PROBING THE PHYSICS OF INFLATION: SEARCHING FOR NON-GAUSSIANITY IN THE COSMIC MICROWAVE BACKGROUND AND LARGE-SCALE STRUCTURE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Joyce Bobae Byun

August 2015

PROBING THE PHYSICS OF INFLATION: SEARCHING FOR NON-GAUSSIANITY IN THE COSMIC MICROWAVE BACKGROUND AND LARGE-SCALE STRUCTURE

Joyce Bobae Byun, Ph.D.

Cornell University 2015

Over thirty years since an early epoch of inflation was proposed as an elegant way solution to the Big Bang's flatness, horizon, and monopole problems, recent theoretical and observational advances have made it possible to robustly test the physics of inflation using a variety of cosmological probes. It is now known that the higher-order correlators of primordial fluctuations beyond the power spectrum, starting with the bispectrum (or 'shape'), is highly sensitive to the specific physical model of inflation, and this has motivated an intense effort to robustly extract bispectrum constraints from current and upcoming data sets. This thesis aims to build on existing methods for reconstructing the amplitude and wavenumber-dependence of primordial bispectra and making robust inferences about the inflationary physics of our Universe, given cosmic microwave background (CMB) data from Planck and future large-scale structure (LSS) surveys, such as DESI and Euclid.

We present forecasted constraints from CMB bispectrum data for a range of inflationary shapes, ranging from fixed individual templates, which are meant to represent distinctive features of large classes of physical models, to more general shapes and joint constraints on sets of shapes, which are important for quantifying how well the data may be able to detect bispectrum signatures that are not already theoretically predicted, and how well it may distinguish be-

tween competing theories. In addition, to better model general bispectra that may have complex behavior on squeezed triangle configurations of (k_1, k_2, k_3), we created an alternative divergent basis for reconstructing bispectra. Next, we extend this analysis to a wholly different probe of inflation that will become available when large galaxy surveys come online in the next decade: the scale-dependent halo bias as measured from galaxy power spectra. Our analysis shows that the two probes of bispectra from the CMB and LSS can be quite complementary, especially if we can obtain a strong prior on the galaxy bias from external data sets. We also show that although previous analyses have focused on the large-scale halo bias, the small-scale halo bias also contains information that can be used to constrain and discriminate between models. Finally, we narrow our focus to a class of inflationary theories that have non-trivial initial quantum states for the inflaton, as these models have in the past not been well-constrained due to the limitations of current reconstruction methods. We present an alternative basis, derived from b-splines, which can reconstruct the sharp localized peaks and oscillatory features typical of bispectra in these models, and show that the b-spline basis is competitive with other choices of basis, partly because b-spline basis functions are more easily generated.

BIOGRAPHICAL SKETCH

Joyce was born in Texas, and grew up in the suburbs of southeast Houston with her parents, brother, and sister. Her early interests in science veered towards physics in high school, and she has been studying astrophysics since learning that many people have successfully made a career out of it. In 2009, she graduated from the University of Texas at Austin with bachelors degrees in Physics and Astronomy, and moved to Ithaca to continue her studies as a graduate student in the Department of Astronomy at Cornell University.

For my parents.

ACKNOWLEDGEMENTS

I am very grateful to my advisor, Rachel Bean, for always making time to discuss research with me and to offer advice and guidance these past six years, with incredible patience and constant encouragement. I also benefitted professionally from the many varied opportunities she gave me, to attend several conferences, meet and work with new collaborators, and to teach and do research with undergraduate students.

I am thankful to have had the opportunity to collaborate with Nishant Agarwal and Richard Holman for the work presented in Chapter 4. If in the future I have a chance to host and work with other collaborators, I will try to follow their example by remembering my discussions with them and my visits to Carnegie Mellon and Penn State, which I enjoyed very much.

I also benefited from many conversations with other researchers, including Tommaso Giannantonio, Tom Loredo, Daan Meerburg, and Sarah Shandera, and my faculty committee consisting of David Chernoff, Eanna Flanagan, and Martha Haynes, all of whom had useful comments and insightful questions that helped to improve this work.

I am indebted to the National Science Foundation, NASA, and Cornell University, for funding me throughout my graduate career, and to the Astronomy department staff, especially Monica Armstrong and Cheryl Neville, who helped to keep me officially enrolled, reimbursed, and well-fed with free food, during my graduate studies.

I think without the support of my family and friends, it would have been much harder, if not impossible, to survive the sometimes isolating and frustrating periods of my studies. I feel very fortunate to have had the cameraderie of Cornell's Astronomy graduate students and postdocs, who are amazingly

smart, cool, and fun. I am especially thankful to Mike Jones for his steady companionship and cheerful humor. I would also like to thank Mike's relatives for making Ithaca feel like my home away from home. Finally, I am most grateful to my parents for their unwavering support and many sacrifices that made all of this possible.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Table of Contents . . . vii
List of Tables . . . ix
List of Figures . . . xi

1 Introduction **1**

2 Non-Gaussian Shape Recognition **12**
2.1 Introduction . . . 12
2.2 Efficient calculation of a general non-Gaussian shape . . . 16
2.2.1 The CMB bispectrum . . . 17
2.2.2 Shape similarity . . . 19
2.2.3 A new separable basis, $\mathcal{K}_n$ . . . 22
2.2.4 Application of the basis to shapes arising in inflationary theory . . . 27
2.2.5 Shape classification and depiction . . . 35
2.3 Forecasting constraints on general shapes . . . 38
2.3.1 Fisher matrix approach . . . 39
2.3.2 Fisher matrix results . . . 42
2.3.3 Drawbacks of normalization at the equilateral configuration 49
2.3.4 Best measured k-configurations . . . 52
2.3.5 Shape determination and distinguishability . . . 55
2.4 Conclusion . . . 59

3 Non-Gaussian Shape Discrimination with Spectroscopic Galaxy Surveys **65**
3.1 Introduction . . . 65
3.2 Formalism . . . 69
3.2.1 Bispectrum shape families and templates . . . 69
3.2.2 Non-Gaussian halo statistics . . . 74
3.2.3 The halo model . . . 77
3.2.4 The galaxy bias model . . . 79
3.2.5 Fisher matrix approach for LSS and CMB . . . 81
3.3 Findings . . . 86
3.3.1 Squeezed limit behavior . . . 86
3.3.2 Configurations probed by the halo bias and mass function 91
3.3.3 Prospective constraints from upcoming LSS and CMB surveys . . . 93
3.4 Discussion & Conclusions . . . 100

4 A New Basis for Non-separable Bispectra **111**
4.1 Introduction . . . 111
4.2 Non-Bunch-Davies shapes . . . 114
4.3 The spline basis . . . 119
4.4 Fitting non-Bunch-Davies shapes . . . 121
4.4.1 $S_{\rm NBD1}$ and $S_{\rm NBD2}$ . . . 124
4.4.2 $S_{\dot\zeta^3}$, $S_{\zeta\dot\zeta^2}$, and $S_{\zeta(\partial\zeta)^2}$ with k-independent $c_s\eta_0$ . . . 126
4.4.3 $S_{\dot\zeta^3}$, $S_{\zeta\dot\zeta^2}$, and $S_{\zeta(\partial\zeta)^2}$ with k-dependent $c_s\eta_0$ . . . 126
4.5 Discussion . . . 127

5 Conclusions and Future Work **130**

A Chapter 1 of appendix **136**
A.1 Power spectrum for general initial states . . . 136
A.2 Bispectrum for general initial states . . . 139
A.2.1 Flattened limit . . . 140
A.2.2 Squeezed limit . . . 141
A.3 B-splines fitting algorithm . . . 141

Bibliography **152**

LIST OF TABLES

2.1 The set of $\mathcal{K}_n$ with the allowed combinations of $\{p, r, s\}$ that satisfy $p + r + s = 0$. The ordering of the modes is according to their divergence behavior in the squeezed limit. The coefficients of their divergent terms in this limit, d_{nm}, are trivially related to the values of $\{p, r, s\}$, but are given here for convenience. We also give the equivalent short-hand notation used in [1] and summarized in (2.23). 26

2.2 Shapes constructed from basis modes with maximal divergence x_{sq}^R ($R < 0$) which, through cancellations of the divergent terms, have a squeezed limit that diverges as x_{sq}^r. These represent an irreducible set of component shapes, for each value of R, from which general, scale invariant, separable shapes can be constructed. 32

2.3 Correlation coefficients between shapes that diverge as x_{sq}^n. These shapes are components in the general template classes, $S_{[-2,r]}$, for $r \leq n$. 44

2.4 The uncertainties on the amplitudes of the component shapes, in the general template classes $S_{[-2,r]}$, that diverge as x_{sq}^r in the squeezed limit. We give both the unmarginalized errors, assuming the underlying shape is exactly described by the component shape, and the marginalized errors if we allow the shape to be a general linear combinations of components consistent with the prior on the divergence properties. 46

2.5 Properties of the principal components for each template class $S_{[-2,r]}$ in terms of their component shapes. The properties in the squeezed limit is determined by the value of r. The table provides uncertainties, for a unit norm eigenvector, $\sigma(b_i)$, and an effective $\sigma(f_{NL}(\hat{e}_i))$, when the eigenvector is normalized consistently at the equilateral configuration. 47

3.1 The $\mathcal{K}_n$ expansion for a variety of templates discussed in the literature. The variables p and α, used in $S_{orth(2)}$ and $S_{enf(2)}$, are chosen to maximize the template's fit to the physical shape. The separable template for S_{SOSF} is used for the CMB analysis, and is constructed from combinations of the basis functions that vanish in the squeezed limits, with coefficients that maximize the cosine with the fitting function given in [2], giving a cosine of > 0.99. . 73

3.2 The squeezed limit expansion, in powers of k/q as $k/q \to 0$, for a variety of templates described in section 3.2.1. 87

3.3 Fisher constraints on f_{NL} for templates with varied divergences and μ-dependent properties in the squeezed limit. We present the forecasted constraints for a Planck-like CMB experiment, and from the halo properties of galaxies measured by a Euclid-like spectroscopic survey [top] and a DESI-like survey [bottom]. The LSS constraints are shown without CMB and with (in parentheses) and with four different assumptions on the amount of information known about the galaxy bias: assuming no knowledge and marginalizing over the bias in independent redshift bins ('marg.'), assuming a 10% and 1% prior knowledge, and assuming the extreme limit, in which the galaxy bias has been measured precisely by a complementary method. 94

LIST OF FIGURES

1.1 A simple inflationary toy model, with a single scalar field, ϕ, called the 'inflaton', and an associated potential, $V(\phi)$. Accelerated expansion occurs when the potential energy dominates the kinetic energy of the field, $\frac{1}{2}\dot{\phi}^2$, which is initially small for a slowly rolling inflaton, and ends when the kinetic energy grows to be comparable to the potential energy, after which the field value will oscillate around the potential's minimum. As the oscillations damp out, the inflaton's energy is transferred into the creation of standard model particles, which heat the universe, and begin the hot Big Bang. Figure taken from [3]. 3

1.2 This diagram, taken from [3], illustrates the changing comoving horizon (red) with respect to a density fluctuation on a fixed scale (blue). During inflation, the shrinking comoving horizon causes density fluctuations to shift outside the horizon, with the fluctuations on the largest scales leaving first. This effectively freezes in the fluctuation until it re-enters the horizon at a later time, after the onset of the Big Bang. Inflation provides a solution to the horizon problem, by allowing the largest-scale fluctuations observable today to have been inside the horizon before the Big Bang. 4

1.3 Parts (a)-(e) of this figure from [4] illustrate a variety of triangle configurations that make up the full space of allowed (k_1, k_2, k_3) combinations, representing the space over which a bispectrum's amplitude, $B_\Phi(k_1, k_2, k_3)$, is defined. Each one is labeled with a common name, such as 'squeezed triangle', that is often used in the literature and throughout this work to correspond to similar-looking triangles. Additionally, bispectra are often plotted on figures like the one in the bottom right panel. Bispectra are always symmetric under interchange of the wavenumbers, and many bispectra are scale-invariant, meaning they can be reduced to functions of k_2/k_1 and k_3/k_1, allowing all of the unique information contained within a bispectrum function to be shown as a contour plot over the domain shown in the last panel. 8

