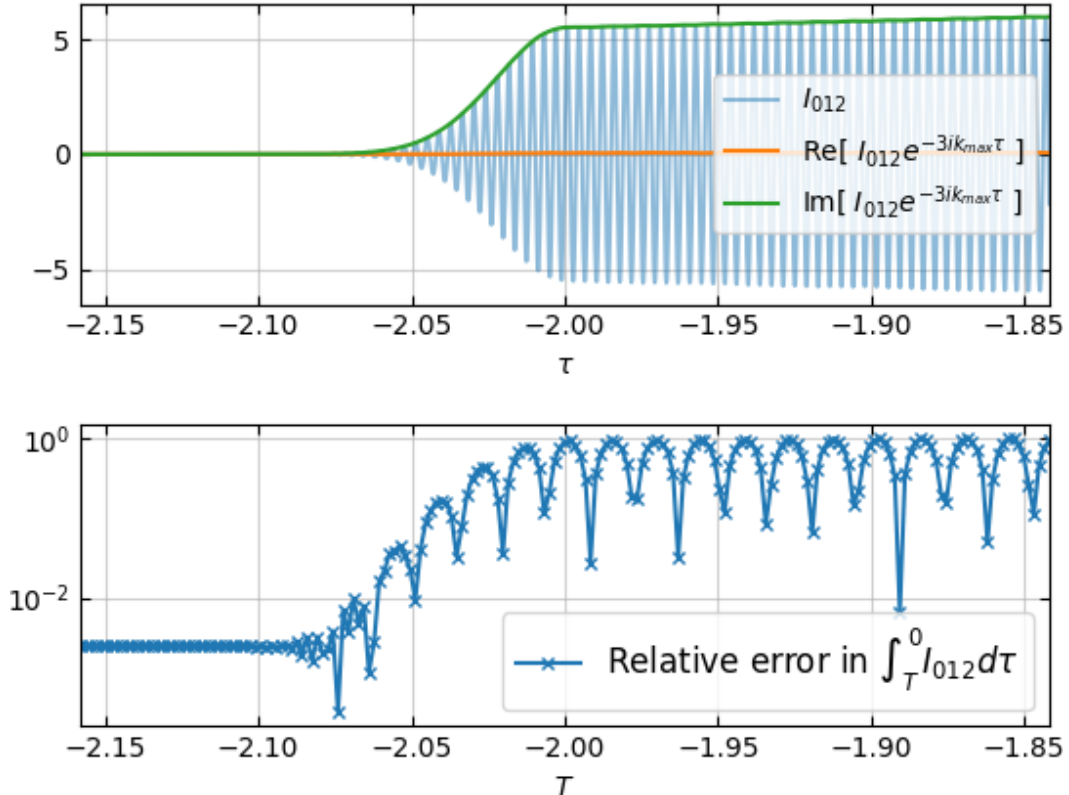
$$f(\tau_s) e^{-\beta^2(\tau_s - T)^2}, \quad (3.29)$$

for  $\tau_s < T$ . In this way, for sufficiently large  $T$ , we obtain the accuracy of the first two terms of the asymptotic expansion ( $O(\beta^2/w^2)$ ,  $w = 3k_{\max}$ ) 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).

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  $\beta$  is not too large their precise values have no significant effect on the final result. For definiteness, we take  $\beta/w = 1 \times 10^{-4}$ , small enough that the integrand has many oscillations while it is “turning on”, so matches the contribution of an infinite limit to high accuracy. We demonstrate this in figure 5 for a toy  $H_{\text{int}} = (-1/\tau)\zeta^3$ , as in (2.9).

While the demonstration in figure 5 is only on one coefficient of a simple model, we find this prescription is successful in accurately obtaining the full bispectrum information for more complicated models too. We use this prescription in section 4 to obtain results for feature and resonance models, which we validate as described in that section. For feature models, one must ensure that the point at which the damping ends ( $\tau_s = T$ ) is sufficiently early, before the time when  $\phi$  crosses the feature. For resonance models, the damping must similarly end before the earliest resonance occurs in any of the integrands, which will depend on the frequency of the feature. The values of  $\beta$  and  $T$  for which the integrals have sufficiently converged will therefore depend on the scenario being considered, but in all cases detailed in section 4 we found values of  $\beta$  and  $T$  which significantly reduced the computational time without a significant decrease in accuracy.

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



**Figure 5.** A toy example demonstrating the considerations involved in performing the time integrals (3.20). By carefully starting the time integrations, using the form (3.29), we can avoid errors that would otherwise swamp our result. The coefficient being calculated is the  $\alpha_{012}$  coefficient of the  $\mathcal{P}_0$  expansion of (2.9).

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 (3.18), 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 (3.18). 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  $\zeta_k$  samples than is practical. We want an expansion of the form

$$z_k e^{-ik\tau_s} = \sum_{n=0}^{\infty} \alpha_n q_n(k). \quad (3.30)$$

We can obtain this by using standard oscillatory quadrature, if the  $\tau_s$  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 (3.28).

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 (3.28)). 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 (3.24) 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 (3.17) of the  $F^{(i)}(\tau, k)$  converge, not until the actual sum (3.17) converges, since for high enough orders the integrals in (3.20) will integrate to zero.

Clearly, once  $\tau_s$  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, (3.20) and (3.18) 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 Validation

### 4.1 Validation methods

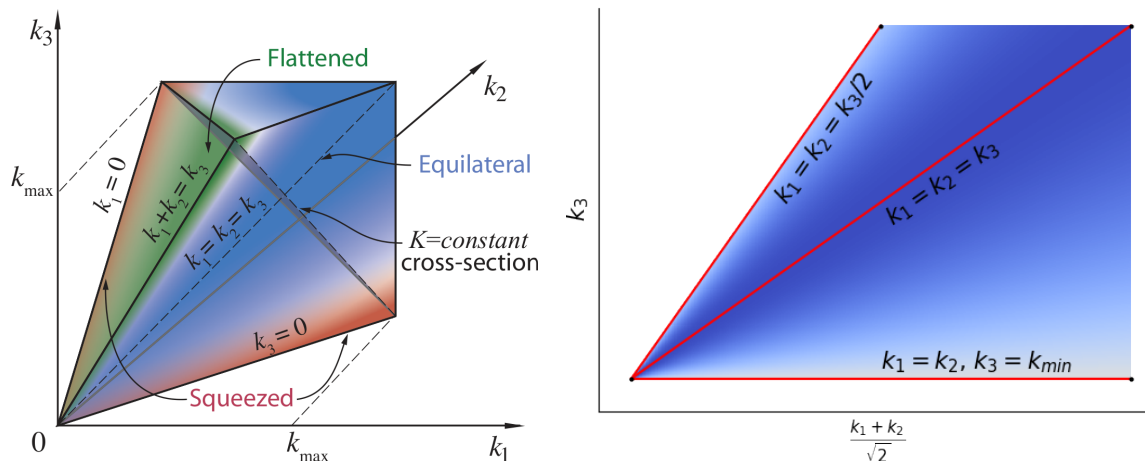
In this section we validate our implementation of our methods on different types of non-Gaussianity, sourced in different ways. While our actual results take the form of a set of mode expansion coefficients  $\alpha_n$ , to make contact with previous results in the literature all of our validation tests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