2.1 The application of the separable basis to describe shapes motivated by theoretically distinct models that span a wide array of configurations in k-space. The local, equilateral, and orthogonal templates are explicitly separable, and are included only for reference. The other shapes are not separable but are approximated by templates using linear combinations of $\{\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2\}$. We construct an orthonormal basis, $\mathcal{R}_n$, in k-space with uniform weighting, using a Gram-Schmidt decomposition for $\mathcal{K}_n$, for $0 \leq n \leq 2$, starting with $n = 0$. [Left panel] The cosines between each shape and the $\mathcal{R}_n$. [Right panel] The cumulative cosine between a constructed template $\sum_{i=0}^{n} c_i \mathcal{R}_i$ and the true shape. 31

2.2 Plots showing the comparative spatial distribution of non-Gaussian shape, $S(k_1, k_2, k_3)$, as a function of k_1/k_3 and k_2/k_3. From left to right we show [top] the local, equilateral, and enfolded separable templates, [middle] the orthogonal(2) template, and non-separable shapes derived from a QSFI model with $\nu = 1.3$ and Galileon inflation, and [bottom] shapes contributing to $S_{[-2,1]}$. All shapes are normalized to unity at the equilateral configuration ($\frac{k_1}{k_3} = \frac{k_2}{k_3} = 1$). The color scales for all but the local and QSFI shapes are the same to aid comparison. 37

2.3 Configurations of the principal components for $S_{[-2,-2]}$, a general shape that can be as divergent as x_{sq}^{-2} in the squeezed limit. The plots show the amplitude of the eigenvectors for the best $\hat{e}_1$ to worst $\hat{e}_7$ measured modes as a function of $\frac{k_1}{k_3}$ versus $\frac{k_2}{k_3}$. The principal components are each normalized to be unity at the equilateral configuration. 48

2.4 As in Figure 2.3, but showing the configurations of the principal components for $S_{[-2,1]}$, a general shape that vanishes in the squeezed limit. 51

2.5 Example contour slices of $\sigma(f_{NL}S)$ under different divergence constraints, [top left] $S_{[-2,-2]}$, [top right] $S_{[-2,1]}$, and [bottom] $S_{[-1,-1]}$. 54

2.6 Detection thresholds on the amplitude of templates, f_{NL}^{T}, for distinguishing between the template and a general shape, $S = (1-\alpha)S_{equil} + \alpha S_X$, at the 1σ confidence level. [Left] $S_X = S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$, [right] $S_X = 2\mathcal{K}_3 - \mathcal{K}_6$. Blue, orange, red, and black curves denote f_{NL}^{equil}, f_{NL}^{orth}, f_{NL}^{local}, and $f_{NL}^{2\mathcal{K}_3-\mathcal{K}_6}$, respectively. Since $S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$ is very similar to $2\mathcal{K}_3 - \mathcal{K}_6$, the case where $S_X = S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$ is not shown. 59

2.7 Detection thresholds on the amplitude of templates, [left] f_{NL}^{equil}, [center] f_{NL}^{orth}, and [right] $f_{NL}^{2\mathcal{K}_3-\mathcal{K}_6}$, for distinguishing between each template and two forms of a general shape S at the 1σ confidence level. The general shapes considered are [top panels] $S_{gen} = (1 - \alpha_X - \alpha_Y)S_{equil} + \alpha_X(S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3) + \alpha_Y(S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3)$ or [bottom panels] $S_{gen} = (1 - \alpha_X - \alpha_Y)S_{equil} + \alpha_X(S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3) + \alpha_Y(2\mathcal{K}_3 - \mathcal{K}_6)$. Contours for f_{NL}^{loc} are not pictured, because the marginalized $\sigma(f_{NL}^{loc})$ remains close to its unmarginalized value over these 2-dimensional spaces. The case where $S_X = S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$ and $S_Y = 2\mathcal{K}_3 - \mathcal{K}_6$ is not pictured because S_X and S_Y are nearly uncorrelated, thus no additional information is revealed beyond that in Figure 2.6. 60

3.1 Fiducial Gaussian galaxy biases in this work from [5] [blue circles] and computed from the halo model we consider for ELGs [black]. Other bias parametrizations in the literature are shown for comparison: $b_{gal}^G(z) = \sqrt{1+z}$ [red] [6], and ELG bias $b_{gal}^G(z) = 0.84/D(z)$ [blue dashed] [7]. 82

3.2 The left panels, from top to bottom, shows the scale-dependent halo bias using the analytic derivation [thick], as in (3.15), and the asymptotic form [8] if extrapolated to small scales [thin], for two halo masses, $10^{12.5}M_\odot/h$ [red] and $10^{13.5}\ M_\odot/h$ [blue], for the local, orthogonal, and equilateral shapes at $z = 1$ for $f_{NL} = 1$. The right panels, from top to bottom, show the scale-dependent bias, for the same halo masses, for the enfolded, anisotropic, and SOSF shapes, respectively. Dashed curves denote negative values. 89

3.3 The halo bias weight function, w_{sd}, for halo masses $m = 10^{12.5}M_\odot/h$ [top] and $m = 10^{13.5}M_\odot/h$ [middle], and scales $k = 10^{-3}\,h$/Mpc [left] and $k = 0.1\,h$/Mpc [right] as a function of the wavenumber q and angle μ used to described the bispectrum shape. We have labeled specific configurations at which each weight shown peaks. Some points are simply labeled as 'Sq' signifying that the triangular configuration is significantly squeezed and not easily illustrated to scale. For weights at $k = 0.1\,h$/Mpc, we label other configurations and illustrate the corresponding triangles, which are much less squeezed than those labeled 'Sq', showing that Δb_{sd} on small scales can probe configurations beyond very squeezed triangles. 104

3.4 The solid and dashed curves respectively show the $\mu = 0$ slice of the weight from the scale-dependent bias for $k = 10^{-3}\,h$/Mpc and $M = 10^{13.5}M_\odot/h$ and $M = 10^{12.5}M_\odot/h$ cases. 105

3.5 The weight function $w_{R_{NG}}$, parameterizing the shape sensitivity in non-Gaussian corrections to the the halo mass function, as defined in 3.29. The function is plotted for a fixed $k = 0.1\,h/\mathrm{Mpc}$ slice, where the weight is peaked, and two halo masses, $m = 10^{12.5} M_\odot/h$ (left) and $m = 10^{13.5} M_\odot/h$ (right). The plots illustrate that there are non-negligible contributions from configurations that are not significantly squeezed, like those labeled E and F. . . 105
3.6 The effect of the non-Gaussian corrections to the halo mass function [left panel] and the scale-independent halo bias [right panel] on the halo power spectrum relative to the Gaussian case P_h^G, for the local [full blue], equilateral [dotted red], and orthogonal [dashed black] shapes at $z = 1$ for $f_{NL} = 100$. 106
3.7 The predicted galaxy bias as a function of scale for luminous red galaxies (LRGs) at $z = 0.75$ [solid lines], and emission line galaxies (ELGs) at $z = 0.75$ [dotted] and $z = 1.55$ [dashed], for a variety of non-Gaussian bispectrum templates with $f_{NL} = 100$: local [blue] and orthogonal [black] (left panel), as well as equilateral [red] (right panel). 106
3.8 Contributions to a Fisher element as a function of k, $F_z(k)$, for (f_{NL}, f_{NL}) in a fixed redshift slice, $z = 2$ [left] or $z = 1$ [right], from the local [blue], equilateral [dotted red], and orthogonal [dashed black] templates. Contributions to the Fisher element for (b_{gal}^G, b_{gal}^G) are also shown [dot dashed gray], multiplied by a scaling factor of 10^{-8} in order to show on the same plot. Unless the primordial shape has a strongly scale-dependent bias on large-scales, most of the resulting constraint on f_{NL} comes from intermediate and small scales, $k \gtrsim 0.05\,h/\mathrm{Mpc}$. The constraint on the Gaussian galaxy bias, while it has a larger overall amplitude, its shape is qualitatively similar to the signal on small scales coming from f_{NL}. 107
3.9 Euclid-like [top] and DESI-like [bottom] Fisher matrix projections for the constraints on the amplitudes of contributions to a general shape constructed from local, equilateral and orthogonal (LEO) templates. Constraints from the halo bias only [blue], CMB only [red], and in combination [black] are shown for a range of assumptions on prior knowledge of the galaxy bias: the most conservative scenario [full lines] in which all Gaussian bias information is obtained from this dataset and b_{gal}^G is marginalized over fully, the opposite regime [dotted lines], in which galaxy bias are wholly determined by an alternative technique, such as those discussed in section 3.2.4, and is not marginalized over, and the optimistic scenario [dashed lines] in which the Gaussian bias is well-constrained which we represent using a 1% prior. . . 108

3.10 Euclid-like Fisher matrix projections for the joint constraints on the amplitudes of a pair of shapes. Constraints from the halo bias only [blue], CMB only [red], and in combination [black] are shown for scenarios in which b_{gal}^{G} is marginalized over fully [full], in which galaxy bias is wholly determined by an alternative technique [dotted lines], and in which the Gaussian bias is well-constrained, which we represent using a 1% prior [dashed lines]. 109
3.11 Joint LSS+CMB k-dependent constraints on a general shape, $\sigma(f_{NL}S(k_1, k_2, k_3))$, assuming a basis set of local, equilateral, and orthogonal templates assuming no [left] and perfect [right] prior knowledge of the Gaussian bias. 110

4.1 Example spline basis generated from knots at $x = \{0, 0, 0, 0, 1/3, 2/3, 1, 1, 1, 1\}$ and polynomials of degree $q = 3$. 120
4.2 Comparisons of shape reconstructions for $S_{\rm NBD1}$ (solid) and $S_{\rm NBD2}$ (dashed) [upper left panel] or $S_{\dot{\zeta}^3}$ (solid) and $S_{\zeta\dot{\zeta}^2}$ (dashed) [other three panels] using the polynomial (black), Fourier (gray), and b-spline methods. *Top row and lower left panels:* The 3D spline basis reconstructions correspond to the colored curves, where the purple, blue, orange, and red colors correspond to spline basis sets derived from using 10, 14, 22, and 40 1D spline functions in each dimension. *Lower right panel:* The 2D spline basis reconstructions correspond to the colored curves, where the purple, blue, orange, and red colors correspond to spline basis sets derived from using 50, 100, 200, and 300 1D spline functions in each dimension. 125
4.3 2D b-spline fits to $S_{\dot{\zeta}^3}$ (solid) and $S_{\zeta\dot{\zeta}^2}$ (dashed), where the purple, blue, orange, and red colors correspond to spline basis sets derived from using 50, 100, 200, and 300 1D spline functions in each dimension. 128

A.1 Example 1D b-spline fits to $f(x) = \sin(10x)/(10x)$. All fits shown have used a basis of 13 splines. In the upper panel, the fits have been computed using 6 data points, sampled at $x_i = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Attempting to construct such a fit without smoothing [solid red] causes the splines to produce spurious features, especially at smaller values of x, while introducing a penalty and a small amount of smoothing [dashed red], $\lambda = 0.1$, restores the fit to a reasonable representation of the true function. The lower panel is the same fit, except with 21 data points, sampled at $x_i = \{0, 0.05, 0.1, \ldots, 0.9, 0.95, 1\}$. In this case, the fits with [dashed blue] and without [solid blue] smoothing are very similar. 145

A.2 Cosines between the enfolded shape and the 2D b-spline fits generated by a basis with 10 splines per dimension. 149
A.3 The original template $S_{\mathrm{enf}}(x, y)$ [upper panel] and its b-spline fit using a set of 55 modes [lower panel]. The amplitudes range from 0 (violet) to 1 (red) in both panels. 149

CHAPTER 1

INTRODUCTION

Cosmology over the past several decades has converged on a concordance model that tells a consistent story of our Universe from its early hot Big Bang phase to its present state. The concordance cosmology is one which obeys Einstein's general relativity, is flat, presently made up of approximately 70% dark energy, 25% dark matter, and 5% baryonic matter, and consistently ties together all existing cosmological measurements to date (e.g. [9, 10, 11]). However, despite the ΛCDM cosmology's many successes in evolving initial conditions from the onset of the Big Bang to the present day, in agreement with observations, it alone does not provide the answers about why or how our Universe's initial conditions were set as they were.

The lack of the ΛCDM model's predictive power in this respect is particularly concerning, because the initial conditions inferred from the model's parameters appear to be very fine-tuned. Historically, these fine-tunings have been called puzzles within the Big Bang paradigm. For example, why is the Universe so flat (the 'flatness problem'), and why have no magnetic monopoles been observed (the 'monopole problem')? Why do causally disconnected regions of the sky look so similar (the 'horizon problem')? In addition to these questions, the concordance cosmology does not comment on the origins of the density perturbations that need to have been present during the early universe in order for subsequent structures to form through gravitational collapse.

Beginning in the early 1980's, it was realized that an early phase of accelerated expansion, or 'inflation', preceding the Big Bang would resolve these puzzles (e.g. [12, 13, 14]). A simple model for inflation can be constructed

by positing the existence of a single scalar field, the 'inflaton', with an associated potential, $V(\phi)$, which dominates the energy budget of the universe during inflation. Then, the inflaton's effective equation of state is $w \equiv p/\rho = (\frac{1}{2}\dot{\phi}^2 - V(\phi))/(\frac{1}{2}\dot{\phi}^2 + V(\phi))$, and can drive accelerated expansion if the negative pressure is sufficiently strong ($w < -1/3$). An example of such a scenario is shown in Figure 1.1. If the accelerated expansion lasts a minimum number of e-foldings, then the flatness, monopole, and horizon problems can be neatly reconciled: the exponential expansion drives generic values of the curvature towards flatness, dilutes the density of monopoles, and shrinks the comoving horizon, such that the largest density fluctuations observable today were previously in causal contact (as illustrated in Figure 1.2).

While this classical behavior of the inflaton field provides a description for the background accelerated expansion of the universe during inflation, quantum fluctuations of the inflaton around the classical solution can provide an elegant physical explanation for the origin of density fluctuations. In this scenario, small differences in the value of the inflaton field in separated regions of space result in different volumes of the universe undergoing accelerated expansion for different lengths of time, creating the density fluctuations that provide the initial conditions for the Big Bang and ΛCDM cosmologies. More quantitatively, an inflationary action encodes the dynamics of the background expansion and quantum fluctuations during inflation, and is the starting point for calculations of the primordial power spectrum and bispectrum. For example, the action in the simplest single-field slow-roll scenario is

$$S = \int d^4x \sqrt{-g}\left[-\frac{1}{2}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right], \tag{1.1}$$

where the first term corresponds to the Einstein-Hilbert action for gravity (R is the Ricci scalar), and $g^{\mu\nu}$ is the metric (with g its determinant). This action can

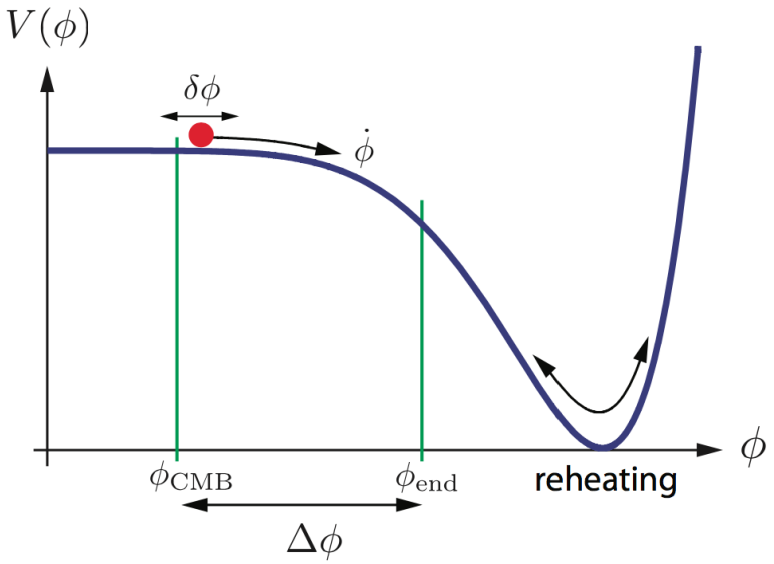


Figure 1.1: A simple inflationary toy model, with a single scalar field, ϕ, called the 'inflaton', and an associated potential, $V(\phi)$. Accelerated expansion occurs when the potential energy dominates the kinetic energy of the field, $\frac{1}{2}\dot{\phi}^2$, which is initially small for a slowly rolling inflaton, and ends when the kinetic energy grows to be comparable to the potential energy, after which the field value will oscillate around the potential's minimum. As the oscillations damp out, the inflaton's energy is transferred into the creation of standard model particles, which heat the universe, and begin the hot Big Bang. Figure taken from [3].

be recast as an expansion in terms of the comoving curvature perturbations, $\mathcal{R}$, which remain constant after leaving the horizon during inflation,

$$S = S_{(0)}[\bar{\phi}, \bar{g}_{\mu\nu}] + S_{(2)}[\mathcal{R}^2] + S_{(3)}[\mathcal{R}^3] + ... \tag{1.2}$$

While the zeroth-order action terms in $S_{(0)}$ determine the background expansion during inflation, such as the values of the Hubble constant and the slow-roll parameters, the second- and third-order terms, $S_{(2)}$ and $S_{(3)}$, determine the generation and evolution of fluctuations captured by the power spectrum and bispectrum, respectively. The quantum mechanical nature of these fluctuations is accounted for by promoting the classical field $\mathcal{R}$ to an operator, schematically

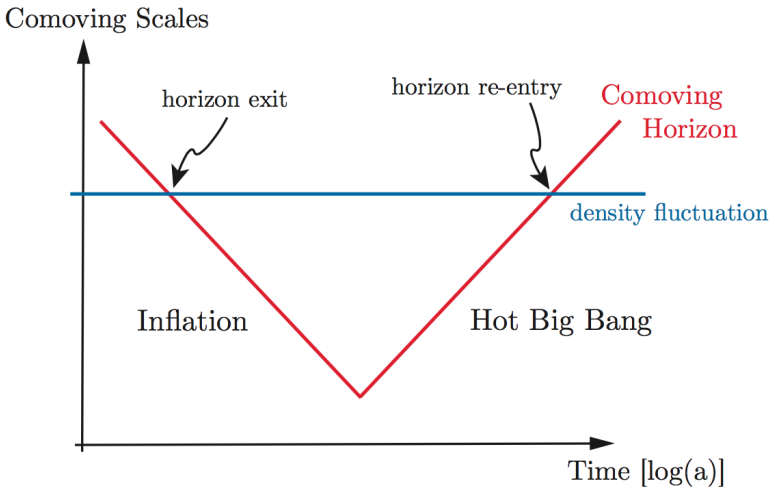


Figure 1.2: This diagram, taken from [3], illustrates the changing comoving horizon (red) with respect to a density fluctuation on a fixed scale (blue). During inflation, the shrinking comoving horizon causes density fluctuations to shift outside the horizon, with the fluctuations on the largest scales leaving first. This effectively freezes in the fluctuation until it re-enters the horizon at a later time, after the onset of the Big Bang. Inflation provides a solution to the horizon problem, by allowing the largest-scale fluctuations observable today to have been inside the horizon before the Big Bang.

written in Fourier-space as

$$\hat{\mathcal{R}}_{\mathbf{k}} \sim \mathcal{R}_{\mathbf{k}}(t)\hat{a}_{\mathbf{k}} + \mathcal{R}^{*}_{\mathbf{k}}(t)\hat{a}^{\dagger}_{-\mathbf{k}}. \tag{1.3}$$

Imposing the canonical commutation relation on the creation and annihilation operators, and choosing the initial quantum state to be the Bunch-Davies vacuum, determines a unique solution for $\hat{\mathcal{R}}_{\mathbf{k}}$. The power spectrum and bispectrum are then defined as correlations in $\hat{\mathcal{R}}_{\mathbf{k}}$ products, evaluated at the time of horizon-crossing,

$$\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') P_{\mathcal{R}}(k) \tag{1.4}$$

$$\langle \mathcal{R}_{\mathbf{k_1}} \mathcal{R}_{\mathbf{k_2}} \mathcal{R}_{\mathbf{k_3}} \rangle = (2\pi)^3 \delta(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}) B_{\mathcal{R}}(k_1, k_2, k_3). \tag{1.5}$$

In this way, different models of inflation, given by a range of inflationary ac-

tions, predict different statistics of the primordial density fluctuations, which we can try to measure in order to test competing inflationary scenarios.

For the simplest single-field slow-roll models, like the one in Figure 1.1, this calculation yields primordial fluctuations that are:

- adiabatic, i.e. fluctuations in different components (photons, baryons, dark matter, neutrinos, etc.) are all sourced by the same perturbations to the gravitational potential, which come from the same inflaton. The violation of this condition would hint at an inflationary model with multiple degrees of freedom, such as a multi-field model of inflation.
- Gaussian, corresponding to the inflaton being a free field in its vacuum ground state. This condition can be broken in a variety of ways, e.g. by introducing additional fields, additional inflaton interactions, or changing the quantum state of the inflaton, each of which will introduce uniquely different non-Gaussian deviations to the resulting primordial fluctuations.
- nearly scale-invariant. Scale-invariant fluctuations are generated when the inflaton's dynamics are time-independent, as it would be if the potential were completely flat. Since inflation must end at some point, there must be some time-dependence, which for a slow-rolling scenario implies that the fluctuations will exhibit small deviations from perfect scale-invariance.
- imprinted on tensor fluctuations of the spacetime metric, which are observable in the late time universe as gravitational waves. For single-field slow-roll models, the amplitude of the tensor power spectrum gives a measure of the energy-scale at which inflation occurred, and the amplitude and near scale-invariance of the tensor power spectrum must satisfy a consistency condition.

These predictions for single-field slow-roll models have been so far found to be consistent with observations of the fluctuations' statistics in the cosmic microwave background (CMB) and large-scale structure (LSS). However, to more fully probe inflation, each of the above predictions for the simplest models of inflation can be constrained with improving data sets, to search for clues to the physical mechanism for inflation. In the last decade, the connections between inflationary theories and the (non-)Gaussianity of the generated primordial fluctuations have been explored for a large number of inflationary models in the literature. While a nearly scale-invariant power spectrum, $P(k)$, is a generic feature of many inflation models, the structure of higher-order correlators, such as the bispectrum, $B(k_1, k_2, k_3)$, or 3-point correlation function, and in particular, their k-dependence, have been shown to be highly sensitive probes of the particulars of the underlying physics of inflation.

One can add physically-motivated complexity to the single-field slow-roll inflaton's action, and see what kinds of bispectra arise. General trends in the derived bispectra of different physical models have become apparent over the last decade of exploring the ties between bispectra and inflationary physics:

- *Local type* non-Gaussianity arises in many models with multiple fields, and these bispectra have large amplitude for squeezed triangles, $k_3 << k_2 \sim k_1$. Squeezed triangles, as well as other commonly discussed triangle configurations, are shown in Figure 1.3.
- *Equilateral and orthogonal type* non-Gaussianity generically come from inflationary actions with non-trivial kinetic terms, or higher-derivative interaction terms, and these bispectra have their largest amplitudes for equilateral triangles, $k_1 \sim k_2 \sim k_3$, and folded/flattened triangles, with $k_2 + k_3 \sim k_1$.

- *Flattened type* non-Gaussianity, also called *enfolded* or *folded* non-Gaussianity, come from models with non-trivial vacuum initial conditions, and these bispectra generically have large amplitude on folded/flattened triangles. Their bispectrum functions are also often highly oscillatory.

These studies, taken together, imply that it is possible to shed light on the physics of the inflation by focusing our efforts on measuring the non-Gaussianities of the primordial fluctuations with future experiments.

Currently running and upcoming surveys of the CMB and LSS aim to do this by constraining the bispectrum of primordial fluctuations with increasingly higher precision. The strongest constraints on non-Gaussianity to date come from the Planck satellite's survey of the CMB sky [15], while galaxy surveys that are planned for the next decade, such as the Dark Energy Survey Instrument (DESI) and Euclid, will provide an alternative and complementary probe of non-Gaussianity. Any robust detections of non-Gaussianity, or the continued non-detections of existing constraints, both have exciting implications for our understanding of inflation, and motivates continued efforts to better understand and analyze the available and upcoming data.