We test that our results have converged using (2.35), 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 (2.35), 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 (2.38). The consistency condition should hold most precisely at the configurations with smallest  $k_L/k_S$ , the most squeezed being the three corners,  $(k_{\max}, k_{\max}, k_{\min})$  and permutations. We want our test to be on an extended region of the tetrapyd however, so we choose the line

$$\frac{k_L}{k_S} = \frac{2k_{\min}}{k_{\max}}, \quad (4.1)$$

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 (2.18) and equilateral-type (2.20). We find that in our chosen basis  $\mathcal{P}_{01}^{n_s}$  our results converge quickly



**Figure 6.** 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 on the right. Horizontal lines on this plot have constant  $k_3$ . The bottom edge is  $k_3 = k_{\min}$ , the top edge is  $k_3 = k_{\max}$ . The right edge is  $k_1 = k_2 = k_{\max}$ , the left edge is  $k_1 = k_2 = k_3/2$ , i.e. the limit imposed by the triangle condition. Plotted in red (in the 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 on the left.

and robustly as we increase the number of modes, where we quantify the convergence using (2.35). We compare the converged results against analytic templates (2.24) and (2.27), using the full shape information (2.35), 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 (2.21). 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 [5], 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 (2.22). 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 [36], 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 6. 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 (2.37). 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.

## 4.2 Quadratic slow-roll

The first model we will consider is slow-roll inflation on a quadratic potential (2.18). 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  $\phi'_0$  according to the slow-roll approximation, we get  $\frac{1}{2}\phi'^2 = \epsilon \approx 0.2 \times 10^{-5}$ . 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  $\epsilon \approx 0.8 \times 10^{-2}$ . The shapes are shown in figure 7.

We choose the first scenario to have such a small value of  $\epsilon$  so that we can use Maldacena's shape (2.24) as a precision test. Indeed, we find that it has a scaled relative difference (2.36) of  $2.7 \times 10^{-5}$  with this shape, contrasting a scaled relative difference of 0.077 with the local template (2.26). 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 (2.24) or the local template (2.26), 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 (2.24) and (2.26), corresponding (in the sense of (2.36)) 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%.

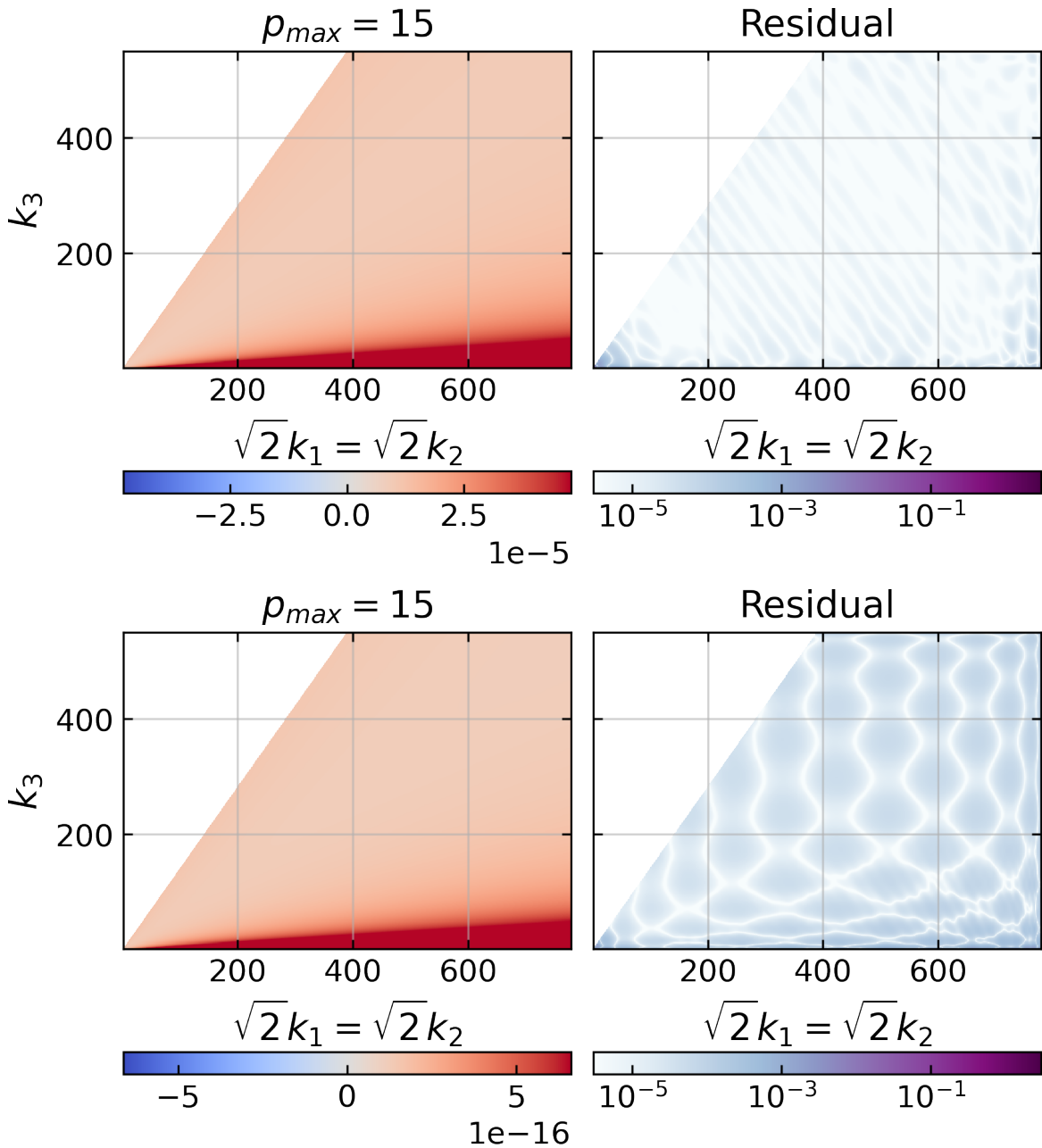
This is a validation of the convergence of our basis, reaffirming the template decomposition results of figure 1 in the setting of the in-in formalism. It is also a stringent validation of our methods of including the higher-order coefficients, as insufficient care taken in the early-time sections of integrals (3.20), or in including the spatial derivatives from  $H_{\text{int}}$ , could have easily swamped the  $p_{\text{max}} = 45$  result.

## 4.3 DBI inflation

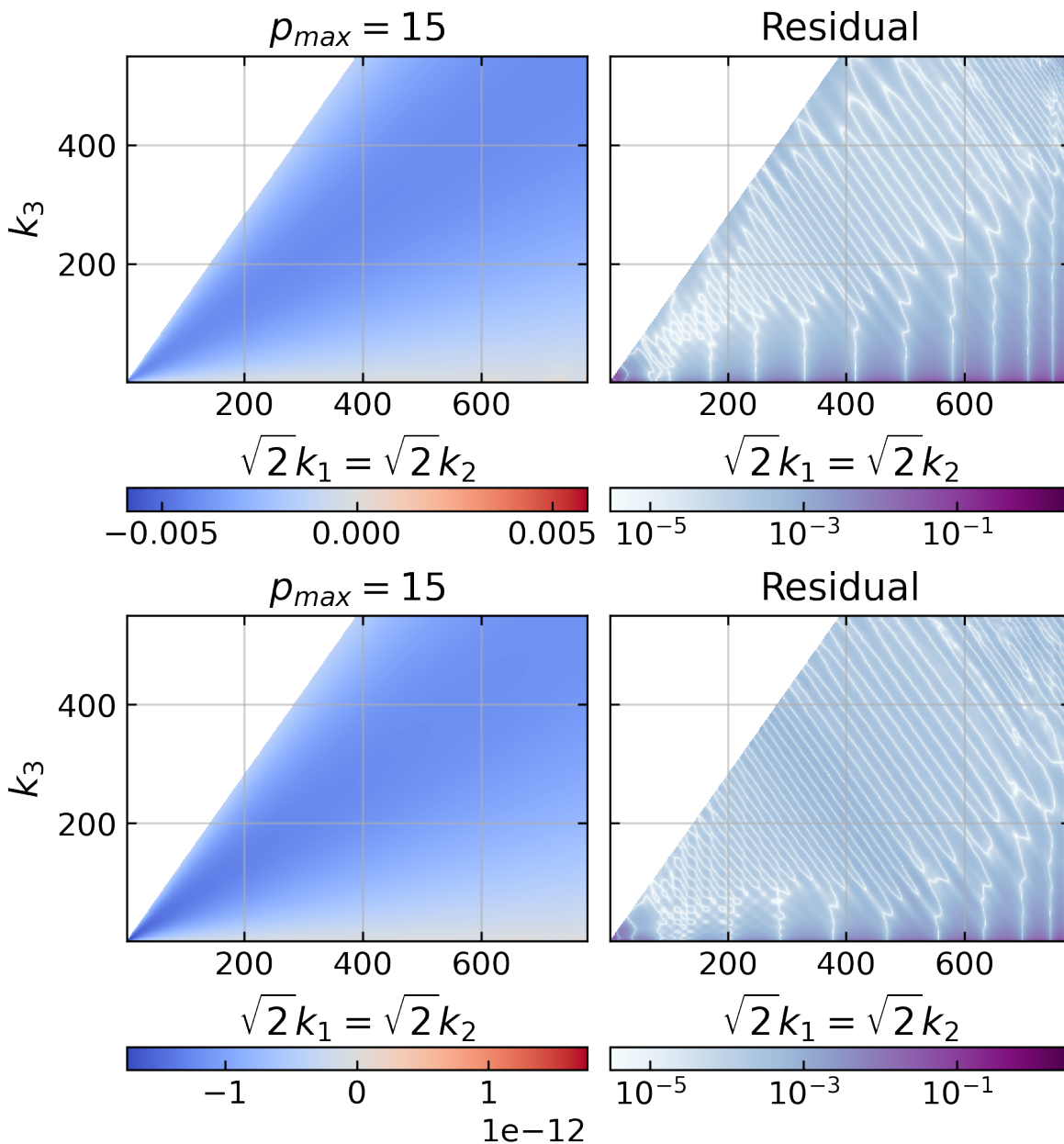
Next, we show results for a similar pair of scenarios for DBI inflation. We choose  $V_0 = 5.2 \times 10^{-12}$  with  $m = \sqrt{0.29V_0/3}$  in (2.19) and (2.20). We choose  $\phi_0 = 0.41$ , and then the starting condition for  $H$  according to the slow-roll approximation, allowing us to choose  $\phi'_0$  such that the Friedmann equation is satisfied to sufficient precision. The first scenario is deep in slow-roll, with  $\lambda_{\text{DBI}} = 1.9 \times 10^{18}$ , while the second scenario saturates the *Planck* limit on  $c_s$ , with  $\lambda_{\text{DBI}} = 1.9 \times 10^{15}$ . The resulting shapes are shown in figure 8.

The scenario deep in slow-roll has a error of 0.082% relative to the DBI shape (2.27), and 13% relative to the equilateral template (2.29). 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 (2.30), we get a relative error of 0.27%. On the line defined by (4.1), 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 3.1. 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.



**Figure 7.** A canonical single-field model on a quadratic potential (2.21), 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  $\varepsilon$ , 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  $2.7 \times 10^{-5}$  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  $7.9 \times 10^{-5}$  between  $p_{\max} = 45$  and  $p_{\max} = 15$ .