Constraining higher order correlators, like the bispectrum, however, is generally not an easy task, and a few subtleties of obtaining and interpreting constraints should be kept in mind. First, calculating observational CMB and LSS predictions for a variety of non-Gaussian initial conditions is computationally challenging, and constructing pipelines that can translate realistic data sets into robust constraints on the amplitude and k-dependence of the true primordial bispectrum adds to the difficulty. Much of the computational burdens can be lifted in cases where the bispectrum of interest can be written, or well-

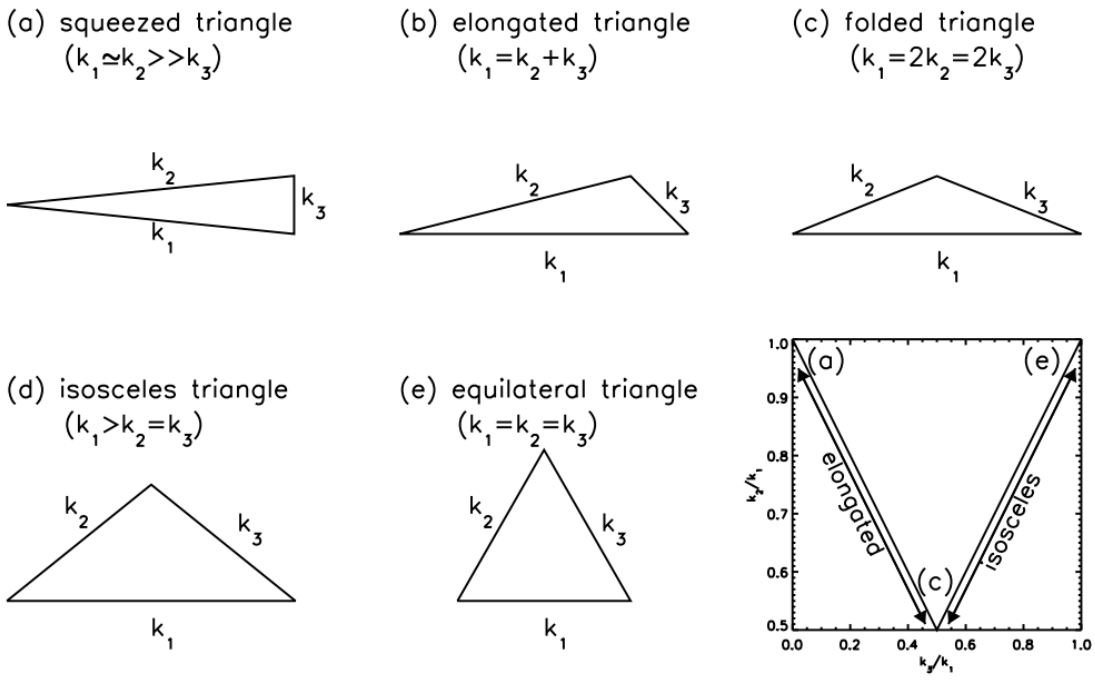


Figure 1.3: Parts (a)-(e) of this figure from [4] illustrate a variety of triangle configurations that make up the full space of allowed (k_1, k_2, k_3) combinations, representing the space over which a bispectrum's amplitude, $B_\Phi(k_1, k_2, k_3)$, is defined. Each one is labeled with a common name, such as 'squeezed triangle', that is often used in the literature and throughout this work to correspond to similar-looking triangles. Additionally, bispectra are often plotted on figures like the one in the bottom right panel. Bispectra are always symmetric under interchange of the wavenumbers, and many bispectra are scale-invariant, meaning they can be reduced to functions of k_2/k_1 and k_3/k_1, allowing all of the unique information contained within a bispectrum function to be shown as a contour plot over the domain shown in the last panel.

approximated as, a separable function of the three wavenumbers, $B_\Phi(k_1, k_2, k_3) = X(k_1)Y(k_2)Z(k_3)$ + permutations of k_1, k_2, and k_3. Exact separable approximations are not guaranteed to exist, but a handful of shape classes (like the local, equilateral, orthogonal, and flattened types discussed above) can be defined, with a representative separable template for each one. However, the danger of this

approach is that there are an infinite number of ways for non-Gaussianities to be realized in the data, and if the true form of non-Gaussianity, if it exists, is not similar to a type that we are looking for, there is a possibility that it will go undetected. It is therefore worthwhile to look for ways of constraining general types of bispectra, including exotic types that do not fit into the classes discussed, and general forms that do not directly correspond to any particular physically-derived model. Fergusson et al. have proposed a strategy of constructing a basis of separable functions, upon which any bispectrum can be reconstructed [16]. In practice, the efficient convergence of such an expansion is sensitive to the bispectra that one desires to the reconstruct, as well as the particular choice of separable basis. Alternative choices of basis can be explored, to probe as much of the theory space as possible, and give us the best chances of uncovering the correct model.

Secondly, in the scenario where some form of non-Gaussianity is eventually detected, care must be taken when interpreting the detection. For example, if we search for a local type of non-Gaussianity and detect it with a statistically significant amplitude, how much confidence can we have that the correct model is truly the local type, and verify that it is not the signal from another type of non-Gaussianity that is leaking into our measurement? There remains a possibility that we are detecting signatures of an altogether different theory, the bispectrum for which shows some similarities with the local template used, but ultimately describes different physics. The degradation of the measurements' error bars can be quantified to include the effect of having to distinguish between multiple competing models on the resulting constraints.

This thesis represents a small part of the current efforts to constrain inflation

through searches for non-Gaussianity, and presents calculations of forecasted constraints on and improving tests of, inflationary models that will be possible with these current and future data sets. In Chapter 2, reproduced from [17], we construct an alternative separable basis, which we call the "divergent basis", that has the advantage of being able to more efficiently describe a wider variety of shapes that peak strongly in squeezed configurations, and using this basis we forecast constraints on general shapes that will be achieved with a CMB survey of temperature and polarization fluctuations, modeled on the data set expected from Planck's final release. Chapter 3 reproduces the analysis in [18], which computes the constraints on inflationary models that upcoming galaxy surveys are likely to yield. The same density perturbations that are reflected in the CMB sky are imprinted on the distribution of structures, such as galaxies, in the late-time universe. Rather than attempting to measure the primordial bispectrum through the bispectrum of galaxies, halos, or dark matter, we instead search for primordial non-Gaussianity indirectly through its effect on the scale-dependence of the halo bias, which is measurable in the power spectrum of galaxies. We forecast the constraints on general shapes that the power spectrum of galaxies will provide, from future galaxy surveys such as DESI and Euclid, focusing on whether these constraints can compete with or complement the constraints from Planck on general shapes, and what theoretical systematics they might be susceptible to. The forecasted constraints in both Chapters 2 and 3 are computed jointly for a subset of non-Gaussian bispectra, to quantify the impact of distinguishing between multiple models on resulting constraints.

In Chapter 4, reproduced from [19],we consider a special class of inflationary models that have typically been less well-probed by existing analyses, due to the unusual features that their bispectra typically display. Inflationary models

with non-trivial initial quantum states for the inflaton are physically motivated extensions of the simplest single-field models, and these bispectra are generically peaked in squeezed and flattened triangular configurations, with oscillatory amplitudes over the entire space of triangles. Their sharp features and rapid oscillations have made them difficult to reconstruct using existing choices of separable basis, such as the polynomial basis adopted by Planck's existing constraints on general shapes. In this work, we derive general shapes with non-Bunch-Davies initial states, and devise a basis built upon b-spline functions that are computationally simple and are able to compete with existing choices of basis for reconstructing these exotic and unusual shapes. Finally, in Chapter 5, we summarize the major results of the work in this thesis, and discuss future directions and extensions of interest on the road to improving constraints on inflation.

CHAPTER 2

NON-GAUSSIAN SHAPE RECOGNITION

2.1 Introduction

An early period of accelerated expansion, perhaps a trillionth of a second after the Big Bang, is proposed to solve a number of problems unresolved by the Big Bang scenario, such as the flatness and horizon problems. The paradigm of single-field slow-roll inflation is the simplest model to describe this acceleration, and makes broad predictions of adiabatic and Gaussian-distributed primordial density (scalar) perturbations, described by a nearly scale-invariant 2-point correlation, and smaller gravitational metric (tensor) perturbations. In this model, the scalar and tensor amplitudes and scale dependence are related through a 'consistency relationship'.

The precision of astrophysical measurements has dramatically improved over the last decade. Cosmic Microwave Background (CMB) measurements, such as from the Wilkinson Microwave Anisotropy Probe [20, 21], small-scale CMB measurements from the Atacama Cosmology Telescope and South Pole Telescope [22, 23], and large-scale structure (LSS) observations such as from the Sloan Digital Sky Survey [24, 25], are all entirely consistent with single-field slow-roll predictions, placing strong constraints on the scalar power spectrum and upper limits on the degree of deviations from Gaussianity and amplitude of tensor modes.

The agreement between single-field inflationary predictions and observations is a profound success for cosmology, but it is as yet, insufficient to inform

us about the underlying theory from which inflation derives. Rapid theoretical progress in high energy effective field theory has led to a wide range of possible Lagrangians for inflation [26, 27, 28, 29]. These often go beyond making distinct predictions for the form of the single-field inflationary potential and can include multiple dynamical fields and non-canonical derivative (kinetic) terms in the action.

These alternative mechanisms can generate new observational signatures, including different consistency relationships relating scalar and tensor perturbations [30, 31], the addition of non-adiabatic (isocurvature) modes [32, 33], and the possibility of observationally measurable non-Gaussian correlations [34]. In particular, the possibility that primordial non-Gaussianity may be detectable as a non-zero 3-point correlation function, or bispectrum, has been a major development in the search for observational signatures of the underlying inflationary theory. What is most exciting is that different theories can give rise to bispectra with distinct scale dependencies, such that measuring not only the amplitude but also the scale-dependence, or 'shape', of the bispectrum could provide a direct insight into the inflationary mechanism.

Much work has focused on potentially measuring the amplitude of commonly predicted shapes, such as the local [35, 36, 34], equilateral [37], and orthogonal [38] templates. Recent theoretical developments have also led to a wider population of bispectra, including those from fast-roll inflation [39, 40, 41, 42, 43], quasi-single field inflation [44, 45], warm inflation [46, 47, 48], and non-Bunch-Davies or excited initial states [39, 49, 50, 51]. There are also hybrids of multi-field and non-slow-roll models [52, 53, 54], and the inclusion of isocurvature modes in the non-Gaussian correlations [55, 56, 57]. These bispec-

tra can have very different shapes, meaning their signal is weighted towards different configurations of the 3 wavenumbers in (Fourier) k-space. How divergent shapes are in the 'squeezed' k-configuration, when one of the three length scales contributing to the 3-point function becomes much larger than the other two, in particular can signal whether inflation is derived from a single-field or multi-field model.

The divergence in the squeezed limit could also be constrained by its effect on large-scale structure. A non-Gaussian signal peaking in the squeezed limit would directly couple large scale modes to small scales, on which non-linear halos are forming [58]. This gives rise to an additional contribution to the halo bias, determining how the number density of halos of a given mass are related to the underlying linear power spectrum. In theory, wide field large-scale structure surveys could provide a sensitive constraint on the divergence properties of non-Gaussianity [59, 60, 61, 62, 63, 64, 65].

Given the diversity of theoretically motivated shapes, an intriguing question is how well one might actually be able to determine the shape of primordial non-Gaussianity, rather than purely assuming a shape template is the true shape a priori. To what extent can the shape of non-Gaussianity be reconstructed using the CMB and LSS 3-point correlations? If a positive detection is made assuming a template, how well would such a detection really constrain the underlying shape and the theoretical model that generated it?

This 'reconstruction' approach has been widely considered in the context of the inflationary power spectrum, both in terms of $P(k)$ reconstruction (e.g. [66, 67, 68]), and measuring the hierarchy of slow-roll parameters (e.g. [69, 70, 71, 72] and references therein), instead of assuming a nearly scale-invariant spectrum

parametrized by a constant tilt n_s and a constant running $dn_s/d\ln k$.

Unfortunately, calculating theoretical predictions for CMB bispectra is computationally cumbersome in its exact form, requiring 4-dimensional integrals to be performed. A formalism to make the calculation tractable for general bispectra was introduced in [16]. The authors proposed a technique to create templates for shapes by expanding non-separable shapes on a basis set of bispectra that are explicitly separable functions of the three wavenumbers. The separability reduces the 4-dimensional integral to a tractable computation without a significant reduction in the accuracy of the computed CMB bispectra. This approach has been used to forecast bispectrum constraints for a variety of fundamental shapes [21] and adapted to other basis sets to describe oscillatory, rather than monotonic, shapes [73]. Furthermore, the method of modal expansions on a separable basis has been shown to be advantageous and applicable in a variety of contexts, for example in studying CMB 3-point correlations with wavelets [74], CMB trispectra [75, 76], and matter density bispectra in LSS [77, 78, 79].

In this work we present an alternative separable basis to efficiently describe and investigate the broad class of nearly scale-invariant general bispectra in terms of their squeezed limit properties. We discuss a way to expand a general shape in the basis, which is specifically tuned to enable us to systematically increase the complexity of the template in a theoretically motivated way. We forecast the potential for determining the underlying non-Gaussian shape given upcoming CMB temperature and E-mode polarization data modeled on the Planck survey.

The format of the paper is as follows. In section 2.2, we review the formalism used to calculate CMB bispectra. We introduce a separable basis to describe

general shapes that are scale-invariant and potentially divergent, and discuss how this basis can be applied to describe a wide variety of shapes in the literature. Using the basis, we develop an expansion that allows us to incrementally investigate classes of bispectra motivated by theories. In section 2.3, we present a Fisher analysis quantifying how well a Planck-like survey will be able to distinguish between and constrain individual bispectrum shapes. Using a principal component analysis, we find the best to worst measured uncorrelated shapes, and compute the overall uncertainties in the bispectrum measurement as a function of k-space configuration under different theoretical priors. We use these results to establish how much we can learn about the bispectrum shape, and hence with what confidence we might be able to narrow down the underlying inflationary theory. In section 2.4, we summarize our findings and discuss implications for future work.

2.2 Efficient calculation of a general non-Gaussian shape

In this section we lay out the formalism to describe and compute general bispectra. In subsections 2.2.1 and 2.2.2, we respectively review the calculation of the CMB bispectrum given the primordial 3-point function and the definitions of covariances in wavenumber and multipole space that roughly quantify the theoretical similarity of two bispectra. In subsection 2.2.3 we introduce a separable basis set to describe general bispectra and develop computationally tractable templates. Subsection 2.2.4 discusses the application of the basis set to a variety of theoretical bispectra and templates in the literature. How bispectra can be classified and presented pictorially is reviewed in subsection 2.2.5.

2.2.1 The CMB bispectrum

While Gaussian fluctuations are wholly described by a 2-point correlation function, a full description of non-Gaussian fluctuations requires higher order correlations that are not trivially related to the 2-point function. The simplest higher order correlation is the 3-point function, where the 3-point Fourier space statistic analogous to the 2-point power spectrum is the bispectrum, B_Φ, defined by

$$\langle \Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3)\rangle \equiv (2\pi)^3\delta(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})B_\Phi(k_1,k_2,k_3). \tag{2.1}$$

$\Phi(\mathbf{k})$ is the primordial gravitational potential, related to the curvature perturbation by $\Phi = \frac{3}{5}\mathcal{R}$. Under the assumptions of statistical isotropy and homogeneity, the bispectrum is dependent on only the magnitudes of the wavenumbers, k_1, k_2, and k_3.

The bispectrum is often parameterized by a shape, $S(k_1,k_2,k_3)$, and an amplitude, f_{NL}, at an arbitrary configuration in k-space which together determine the bispectrum at all scales,

$$\frac{(k_1k_2k_3)^2}{N}B_\Phi(k_1,k_2,k_3) \quad = \quad f_{NL}S(k_1,k_2,k_3). \tag{2.2}$$

The typical convention is to choose $N = 6[2\pi^2\left(\frac{3}{5}\right)^2\Delta^2_{\mathcal{R}}(k_0)]^2$, where $\Delta^2_{\mathcal{R}}(k_0)$ is the amplitude of the primordial power spectrum of the curvature perturbations at a pivot scale k_0. Shapes are typically normalized such that $S(k_0,k_0,k_0) = 1$.

CMB statistics are commonly described by correlations between angular moments on the sky, $a_{\ell m}$, calculated through a spherical harmonic decomposition of the photon transfer functions, $\Delta_\ell(k)$, integrated along the line of sight and sourced by the primordial perturbations,

$$a_{\ell m} \quad \equiv \quad 4\pi(-i)^\ell\int\frac{d^3k}{(2\pi)^3}\Delta_\ell(k)\Phi(\mathbf{k})Y_{\ell m}(\hat{\mathbf{k}}). \tag{2.3}$$

The CMB 3-point correlation function is given by

$$\begin{aligned}\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}\rangle &= \left(\frac{2}{\pi}\right)^3 \int dx x^2 \int dk_1 dk_2 dk_3 (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3) \\ &\times \Delta_{\ell_1}(k_1)\Delta_{\ell_2}(k_2)\Delta_{\ell_3}(k_3) j_{\ell_1}(k_1 x) j_{\ell_2}(k_2 x) j_{\ell_3}(k_3 x) \\ &\times \int d\Omega_{\hat{\mathbf{x}}} Y_{\ell_1 m_1}(\hat{\mathbf{x}}) Y_{\ell_2 m_2}(\hat{\mathbf{x}}) Y_{\ell_3 m_3}(\hat{\mathbf{x}}). \end{aligned} \tag{2.4}$$

To perform the integrals over k and x, we use the CAMB[1] code [80], which uses the line of sight approximation [81] to calculate the photon transfer functions Δ_ℓ.

We consider purely isotropic bispectra for which the integral over $\int d\Omega_{\hat{\mathbf{x}}}$ is a separable geometrical factor called the Gaunt integral. The properties of the Gaunt integral require that the non-zero correlations have ℓ_1, ℓ_2, and ℓ_3 satisfying an even sum $\ell_1 + \ell_2 + \ell_3$ and $|\ell_1 - \ell_2| \leq \ell_3 \leq \ell_1 + \ell_2$ for ℓ_1, $\ell_2 \leq \ell_3$. Under the assumption of isotropy, the angle-averaged angular bispectrum is the 3-point analogue to the C_ℓ,

$$B_{\ell_1 \ell_2 \ell_3} = \sum_{m_i} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}\rangle, \tag{2.5}$$

where the bracketed term is the Wigner-3j symbol. To further separate out a purely geometrical factor from the angular-averaged bispectrum, it is convenient to work with the reduced bispectrum $b_{\ell_1 \ell_2 \ell_3}$,

$$B_{\ell_1 \ell_2 \ell_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1 \ell_2 \ell_3}, \tag{2.6}$$

so that

$$\begin{aligned} b_{\ell_1 \ell_2 \ell_3} &= \left(\frac{2}{\pi}\right)^3 \int dx x^2 \int dk_1 dk_2 dk_3 (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3) \\ &\times \Delta_{\ell_1}(k_1)\Delta_{\ell_2}(k_2)\Delta_{\ell_3}(k_3) j_{\ell_1}(k_1 x) j_{\ell_2}(k_2 x) j_{\ell_3}(k_3 x). \end{aligned} \tag{2.7}$$

[1] http://camb.info

Here x is a dummy variable which should be integrated between zero and infinity. We note that in an analogous evaluation of C_ℓ, x has a physical interpretation as the comoving distance to the surface of last scattering. One might assume therefore that in the 3-point integral the upper limit x_{max} could be set to $(\tau_0 - \tau_{rec})$. However, as others have previously also commented [38], the integral over x arises out of rewriting the delta function in 2.1 as an integral over a product of Bessel functions. Numerically, for a general bispectrum we find the value of x_{max} ensuring the required degree of convergence is ℓ-dependent, and typically needs to be greater than $(\tau_0 - \tau_{rec})$.

2.2.2 Shape similarity

The degree to which a bispectrum B is theoretically similar to another, B', can be quantified by a k-space correlation coefficient, or 'k-space cosine', corr_k, integrated over the k-space tetrapyd volume $\mathcal{V}$ with weight w [82, 16],

$$\langle S, S' \rangle \equiv \int_{\mathcal{V}} S(k_1, k_2, k_3) S'(k_1, k_2, k_3) w(k_1, k_2, k_3) dk_1 dk_2 dk_3, \tag{2.8}$$

$$\mathrm{corr_k}(S, S') \equiv \frac{\langle S, S' \rangle_k}{\sqrt{\langle S', S' \rangle_k \langle S, S \rangle_k}}. \tag{2.9}$$

An analogous statistic describing the similarity of two bispectra in multipole space can be quantified by their ℓ-space correlation coefficient, or 'ℓ-space cosine', corr_ℓ [16],

$$\langle S, S' \rangle_\ell \equiv \sum_{\ell_1 \ell_2 \ell_3} \frac{B_{\ell_1 \ell_2 \ell_3} B'_{\ell_1 \ell_2 \ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \tag{2.10}$$

$$\mathrm{corr}_\ell(B, B') \equiv \frac{\langle S, S' \rangle_\ell}{\sqrt{\langle S, S \rangle_\ell \langle S', S' \rangle_\ell}}. \tag{2.11}$$

These correlation statistics are frequently used within non-Gaussian shape studies to quantify how well a template matches a given shape. A template can

be obtained, given a set of n basis shapes, in our case $\{\mathcal{K}_n\}$, by first applying the Gram-Schmidt algorithm to give a set of orthonormal basis functions $\{\mathcal{R}_n\}$ (in either k or ℓ space),

$$\mathcal{R}'_0 = \mathcal{K}_0 \tag{2.12}$$

$$\mathcal{R}'_{n+1} = \mathcal{K}_{n+1} - \sum_{i=0}^{n} \langle \mathcal{K}_{n+1}, \mathcal{R}_i \rangle \tag{2.13}$$

$$\mathcal{R}_j = \frac{\mathcal{R}'_j}{\sqrt{\langle \mathcal{R}'_j, \mathcal{R}'_j \rangle}} = \sum_{i=0}^{j} \lambda_{ji} \mathcal{K}_i, \tag{2.14}$$

where the last line defines the matrix λ. The new basis can then be used to create a matched template for a specific non-separable shape S,

$$S^{(n)}_{template} = \sum_{i=0}^{n} \langle \mathcal{R}_i, S \rangle \mathcal{R}_i = \sum_{i=0}^{n} \alpha_i \mathcal{R}_i. \tag{2.15}$$

We note that in general the classical Gram-Schmidt algorithm can be numerically unstable, resulting in $\{\mathcal{R}_n\}$ that are not exactly orthogonal. This issue can be abated by implementing the well-known modified Gram-Schmidt algorithm, and any numerical issues that may remain can be checked by verifying that all of the $\{\mathcal{R}_n\}$ are orthogonal to each other, i.e. $\langle \mathcal{R}_i, \mathcal{R}_j \rangle = \delta_{ij}$. For our case this was true to within a few $\times 10^{-6}$ at worst, across the first 7 modes. In addition, a faster check can be conducted for those shapes for which the coefficients on the original basis are known, since computing $\alpha\lambda$ should return the input coefficients on the $\{\mathcal{K}_n\}$ basis. We verified that this was the case for the equilateral, orthogonal, and enfolded templates, with the worst coefficients being off by a fractional error of a few $\times 10^{-5}$. We find this accuracy is more than sufficient for the forecasting analyses to constrain shape measurements with upcoming surveys as we discuss in section 3.

The efficacy of the template to match a given shape S can be quantified by the cumulative cosine, $\mathrm{corr}(S, S^{(n)}_{template})$ as in (2.9) or (2.11). A high correlation

coefficient signals a good fit. If this cosine is close to one then the two shapes are sufficiently alike that one might expect constraints on the amplitude of B can be taken as constraints on B' as well, without having to do a separate analysis of the data. If on the other hand, the cosine is low, then it is likely that separate analyses of the data are needed for B and B', because a template for B will not be able to pick out a non-zero signal for B', and vice versa. We note though that while a template may have a large cosine with the shape, this does not automatically mean that the template will be able to accurately model a correlation between the true shape and a third shape, with which it is not similar. The extreme example of this would be if the third shape were exactly proportional to the discrepancy between the template and shape. In constructing a template, if this was a concern, one might want to tailor it to the purpose by altering the weight in the Gram-Schmidt decomposition to ensure a minimization of the covariance between the shape and template over a given region of (k or ℓ) configuration space in which a third shape was relevant.

While it is extremely useful to establish the similarity of shapes, it is the converse of this, how well two shapes can be distinguished from one another, using data, that is the main focus of this work. This provides a motivation to consider an efficient way to generate ℓ-space bispectra explicitly by creating templates described by basis functions separable in k_1, k_2 and k_3 as we discuss below. To do this, in sections 2.2.3 and 2.2.4 we develop a framework to describe the possible degrees of freedom that a general shape might have under a specific theoretical prior.