**Figure 8.** The upper plot shows the shape function for a DBI model deep in slow-roll. We set  $\lambda_{\text{DBI}}$  in (2.20) to  $1.9 \times 10^{18}$ , 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  $2.1 \times 10^{-3}$  between  $p_{\text{max}} = 45$  and  $p_{\text{max}} = 15$ . The lower plot shows a DBI model that saturates the *Planck* limit on  $c_s$ . We set  $\lambda_{\text{DBI}}$  in (2.20) to  $1.9 \times 10^{15}$ , 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  $1.1 \times 10^{-3}$  between  $p_{\text{max}} = 45$  and  $p_{\text{max}} = 15$ .

#### 4.4 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 figure 7. In (2.21) we fix  $d = 1 \times 10^{-2}$  and  $\phi_f = 15.55$  (as with the second canonical quadratic example,  $\phi_0 = 16.5$ ). Figure 9 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 [5], presented there in equations (48), (54), (55). We plot a comparison of the result of [5] and our result in figure 11. For larger step size, we check the squeezed limit in figure 10, 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 [20]. They validate the calculation of the high order coefficients, and show that our code as implemented can handle sharp deviations from slow-roll, generating non-Gaussianity around horizon crossing.

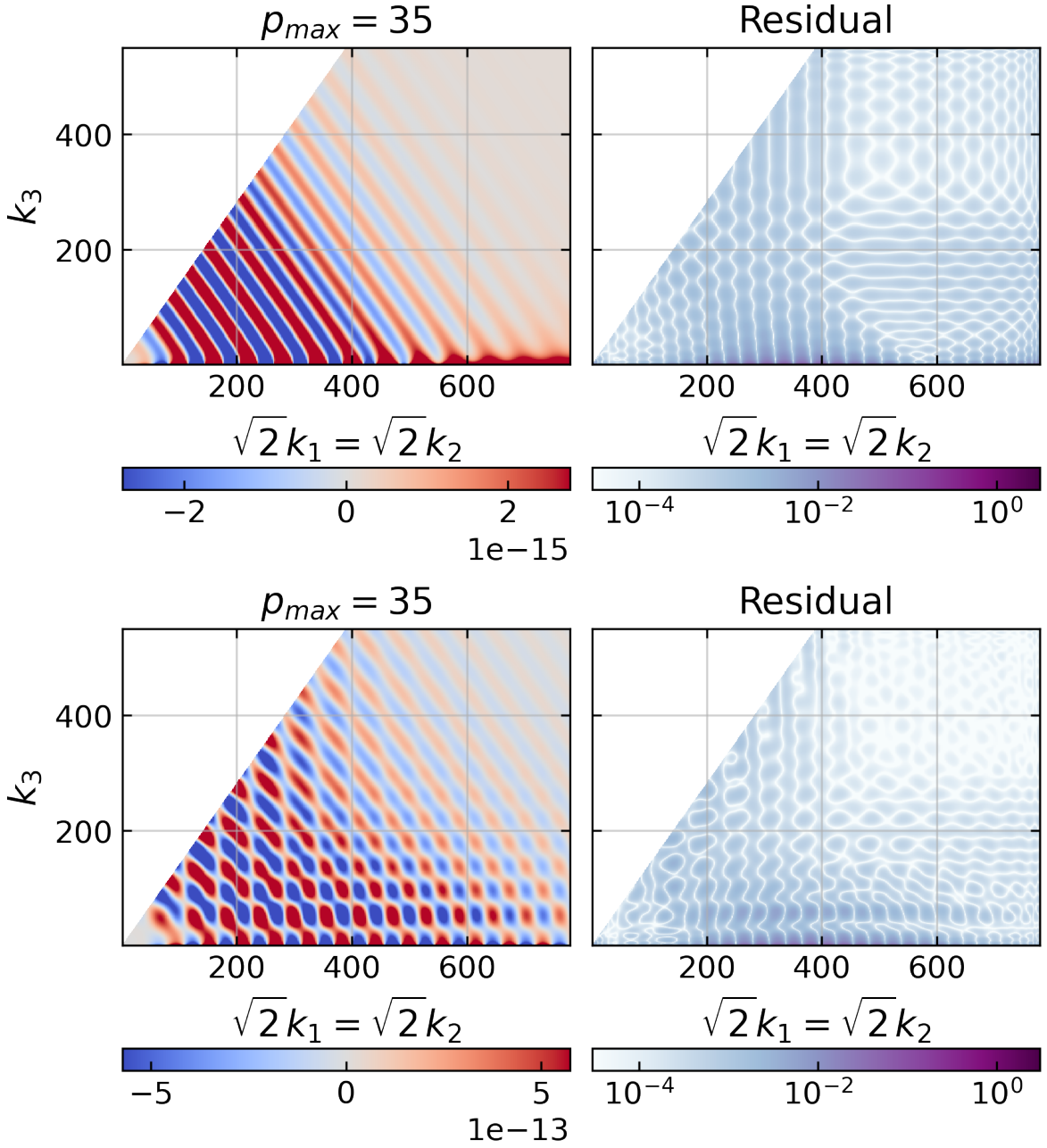
#### 4.5 Resonance features

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 12, is a model with a canonical kinetic term, on a quadratic potential with a superimposed oscillation (2.22). We take  $bf = 10^{-7}$ , and  $f = 10^{-2}$ . The resulting bispectrum has oscillations logarithmic in  $k_1 + k_2 + k_3$ . In figure 12 we see the excellent agreement between our result and the PyTransport result, once initial conditions in both codes are set early enough to achieve convergence. This validates the code on non-Gaussianity generated deeper in the horizon. Note the change of phase in the squeezed limit, though this is expected to be unobservable. We obtain a full tetrapyd convergence test result (between  $p_{\max} = 65$  and  $p_{\max} = 35$ ) of 0.93%, a squeezed limit test result of 1.1% (along the line defined by (4.1)), 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 12.