We note that corr_k and corr_ℓ represent simplified correlation statistics that purely take into account the cosmic variance limitations. Neither statistic, as

they are written above, takes into account the noise, sky coverage or resolution characteristics of a particular survey. As described in [16], corr_ℓ can be modified to include these experimental effects by changing the weighted sum over ℓ_1, ℓ_2, ℓ_3 to reflect the measurement covariance matrix. The modified corr_ℓ is then a refined, survey-dependent extension of 2.11 that tailors the correlation statistic to reflect the observational, rather than intrinsic, distinguishability of shapes. Distinguishability between several shapes can be done by conducting a Fisher or χ^2 analysis that includes the measurement covariance matrix [34]. In section 2.3, we perform a Fisher analysis and use the correlation statistics, including experimental effects such as instrument noise, beam size, and incomplete sky coverage, to quantify how well upcoming surveys might distinguish one shape from another.

2.2.3 A new separable basis, $\mathcal{K}_n$

The 4-dimensional integral over the product of highly oscillatory functions given in 2.7 is computationally intensive. This has been a barrier to efficiently calculating observational predictions for the CMB bispectrum. As a result many studies have focused on models for which the primordial bispectrum can be written as, or well-approximated by, a separable (symmetric) function of (k_1, k_2, k_3),

$$S(k_1, k_2, k_3) = f(k_1)g(k_2)h(k_3) + \text{ cyclic perms}, \tag{2.16}$$

such that the 3-dimensional integral over k_1, k_2, and k_3 in 2.7 is reduced to a product of three 1-dimensional integrals.

In [16] the authors proposed a way to reduce the computation time for gen-

eral models by expanding the shape in terms of a separable basis,

$$S(k_1, k_2, k_3) \quad = \quad \sum_i \alpha_i Q_i(k_1, k_2, k_3). \tag{2.17}$$

Each Q_i is constructed from symmetrized products of three 1-dimensional polynomials of k_1, k_2, and k_3,

$$Q_n(k_1, k_2, k_3) \propto \sum_{i=0}^{p} c_{pi} \sum_{j=0}^{r} c_{rj} \sum_{k=0}^{s} c_{sk} \left(k_1^i k_2^j k_3^k + k_1^i k_2^k k_3^j + 4 \text{ perms} \right), \tag{2.18}$$

where n maps onto a combination of $\{p, r, s\} \geq 0$, and c_{pi}, c_{rj}, c_{sk} are constants. Using the Q_n basis, the equilateral template can be reconstructed to 98% accuracy (according to the cumulative cosine) using 6 basis functions [16], while other shapes motivated by single-field inflation models can require 20 or more mode functions to get > 95% convergence [83].

An analysis of data to constrain the bispectrum depends not only on the uncertainties inherent in the data itself, but also the theoretical priors determining the model being compared with the data. The choice of a separable basis set to describe the theory is therefore also influenced by this prior. An analysis allowing the primordial bispectrum to take any form (i.e. with no shape prior on the forms of the separable functions f_i, g_i, h_i) would use a discrete set of k-space bins to describe the uncorrelated amplitudes at each scale and configuration. In such a scenario, no theoretical prior is applied and the constraints on the bispectrum are simply those determined by the data. In studying theoretically motivated models of inflation, however, there can be broad or specific characteristics of the bispectra that it would be reasonable to impose in conjunction with the data that suggest a form for the separable basis functions. For example, a Fourier basis may be more efficient than a polynomial one for describing bispectra with oscillatory features [73]. The two minimal assumptions we consider here as theoretical priors are that the bispectrum i) has a roughly monotonically changing

amplitude as a function of scale, and ii) like the power spectrum, it is nearly scale-invariant.

The polynomial basis of [16] does not naturally confine shapes to these two common theoretical properties of bispectra. Firstly, the polynomial basis does not naturally restrict itself to scale-invariant shapes, because $i + j + k \geq 0$ in 2.18; resulting sums of the basis functions are thus scale-dependent in general. Most theoretically-motivated bispectra in the literature, however, are nearly scale-invariant, with $i + j + k \approx 0$ (see [84] for a review). Such shapes can be reduced to functions of two variables, k_1/k_3 and k_2/k_3. There are exceptions to this, of course, such as non-Gaussianity from particle production [85] or from features in the inflationary potential [86, 87]. These models can strongly deviate from scale-invariance because modes leaving the horizon at a specific moment, when particle production is occurring or a feature in the potential is important, are preferentially populated. Secondly, different types of theoretical mechanisms generating bispectra predict different divergence properties in the squeezed limit, where $k_1 \ll k_2 \approx k_3$. We consider the squeezed limit as $k_1 = k_\ell$, the long wavelength mode, and $k_2 = k_3 = k_s$, the short wavelength modes, so that for scale-invariant shapes the squeezed limits purely dependent on $x_{sq} \equiv k_\ell/k_s$. Single-field inflation models, through a consistency relation [88] predict the bispectrum will vanish in this limit. Local bispectra, typically arising in multi-field models, have a x_{sq}^{-1} divergence, while excited states can have x_{sq}^{-2} divergence. Since the powers i, j, and k in 2.18 are ≥ 0, the $\{Q_n\}$ all tend towards a constant value in the squeezed limit, and thus cannot effectively describe shapes diverging in the squeezed limit. As a result, this basis is not immediately suited to reconstructing templates that display specific divergence behaviors in this limit, without a further prior being imposed. For example, the compelling and

well-studied local template cannot easily be recovered using the $\{Q_n\}$ basis without either using many more basis functions or ignoring the divergent part of the shape that makes the local template distinct from others [83].

In this paper we introduce a set of separable basis functions, $\{\mathcal{K}_n\}$, that efficiently describe nearly scale-invariant and potentially divergent shapes, and explicitly consider the forms of the shapes generated using this basis under various divergence constraints. Explicitly, we consider

$$f_{NL}S(k_1,k_2,k_3) = \sum_n f_{NL}^n \mathcal{K}_n(k_1,k_2,k_3). \tag{2.19}$$

Here f_{NL}^n are expansion coefficients and n again denotes a combination of powers $\{p,r,s\}$ of the wavenumbers (k_1,k_2,k_3). $\mathcal{K}_n$ is defined as

$$\mathcal{K}_n(k_1,k_2,k_3) \equiv \frac{1}{\mathcal{N}_n k_0^{2(n_s-1)}}\left[k_1^{p'}k_2^{r'}k_3^{s'} + \{prs\}\text{perms}\right], \tag{2.20}$$

where $\mathcal{N}_n$ is the number of distinct permutations of $\{p,r,s\}$. p' is defined as

$$p' = 2 + \frac{(p-2)(4-n_s)}{3}, \tag{2.21}$$

and similarly for r' and s'.

To allow for potentially divergent shapes, we allow the powers $\{p,r,s\}$ to be negative as well as positive, and to make each $\mathcal{K}_n$ nearly scale-invariant, we require the powers to satisfy $p+r+s=0$. Equations 2.20–2.21 then ensure that each $\mathcal{K}_n$ is normalized in the conventional way, with $\mathcal{K}_n(k_1,k_2,k_3)=1$ at $k_1=k_2=k_3=k_0$. In the scale-invariant case where $n_s=1$, $\mathcal{K}_n$ only depends on k_1/k_3 and k_2/k_3, and $\mathcal{K}_n(k_1,k_2,k_3)=1$ for all $k_1=k_2=k_3$.

Each shape has a well-determined behavior in the squeezed limit,

$$\mathcal{K}_n^{sq} = \sum_{m=R}^{1} d_{nm}x_{sq}^m + O(x_{sq}^2). \tag{2.22}$$

$R = min(p,r,s)$	$\mathcal{K}_n$	$\{p,r,s\}$	$d_{nm}, m =$				K_{pqr}
			-3	-2	-1	0	
0	$\mathcal{K}_0$	(0,0,0)				1	K_{000}
-1	$\mathcal{K}_1$	(-1,0,1)			$\frac{1}{3}$	$\frac{1}{3}$	$\frac{K_{12}}{K_{111}}$
	$\mathcal{K}_2$	(-1,-1,2)			$\frac{2}{3}$		$\frac{K_3}{K_{111}}$
-2	$\mathcal{K}_3$	(-2,1,1)		$\frac{1}{3}$			$\frac{K_{33}}{K_{222}}$
	$\mathcal{K}_4$	(-2,0,2)		$\frac{1}{3}$		$\frac{1}{3}$	$\frac{K_{24}}{K_{222}}$
	$\mathcal{K}_5$	(-2,-1,3)		$\frac{1}{3}$	$\frac{1}{3}$		$\frac{K_{15}}{K_{222}}$
	$\mathcal{K}_6$	(-2,-2,4)		$\frac{2}{3}$			$\frac{K_6}{K_{222}}$
-3	$\mathcal{K}_7$	(-3,1,2)	$\frac{1}{3}$				$\frac{K_{45}}{K_{333}}$
	$\mathcal{K}_8$	(-3,0,3)	$\frac{1}{3}$			$\frac{1}{3}$	$\frac{K_{36}}{K_{333}}$
	$\mathcal{K}_9$	(-3,-1,4)	$\frac{1}{3}$		$\frac{1}{3}$		$\frac{K_{27}}{K_{333}}$
	$\mathcal{K}_{10}$	(-3,-2,5)	$\frac{1}{3}$	$\frac{1}{3}$			$\frac{K_{18}}{K_{333}}$
	$\mathcal{K}_{11}$	(-3,-3,6)	$\frac{2}{3}$				$\frac{K_9}{K_{333}}$
	etc.						

Table 2.1: The set of $\mathcal{K}_n$ with the allowed combinations of $\{p,r,s\}$ that satisfy $p+r+s=0$. The ordering of the modes is according to their divergence behavior in the squeezed limit. The coefficients of their divergent terms in this limit, d_{nm}, are trivially related to the values of $\{p,r,s\}$, but are given here for convenience. We also give the equivalent short-hand notation used in [1] and summarized in (2.23).

The set of $\{\mathcal{K}_n\}$ with the allowed combinations of $\{p,r,s\}$ is given in Table 2.1 along with their divergence properties. The basis modes are also written in the equivalent short-hand notation used by [1]. Since we will use this notation to describe non-separable shapes in the next section we summarize it here:

$$K_p = \sum_{i=1}^{3} k_i^p, \quad K_{pq} = \frac{1}{\Delta_{pq}} \sum_{i=1}^{3} k_i^p \sum_{j\neq i} k_j^q, \quad K_{pqr} = \frac{1}{\Delta_{pqr}} \sum_{i=1}^{3} k_i^p \sum_{j\neq i} k_j^q \sum_{\ell\neq i,j} k_\ell^r, \tag{2.23}$$

with

$$\Delta_{pq} = 1+\delta_{pq}, \qquad \Delta_{pqr} = \Delta_{pq}(\Delta_{qr}+\delta_{pr}). \tag{2.24}$$

2.2.4 Application of the basis to shapes arising in inflationary theory

In this subsection, we illustrate the efficiency and accuracy allowed by our basis in describing shapes in the literature. First we discuss cases involving shapes and templates which are exactly expressed in terms of, or well-approximated by, linear combinations of the first 3 modes of the basis. Then we extend the basis to include more divergent modes, and present the basis of shapes we will use under different divergence priors.

Shapes exactly expressed in terms of $\{\mathcal{K}_0 - \mathcal{K}_2\}$

Some commonly considered templates can be exactly expressed in terms of the first three modes of the basis $\{\mathcal{K}_0 - \mathcal{K}_2\}$.

The local shape, $S_{local} = \mathcal{K}_2$, can be derived from a simple ansatz for describing the nonlinear contribution to the primordial curvature perturbation in real space as a local effect [34],

$$\Phi(\mathbf{x}) = \Phi_L(\mathbf{x}) + f_{NL}\left(\Phi_L^2(\mathbf{x}) - \left\langle \Phi_L^2(\mathbf{x}) \right\rangle\right). \tag{2.25}$$

Local shapes arise out of single-field slow-roll models, though the amplitude of the bispectrum in this case is predicted to be undetectably small [88, 89]. Large, local non-Gaussianity is predicted by a wide variety of other physically-motivated models, such as multifield inflation (e.g. curvaton scenario) [90, 91], (p)reheating mechanisms [92], and ekpyrotic inflation [91, 93].

The constant shape, $S_{const} = 1 = \mathcal{K}_0$, was originally studied for its very simple form [1]. More recently the shape has been studied in the context of shapes

arising from quasi-single field inflation (QSFI) models [44, 45, 94, 63]. The more general shape of QSFI is discussed in more detail below.