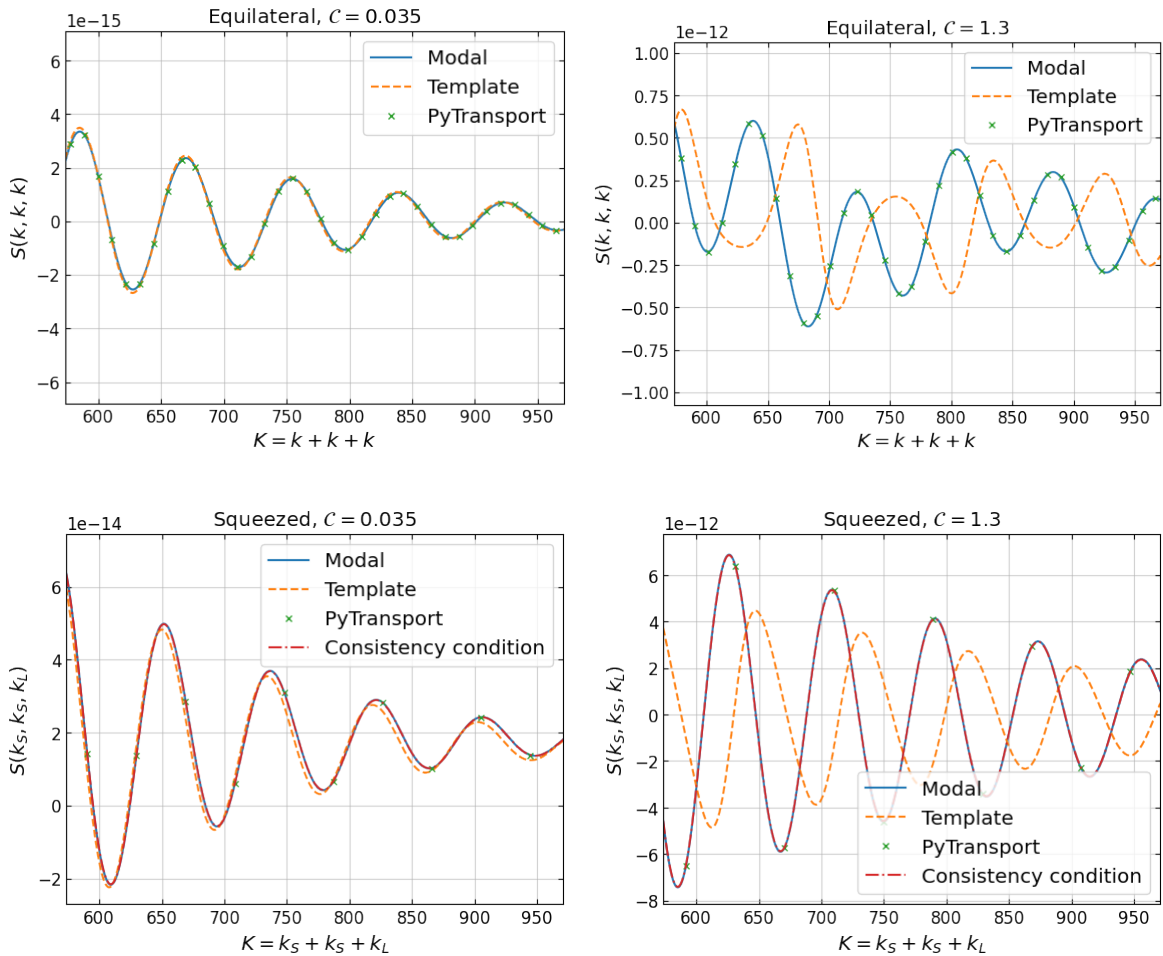
The time taken for the PyTransport code (per configuration) varies by a factor of around forty between the equilateral limit and the squeezed limit, as we show in figure 12. While the PyTransport code is extremely fast at calculating the shape function for a single  $k$ -configuration, to obtain this 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 13. 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





**Figure 9.** The tree-level shape function of a feature model (2.21), 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 [5],  $\mathcal{C} = 6c/(\varepsilon + 3c)$ , 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  $1.6 \times 10^{-3}$  and  $1.5 \times 10^{-3}$ , respectively, between  $p_{\max} = 65$  and  $p_{\max} = 35$ .

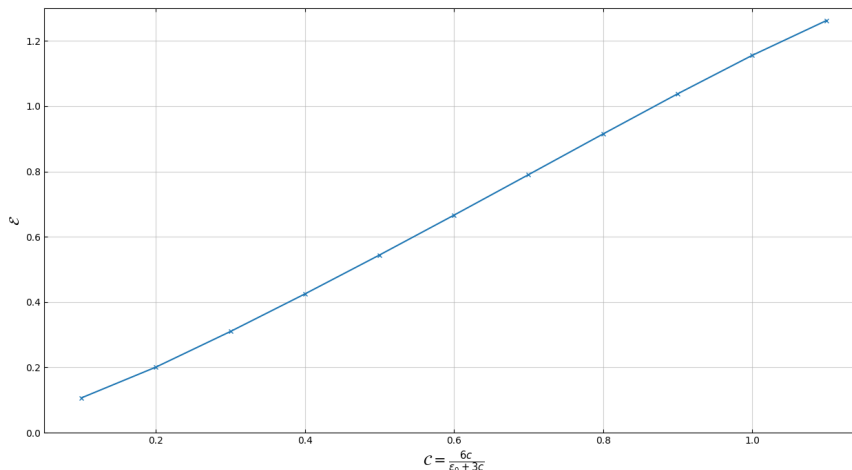


**Figure 10.** In the equilateral limit for the feature models (the top two figures) we validate our modal result against the PyTransport result. In the squeezed limit (the bottom two figures) we validate against PyTransport, and the consistency condition. In both limits, for both step sizes shown, we find excellent agreement. For the small step size (the two plots to the left), we additionally see a good match to the template of [5]. 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.

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 [30, 36]. 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%.