Models with higher-derivative kinetic terms and/or non-trivial speeds of sound in the inflationary Lagrangian generally produce non-separable shapes, sensitive to the sum $k_t = k_1 + k_2 + k_3$ in the denominator, and thus cannot be exactly written in terms of a separable basis. The equilateral template [37], $S_{equil} = -2\mathcal{K}_0 + 6\mathcal{K}_1 - 3\mathcal{K}_2$, is widely used as a template to detect evidence of such shapes. Examples include generalized single-field models [95, 96, 39, 26], k-inflation [97, 30, 98], ghost inflation [99], DBI inflation [100, 101], single-field non-slow roll and bimetric theories [42, 102, 41].

A general, effective field theory of inflation is dominated by contributions from two shapes [39],

$$S_{DBI} = \frac{3}{7K_{111}}\left(\frac{8K_{22}}{k_t} - \frac{4K_{23}}{k_t^2} - K_3\right), \tag{2.26}$$

$$S_{single} = \frac{27K_{111}}{k_t^3}. \tag{2.27}$$

While each can typically be well-described by the equilateral template, a linear combination of these picking out the differences between them can yield a very different shape. This realization led to the generation of the 'orthogonal' template, $S_{orth} = -8\mathcal{K}_0 + 18\mathcal{K}_1 - 9\mathcal{K}_2$ [38].

While inflation derived from a Bunch-Davies vacuum can be written in terms of a plane wave basis with positive k modes, excited states that are not in the Bunch Davies-vacuum can have initial states with both positive and negative k. Models motivated by non-trivial vacuum states can produce shapes with denominators containing $k_1 + k_2 - k_3$ (and its permutations), rather than k_t [39, 49, 50, 51]. Unlike the equilateral and local templates, these shapes peak in

the flattened configuration, when $k_3 = k_1 + k_2$. While this shape again cannot be reconstructed perfectly using separable basis functions, an ansatz proposed as a proxy to this shape can be given by $S_{enf} = -3\mathcal{K}_0 + 6\mathcal{K}_1 - 3\mathcal{K}_2$ [50]. The shape has zero amplitude at $k_1 = k_2 = k_3$, making the conventional normalization at this configuration unsuitable for this template. Though flattened shapes such as this one are usually associated with generalized initial states, it is in some cases possible to obtain flattened shapes through single-field inflation [29].

Shapes well-approximated by $\{\mathcal{K}_0 - \mathcal{K}_2\}$

Non-Gaussian templates to describe single-field theories are not limited to equilateral and orthogonal shapes. Fast-roll single-field non-Gaussian models [40, 41] retain the scale-invariant spectra but relax the condition for slow-roll inflation. [83] showed these can be written in terms of seven constituents, four of which are S_{local}, S_{const}, $\mathcal{K}_1$, and S_{single}. The remaining three constituent shapes are [2]

$$S_3 = \frac{K_{22}}{K_{111}k_t}, \quad S_4 = \frac{K_{23}}{K_{111}k_t^2}, \quad S_5 = \frac{K_6}{K_{111}k_t^3}, \tag{2.28}$$

all of which have significant cosines with the local template.

Other shapes exist in the literature that, while not separable, to some degree interpolate between the templates discussed above and hence can be reasonably-well described by linear combinations of $\{\mathcal{K}_0 - \mathcal{K}_2\}$. For example, non-Bunch-Davies vacua generate shapes that can be equilateral, local, or enfolded [51].

Quasi-single field (QFSI) models [44, 45, 94, 63] motivated by string theory

[2]Our shape, S_n, is related to Battefeld and Grieb's $\mathcal{A}_n$, by $S_n = \mathcal{A}_n/K_{111}$.

and supergravity inspired inflation contain multiple fields, but the extra fields have masses comparable to the Hubble scale. These models can be well described by a family of bispectrum templates dependent on a single parameter, ν,

$$S_{QSFI}(\nu) \;=\; \left(\frac{3k_1k_2k_3}{k_t}\right)^{3/2} \frac{N_\nu\left[8k_1k_2k_3k_t^{-3}\right]}{N_\nu[8/27]}, \tag{2.29}$$

where N_ν is the Neumann function of order ν. This shape interpolates between the constant and local templates. Another set of models that combine multiple fields and higher-derivative terms [53, 52, 54] also generate configurations that interpolate between standard shapes, spanning the local and equilateral templates.

We use the basis modes, $\{\mathcal{K}_0 - \mathcal{K}_2\}$ to create templates for these non-separable shapes, $S_{3\text{–}5}$, S_{DBI}, S_{single}, and $S_{QFSI}(\nu)$. To demonstrate this, we generate an orthonormal basis $\{\mathcal{R}_n\}$ using the Gram-Schmidt algorithm in k-space, taking $\mathcal{R}_0 = \mathcal{K}_0$, and create a template $S_{template} = \sum_{i=0}^{n} \alpha_n \mathcal{R}_n$ as in (2.15) that reduces the covariance between the shape and template. The effectiveness of the template's fit can be quantified by the cumulative cosine. In Figure 2.1 we show how the shapes discussed above can be well modeled by templates using linear combinations of the $\{\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2\}$ templates. In each case the cumulative cosine for the template and shape exceeds 0.98.

Shapes well-approximated by more divergent $\{\mathcal{K}_0 - \mathcal{K}_n\}$

There are two strong motivations to extend template design beyond these three core templates. Firstly, expansions using the first three templates do not necessarily ensure that theoretical priors on the divergence properties are satisfied

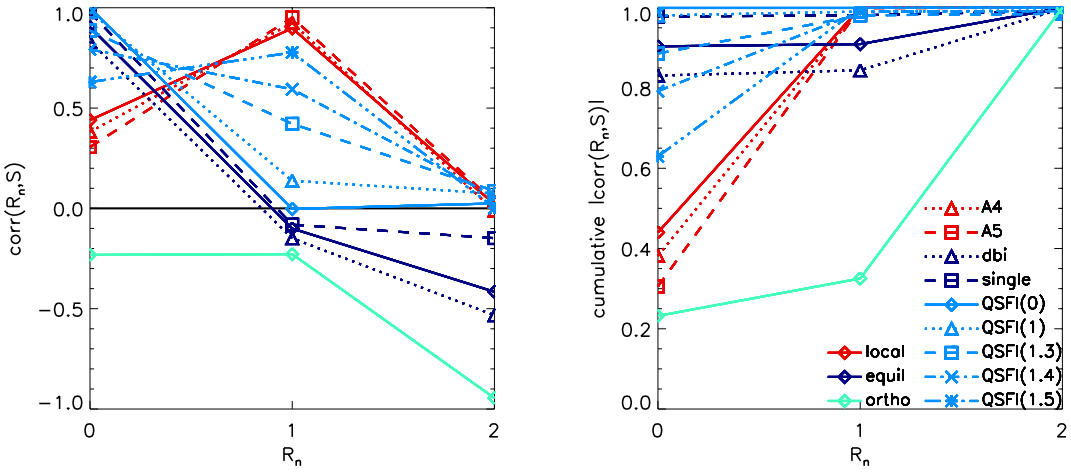


Figure 2.1: The application of the separable basis to describe shapes motivated by theoretically distinct models that span a wide array of configurations in k-space. The local, equilateral, and orthogonal templates are explicitly separable, and are included only for reference. The other shapes are not separable but are approximated by templates using linear combinations of $\{\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2\}$. We construct an orthonormal basis, $\mathcal{R}_n$, in k-space with uniform weighting, using a Gram-Schmidt decomposition for $\mathcal{K}_n$, for $0 \leq n \leq 2$, starting with $n = 0$. [Left panel] The cosines between each shape and the $\mathcal{R}_n$. [Right panel] The cumulative cosine between a constructed template $\sum_{i=0}^{n} c_i \mathcal{R}_i$ and the true shape.

by the template. An example of this is the consistency relation that requires shapes of single-field inflation to vanish in the squeezed limit [88]. However, the orthogonal and enfolded templates constructed to describe single-field shapes tend toward a constant value in the squeezed limit. [38] proposed an orthogonal template, $S_{ortho(2)}$, and [103] an enfolded template, $S_{enf(2)}$, that are somewhat more complex, using linear combinations of shapes that diverge as ${x_{sq}}^{-2}$, but they have the benefit of showing the correct divergence properties and more accurately reproducing the original non-separable shape. They can be written in

R	r			
	-2	-1	0	1
-1		$\mathcal{K}_2 = S_{local}$	$\mathcal{K}_0$ $2\mathcal{K}_1 - \mathcal{K}_2$	$-2\mathcal{K}_0 + 6\mathcal{K}_1 + 3\mathcal{K}_2 = S_{equil}$
-2	$\mathcal{K}_6$	$2\mathcal{K}_5 - \mathcal{K}_6$	$2\mathcal{K}_4 - \mathcal{K}_6$	$\mathcal{K}_0 + 3\mathcal{K}_3 - 3\mathcal{K}_4$ $\mathcal{K}_2 + 2\mathcal{K}_3 - 2\mathcal{K}_5$ $2\mathcal{K}_3 - \mathcal{K}_6$

Table 2.2: Shapes constructed from basis modes with maximal divergence x_{sq}^R ($R < 0$) which, through cancellations of the divergent terms, have a squeezed limit that diverges as x_{sq}^r. These represent an irreducible set of component shapes, for each value of R, from which general, scale invariant, separable shapes can be constructed.

terms of the $\mathcal{K}_n$ modes as

$$\begin{aligned} S_{ortho(2)} &= (1+p)S_{equil} \\ &\quad -p\left(\frac{2}{9}\mathcal{K}_0 + \frac{8}{3}\mathcal{K}_1 - 2\mathcal{K}_2 + \frac{20}{9}\mathcal{K}_3 - \frac{10}{3}\mathcal{K}_4 + \frac{4}{3}\mathcal{K}_5 - \frac{1}{9}\mathcal{K}_6\right) \qquad (2.30) \\ S_{enf(2)} &= (1+p)S_{equil} - p\left(\frac{6}{5}\mathcal{K}_0 + \frac{16}{5}\mathcal{K}_3 - \frac{18}{5}\mathcal{K}_4 + \frac{1}{5}\mathcal{K}_6\right), \qquad (2.31) \end{aligned}$$

where p is a variable chosen to maximize the template's fit to the physical shape.

Using our basis we can generalize this approach and write down classes of templates, denoted $S_{[R,r]}$, constructed from basis modes with maximal divergence R that in the squeezed limit diverge as x_{sq}^r, where $r \geq R$. In general, a shape written in terms of the basis will have a squeezed limit behavior given by

$$S_{sq} = \alpha_n \sum_{m=R}^{1} d_{nm} x_{sq}^m + \mathcal{O}(x_{sq}^2), \qquad (2.32)$$

with d_{nm} summarized in Table 2.1. We find $S_{[R,r]}$ can be written in terms of an irreducible set of shapes given in Table 2.2 for which $\alpha_n d_{nm} = 0$ for $R \leq m < r$.

S_{local} and S_{equil} are the only shapes constructed from $R = -1$ modes that respectively have -1 and vanishing divergence. There are an infinite set of shapes,

however, with constant divergence described by $\beta\mathcal{K}_0+(1-\beta)(2\mathcal{K}_1-\mathcal{K}_2)$ where β is free parameter which could take any value except $\beta = -2$, for which the equilateral template is recovered. Instead of varying the parameter β, we could instead select a value of β to generate a template from the set. $\beta = -8$ corresponds to the orthogonal template chosen by [38] to maximize the resulting shape's orthogonality with S_{local} and S_{equil}. We could then choose to write general shapes in terms of linear combinations of $\{S_{equil}, S_{ortho}, S_{local}\}$, rather than $\alpha_n\mathcal{K}_n$,

$$S_{[-1,0]} = \alpha_E S_{equil} + \alpha_O S_{ortho}, \tag{2.33}$$

$$S_{[-1,-1]} = \alpha_E S_{equil} + \alpha_O S_{ortho} + \alpha_L S_{local}. \tag{2.34}$$

If these are the only shapes being used, the normalization constraint $S_{[R,-r]}(k_0, k_0, k_0) = 1$ fixes one α coefficient.