## 5 Discussion

In this work we have extended the modal methods of [17–19] 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

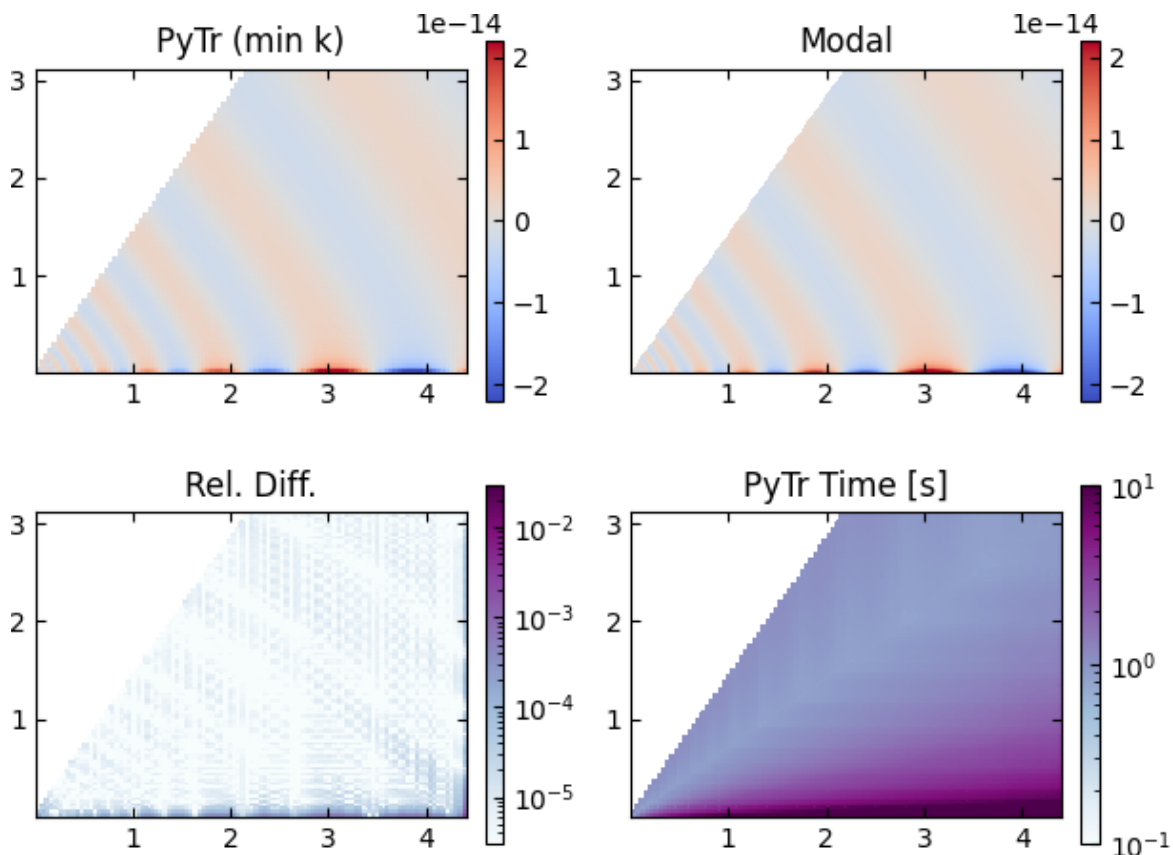


**Figure 11.** We sample more shapes with step sizes between the two feature models shown in figure 9. We plot the relative difference, integrated over the full tetrapyd in the sense of (2.35), between the modal result and the analytic template of [5], as a function of the template parameter  $\mathcal{C} = \frac{6c}{\epsilon_0 + 3c}$  (where  $c$  is the step size and  $\epsilon_0$  is the value of the slow-roll parameter  $\epsilon$  at  $\phi_{step}$  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 (2.21) occurs, which is kept constant across this scan. The width of the feature was also kept constant.

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 (2.2), the integral (2.7) 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 [20] 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 (3.20) 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-



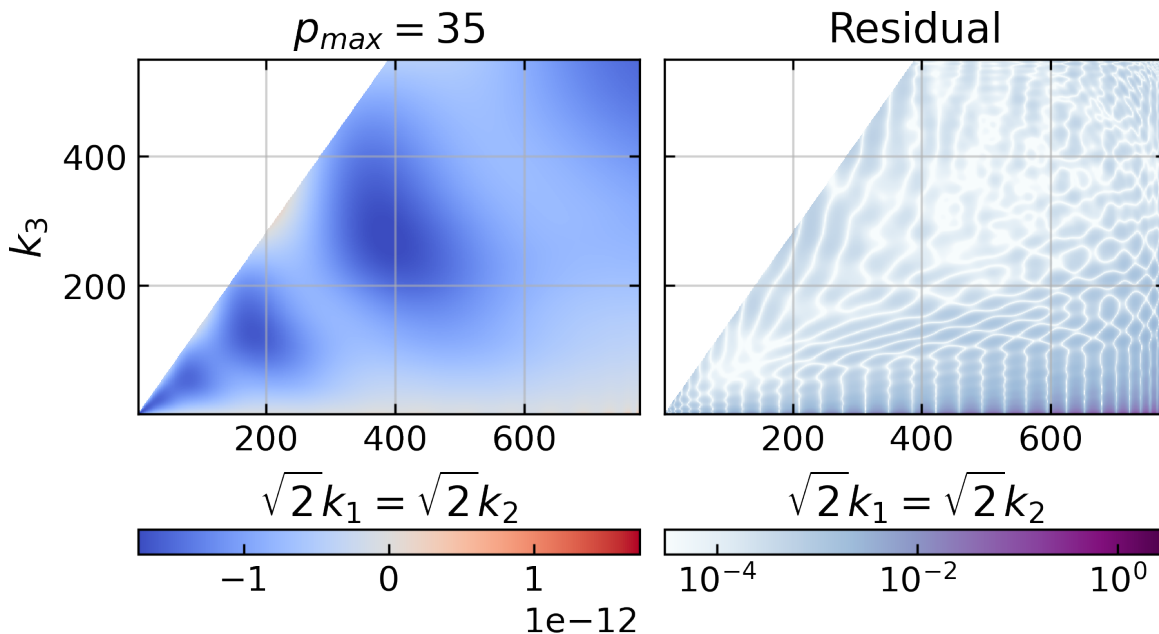
**Figure 12.** Resonance on a quadratic potential (2.22), testing our result using point tests against the PyTransport code. The logarithmic oscillations in the shape function are generated by periodic features deep in the horizon. The differences between our result and the PyTransport result are sufficiently 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  $9.6 \times 10^{-3}$  between  $p_{\max} = 65$  and  $p_{\max} = 35$ .

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 [10]. 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 [21]. 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.

## Acknowledgments

We are grateful to Wuhyun Sohn and James Fergusson for enlightening discussions and useful comments. PC thanks the Robert Gardiner Memorial Scholarship and the Cambridge European Scholarship for their support, through the Cambridge Trust, as well as the UK STFC. PS acknowledges funding from STFC Consolidated Grant ST/P000673/1.



**Figure 13.** Non-Gaussianity generated by periodic features in a DBI model, including a phase difference in the flattened limit as described in [36]. 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  $1.9 \times 10^{-3}$  between  $p_{\max} = 65$  and  $p_{\max} = 35$ .

## References

- [1] J.M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, *JHEP* **05** (2003) 013 [[astro-ph/0210603](#)] [[INSPIRE](#)].
- [2] C. Burrage, R.H. Ribeiro and D. Seery, *Large slow-roll corrections to the bispectrum of noncanonical inflation*, *JCAP* **07** (2011) 032 [[arXiv:1103.4126](#)] [[INSPIRE](#)].
- [3] M. Alishahiha, E. Silverstein and D. Tong, *DBI in the sky*, *Phys. Rev. D* **70** (2004) 123505 [[hep-th/0404084](#)] [[INSPIRE](#)].
- [4] C.T. Byrnes and K.-Y. Choi, *Review of local non-gaussianity from multifield inflation*, *Adv. Astron.* **2010** (2010) 1.
- [5] P. Adshead, C. Dvorkin, W. Hu and E.A. Lim, *Non-Gaussianity from Step Features in the Inflationary Potential*, *Phys. Rev. D* **85** (2012) 023531 [[arXiv:1110.3050](#)] [[INSPIRE](#)].
- [6] R. Flauger and E. Pajer, *Resonant Non-Gaussianity*, *JCAP* **01** (2011) 017 [[arXiv:1002.0833](#)] [[INSPIRE](#)].
- [7] E. Komatsu, D.N. Spergel and B.D. Wandelt, *Measuring primordial non-Gaussianity in the cosmic microwave background*, *Astrophys. J.* **634** (2005) 14 [[astro-ph/0305189](#)] [[INSPIRE](#)].
- [8] M. Münchmeyer, F. Bouchet, M.G. Jackson and B. Wandelt, *The Komatsu Spergel Wandelt estimator for oscillations in the cosmic microwave background bispectrum*, *Astron. Astrophys.* **570** (2014) A94 [[arXiv:1405.2550](#)] [[INSPIRE](#)].
- [9] PLANCK collaboration, *Planck 2015 results. XVII. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **594** (2016) A17 [[arXiv:1502.01592](#)] [[INSPIRE](#)].
- [10] PLANCK collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [[arXiv:1905.05697](#)] [[INSPIRE](#)].

- [11] SIMONS OBSERVATORY collaboration, *The Simons Observatory: Science goals and forecasts*, *JCAP* **02** (2019) 056 [[arXiv:1808.07445](#)] [[INSPIRE](#)].
- [12] CMB-S4 collaboration, *CMB-S4 Science Book, First Edition*, [arXiv:1610.02743](#) [[INSPIRE](#)].
- [13] T. Baldauf, U. Seljak and L. Senatore, *Primordial non-Gaussianity in the Bispectrum of the Halo Density Field*, *JCAP* **04** (2011) 006 [[arXiv:1011.1513](#)] [[INSPIRE](#)].
- [14] 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*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 1341 [[arXiv:1801.09280](#)] [[INSPIRE](#)].
- [15] X. Chen, C. Dvorkin, Z. Huang, M.H. Namjoo and L. Verde, *The Future of Primordial Features with Large-Scale Structure Surveys*, *JCAP* **11** (2016) 014 [[arXiv:1605.09365](#)] [[INSPIRE](#)].
- [16] R. Scoccimarro, L. Hui, M. Manera and K.C. Chan, *Large-scale Bias and Efficient Generation of Initial Conditions for Non-Local Primordial Non-Gaussianity*, *Phys. Rev. D* **85** (2012) 083002 [[arXiv:1108.5512](#)] [[INSPIRE](#)].
- [17] J.R. Fergusson and E.P.S. Shellard, *Primordial non-Gaussianity and the CMB bispectrum*, *Phys. Rev. D* **76** (2007) 083523 [[astro-ph/0612713](#)] [[INSPIRE](#)].
- [18] J.R. Fergusson and E.P.S. Shellard, *The shape of primordial non-Gaussianity and the CMB bispectrum*, *Phys. Rev. D* **80** (2009) 043510 [[arXiv:0812.3413](#)] [[INSPIRE](#)].
- [19] J.R. Fergusson, M. Liguori and E.P.S. Shellard, *The CMB Bispectrum*, *JCAP* **12** (2012) 032 [[arXiv:1006.1642](#)] [[INSPIRE](#)].
- [20] H. Funakoshi and S. Renaux-Petel, *A Modal Approach to the Numerical Calculation of Primordial non-Gaussianities*, *JCAP* **02** (2013) 002 [[arXiv:1211.3086](#)] [[INSPIRE](#)].
- [21] W. Sohn and J.R. Fergusson, *High-resolution CMB bispectrum estimator*, in preparation.
- [22] A. Slosar et al., *Scratches from the Past: Inflationary Archaeology through Features in the Power Spectrum of Primordial Fluctuations*, [arXiv:1903.09883](#) [[INSPIRE](#)].
- [23] P.D. Meerburg et al., *Primordial Non-Gaussianity*, [arXiv:1903.04409](#) [[INSPIRE](#)].
- [24] S. Weinberg, *Quantum contributions to cosmological correlations*, *Phys. Rev. D* **72** (2005) 043514 [[hep-th/0506236](#)] [[INSPIRE](#)].
- [25] S. Renaux-Petel, *On the redundancy of operators and the bispectrum in the most general second-order scalar-tensor theory*, *JCAP* **02** (2012) 020 [[arXiv:1107.5020](#)] [[INSPIRE](#)].
- [26] X. Chen, R. Easther and E.A. Lim, *Large Non-Gaussianities in Single Field Inflation*, *JCAP* **06** (2007) 023 [[astro-ph/0611645](#)] [[INSPIRE](#)].
- [27] X. Chen, R. Easther and E.A. Lim, *Generation and Characterization of Large Non-Gaussianities in Single Field Inflation*, *JCAP* **04** (2008) 010 [[arXiv:0801.3295](#)] [[INSPIRE](#)].
- [28] X. Chen, M.-x. Huang, S. Kachru and G. Shiu, *Observational signatures and non-Gaussianities of general single field inflation*, *JCAP* **01** (2007) 002 [[hep-th/0605045](#)] [[INSPIRE](#)].
- [29] D. Seery and J.E. Lidsey, *Primordial non-Gaussianities in single field inflation*, *JCAP* **06** (2005) 003 [[astro-ph/0503692](#)] [[INSPIRE](#)].
- [30] N. Bartolo, D. Cannone and S. Matarrese, *The Effective Field Theory of Inflation Models with Sharp Features*, *JCAP* **10** (2013) 038 [[arXiv:1307.3483](#)] [[INSPIRE](#)].
- [31] S. Passaglia and W. Hu, *Scalar Bispectrum Beyond Slow-Roll in the Unified EFT of Inflation*, *Phys. Rev. D* **98** (2018) 023526 [[arXiv:1804.07741](#)] [[INSPIRE](#)].
- [32] P. Adshead, W. Hu and V. Miranda, *Bispectrum in Single-Field Inflation Beyond Slow-Roll*, *Phys. Rev. D* **88** (2013) 023507 [[arXiv:1303.7004](#)] [[INSPIRE](#)].

- [33] V. Miranda, W. Hu and P. Adshead, *Warp Features in DBI Inflation*, *Phys. Rev. D* **86** (2012) 063529 [[arXiv:1207.2186](#)] [[INSPIRE](#)].
- [34] E. Silverstein and D. Tong, *Scalar speed limits and cosmology: Acceleration from D-cceleration*, *Phys. Rev. D* **70** (2004) 103505 [[hep-th/0310221](#)] [[INSPIRE](#)].
- [35] V. Miranda and W. Hu, *Inflationary Steps in the Planck Data*, *Phys. Rev. D* **89** (2014) 083529 [[arXiv:1312.0946](#)] [[INSPIRE](#)].
- [36] X. Chen, *Folded Resonant Non-Gaussianity in General Single Field Inflation*, *JCAP* **12** (2010) 003 [[arXiv:1008.2485](#)] [[INSPIRE](#)].
- [37] S. Ávila, J. Martin and D. Steer, *Superimposed Oscillations in Brane Inflation*, *JCAP* **08** (2014) 032 [[arXiv:1304.3262](#)] [[INSPIRE](#)].
- [38] S. Garcia-Saenz, S. Renaux-Petel and J. Ronayne, *Primordial fluctuations and non-Gaussianities in sidetracked inflation*, *JCAP* **07** (2018) 057 [[arXiv:1804.11279](#)] [[INSPIRE](#)].
- [39] J. Garriga and V.F. Mukhanov, *Perturbations in k-inflation*, *Phys. Lett. B* **458** (1999) 219 [[hep-th/9904176](#)] [[INSPIRE](#)].
- [40] G. Cabass, E. Pajer and F. Schmidt, *How Gaussian can our Universe be?*, *JCAP* **01** (2017) 003 [[arXiv:1612.00033](#)] [[INSPIRE](#)].
- [41] M. Münchmeyer, P.D. Meerburg and B.D. Wandelt, *Optimal estimator for resonance bispectra in the CMB*, *Phys. Rev. D* **91** (2015) 043534 [[arXiv:1412.3461](#)] [[INSPIRE](#)].
- [42] D.K. Hazra, L. Sriramkumar and J. Martin, *BINGO: A code for the efficient computation of the scalar bi-spectrum*, *JCAP* **05** (2013) 026 [[arXiv:1201.0926](#)] [[INSPIRE](#)].
- [43] J.S. Horner and C.R. Contaldi, *Non-Gaussianity Unleashed*, [arXiv:1303.2119](#) [[INSPIRE](#)].
- [44] J.S. Horner and C.R. Contaldi, *Non-Gaussian signatures of general inflationary trajectories*, *JCAP* **09** (2014) 001 [[arXiv:1311.3224](#)] [[INSPIRE](#)].
- [45] J.S. Horner and C.R. Contaldi, *The bispectrum of single-field inflationary trajectories with  $c_s \neq 1$* , [arXiv:1503.08103](#) [[INSPIRE](#)].
- [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* **12** (2016) 033 [[arXiv:1609.00379](#)] [[INSPIRE](#)].
- [47] D.J. Mulryne and J.W. Ronayne, *PyTransport: A Python package for the calculation of inflationary correlation functions*, *J. Open Source Softw.* **3** (2018) 494 [[arXiv:1609.00381](#)] [[INSPIRE](#)].
- [48] J.W. Ronayne and D.J. Mulryne, *Numerically evaluating the bispectrum in curved field-space — with PyTransport 2.0*, *JCAP* **01** (2018) 023 [[arXiv:1708.07130](#)] [[INSPIRE](#)].
- [49] S. Butchers and D. Seery, *Numerical evaluation of inflationary 3-point functions on curved field space — with the transport method  $\mathcal{E}$  CppTransport*, *JCAP* **07** (2018) 031 [[arXiv:1803.10563](#)] [[INSPIRE](#)].
- [50] 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:1902.01830](#) [[INSPIRE](#)].
- [51] P. Creminelli and M. Zaldarriaga, *Single field consistency relation for the 3-point function*, *JCAP* **10** (2004) 006 [[astro-ph/0407059](#)] [[INSPIRE](#)].
- [52] P. Creminelli, G. D’Amico, M. Musso and J. Noreña, *The (not so) squeezed limit of the primordial 3-point function*, *JCAP* **11** (2011) 038 [[arXiv:1106.1462](#)] [[INSPIRE](#)].
- [53] N. Arkani-Hamed, D. Baumann, H. Lee and G.L. Pimentel, *The Cosmological Bootstrap: Inflationary Correlators from Symmetries and Singularities*, *JHEP* **04** (2020) 105 [[arXiv:1811.00024](#)] [[INSPIRE](#)].

- [54] F. Arroja and T. Tanaka, *A note on the role of the boundary terms for the non-Gaussianity in general  $k$ -inflation*, *JCAP* **05** (2011) 005 [[arXiv:1103.1102](#)] [[INSPIRE](#)].
- [55] G. Rigopoulos, *Gauge invariance and non-Gaussianity in Inflation*, *Phys. Rev. D* **84** (2011) 021301 [[arXiv:1104.0292](#)] [[INSPIRE](#)].
- [56] A. Dixit, L. Jiu, V.H. Moll and C. Vignat, *The finite Fourier transform of classical polynomials*, [arXiv:1402.5544](#).
- [57] A. Iserles and S.P. Norsett, *Efficient quadrature of highly efficient integrals using derivatives*, *Proc. Roy. Soc. Lond. A* **461** (2005) 1383.