We can extend this approach to include basis modes that diverge as ${x_{sq}}^{-2}$,

$$\begin{aligned} S_{[-2,1]} &= \alpha_E S_{equil} + \alpha_O(S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3) + \alpha_L(S_{local} + 2\mathcal{K}_3 - 2\mathcal{K}_5) \\ &\quad +(1 - \alpha_E - \alpha_O - \alpha_L)(2\mathcal{K}_3 - \mathcal{K}_6), \end{aligned} \tag{2.35}$$

$$\begin{aligned} S_{[-2,0]} &= \alpha_E S_{equil} + \alpha_O S_{ortho} + \alpha_L(S_{local} + 2\mathcal{K}_3 - 2\mathcal{K}_5) + \beta_3(2\mathcal{K}_3 - \mathcal{K}_6) \\ &\quad +(1 - \beta_3 - \alpha_L - \alpha_E - \alpha_O)(2\mathcal{K}_4 - \mathcal{K}_6) \end{aligned} \tag{2.36}$$

$$\begin{aligned} S_{[-2,-1]} &= \alpha_E S_{equil} + \alpha_O S_{ortho} + \alpha_L S_{local} + \beta_3(2\mathcal{K}_3 - \mathcal{K}_6) + \beta_4(2\mathcal{K}_4 - \mathcal{K}_6) \\ &\quad +(1 - \beta_3 - \beta_4 - \alpha_L - \alpha_E - \alpha_O)(2\mathcal{K}_5 - \mathcal{K}_6), \end{aligned} \tag{2.37}$$

$$\begin{aligned} S_{[-2,-2]} &= \alpha_E S_{equil} + \alpha_O S_{ortho} + \alpha_L S_{local} + \beta_3(2\mathcal{K}_3 - \mathcal{K}_6) + \beta_4(2\mathcal{K}_4 - \mathcal{K}_6) \\ &\quad +\beta_5(2\mathcal{K}_5 - \mathcal{K}_6) + (1 - \beta_3 - \beta_4 - \beta_5 - \alpha_L - \alpha_E - \alpha_O)\mathcal{K}_6. \end{aligned} \tag{2.38}$$

To tie this general approach to specific shapes in the literature, $S_{ortho(2)}$ and $S_{enf(2)}$ can be written in this form by the following choice of coefficients:

$$S_{ortho(2)} = (1 + p)S_{equil} - pS_{[-2,1]}\left[\alpha_E = -\frac{19}{9}, \alpha_O = \frac{5}{9}, \alpha_L = \frac{2}{3}\right] \tag{2.39}$$

$$S_{enf(2)} = (1+p)S_{equil} - pS_{[-2,1]}\left[\alpha_E = \frac{9}{5}, \alpha_O = -\frac{3}{5}, \alpha_L = 0\right]. \tag{2.40}$$

The inclusion of extra basis shapes can be particularly important when the shape has undulations and is not just a smooth monotonic function. Shapes arising out of Galileon inflation are a good example of this. Imposing a Galilean symmetry on a single-field inflation model [104, 105, 103, 106] gives rise to a non-Gaussian shape generated by three cubic interaction terms in the inflaton Lagrangian. While the shapes associated with each of these three operators, individually, are well-approximated by S_{equil} and $S_{enf(2)}$, there exist combinations of them for which the resulting Galileon shape has little overlap with any of the shapes we have mentioned so far. Non-separable templates for Galileon inflation have been developed in [103] and [106] which have high cosines both with the underlying shape and each other.

For illustrative purposes, we consider the shape presented in [103], based on equations (26)-(28) of this reference. When we use the Gram-Schmidt decomposition to construct a template with only the first three modes, we find a poor fit with a cumulative cosine of only 0.13. The Galileon shape derives from a single-field action and a Bunch-Davies vacuum so theoretical consistency requires that it vanishes in the squeezed limit. Motivated by this, if we fit the Galileon model using the 4 shapes in $S_{[-2,1]}$, we obtain a template with a cosine of 0.93. This reconstruction is not improved if we allow an unconstrained combination of the seven $\mathcal{K}_0 - \mathcal{K}_6$ modes.

We can extend our approach to $R = -3$ modes, and for example consider the following general shape that vanishes in the squeezed limit:

$$S_{[-3,1]} = \alpha_E S_{equil} + \alpha_O(S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3) + \alpha_L(S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3) + \beta_3(2\mathcal{K}_3 - \mathcal{K}_6)$$

$$+\beta_7(2\mathcal{K}_7 - \mathcal{K}_{11}) + \beta_8(S_{ortho} + 6\mathcal{K}_8 - 6\mathcal{K}_7) + \beta_9(S_{local} - 2\mathcal{K}_9 + 2\mathcal{K}_7)$$
$$+(1 - \alpha_E - \alpha_O - \alpha_L - \beta_3 - \beta_4 - \beta_7 - \beta_8 - \beta_9)(\mathcal{K}_6 - 2\mathcal{K}_{10} + 2\mathcal{K}_7). \qquad (2.41)$$

Fitting these eight distinct shapes in $S_{[-3,1]}$ to the Galileon shape, we obtain an improved template with cosine of 0.99.

The second reason to consider a basis including more divergent terms is that some inflationary scenarios, such as excited initial states and warm inflation, in which inflation occurs in a warm radiation bath [46, 47, 48] (see [107] for a review), can give rise to shapes that are more divergent than the local shape, with an overall divergence of ${x_{sq}}^{-2}$. This would suggest using an unconstrained combination of $\mathcal{K}_0 - \mathcal{K}_6$ modes, or using constrained combinations of the $R = -3$ modes for which the x_{sq}^{-3} divergent term vanishes. One such example of this is a template for warm inflation proposed by [83],

$$S_{warm} = \mathcal{K}_2 + \mathcal{K}_7 - \mathcal{K}_9. \qquad (2.42)$$

The realization that the differences between similar shapes can be important and provide an additional insight into the underlying model, implies that we should not just compare a small number of templates to the data. It is reasonable to extend beyond this and create more refined templates, sensitive to more than just properties that models have in common with the equilateral, orthogonal, and local templates.

2.2.5 Shape classification and depiction

The models discussed in the previous section reflect only a sample of the wide range of non-Gaussian inflationary shapes arising in the literature. Putting a

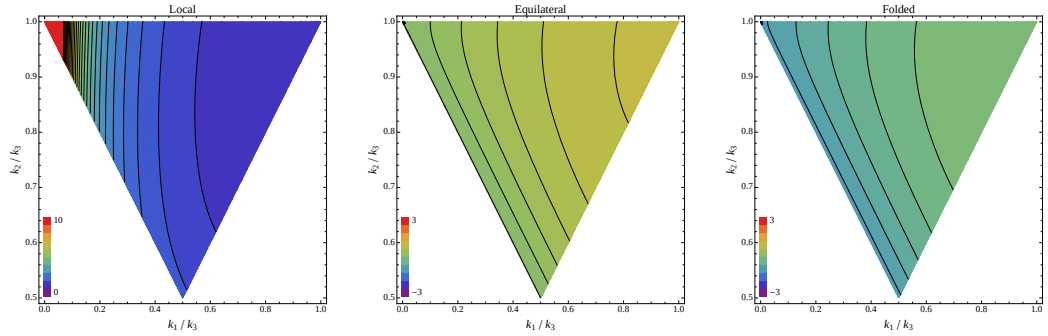
coarse filter on their properties, one might characterize them using three descriptors: i) their divergence in the squeezed limit, ii) how many modes it takes to accurately describe them, and iii) the "family" to which they belong.

Many of the physical shapes tend to be grouped in terms of a "family" resemblance to an existing template, reflecting the type of configurations of triangles with side lengths k_1, k_2, and k_3 where the shapes have most of their power [1, 108]. For scale invariant shapes this is equivalent to studying the distribution of power over the space $\{\frac{k_1}{k_3}, \frac{k_2}{k_3}\}$ for a fixed $k_3 > k_1, k_2$. This space can be pictorially represented by a triangle with sides $0 \leq \frac{k_1}{k_3} \leq 1$ and $1/2 \leq \frac{k_2}{k_3} \leq 1$. We introduce it here in the context of the shapes already discussed, because we use this format to present some of our forecasting results.

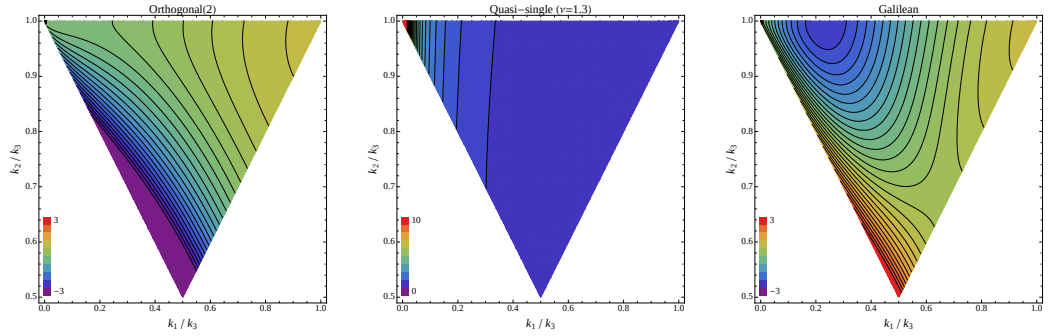
In Figure 2.2 we show examples of the shapes discussed in the previous section. $S_{const} = \mathcal{K}_0$ is the archetypal component of a family with similar power over all scales, homogeneous over the whole triangular region plotted. "Squeezed" shapes have a bispectrum amplitude that is peaked in the top left-hand corner of the plot where $\frac{k_1}{k_3} \ll 1$ and $\frac{k_2}{k_3} = 1$, while "equilateral" type shapes peak in the top right-hand corner where $\frac{k_1}{k_3} = \frac{k_2}{k_3} = 1$. "Flattened" shapes peak along the left edge, where $\frac{k_1}{k_3} + \frac{k_2}{k_3} = 1$.

Of the shapes we've discussed so far, some clearly fall within these family categories: S_{local}, S_{warm}, S_4 and S_5 are "squeezed" shapes, while S_{equil}, S_{DBI}, S_{single} are "equilateral" and S_{enf} is "flattened".

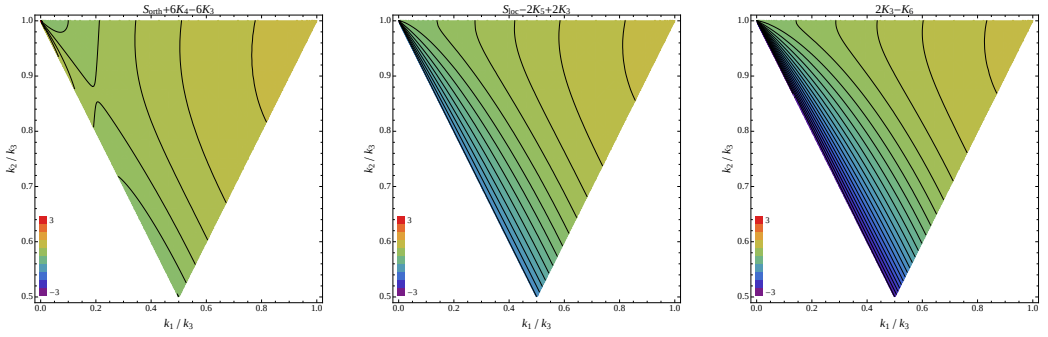
There exist other additional shapes generated by modes $\mathcal{K}_3$ through $\mathcal{K}_7$. For example, Figure 2.2 also includes three shapes that contribute to $S_{[-2,1]}$ that could describe a general single-field model with Bunch Davies vacuum. While each



(a) S_{local} (b) S_{equil} (c) S_{enf}



(d) $S_{ortho(2)}$ (e) $S_{QSFI}(\nu = 1.3)$ (f) $S_{Galileon}$



(g) $S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$ (h) $S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$ (i) $2\mathcal{K}_3 - \mathcal{K}_6$

Figure 2.2: Plots showing the comparative spatial distribution of non-Gaussian shape, $S(k_1, k_2, k_3)$, as a function of k_1/k_3 and k_2/k_3. From left to right we show [top] the local, equilateral, and enfolded separable templates, [middle] the orthogonal(2) template, and non-separable shapes derived from a QSFI model with $\nu = 1.3$ and Galileon inflation, and [bottom] shapes contributing to $S_{[-2,1]}$. All shapes are normalized to unity at the equilateral configuration ($\frac{k_1}{k_3} = \frac{k_2}{k_3} = 1$). The color scales for all but the local and QSFI shapes are the same to aid comparison.

vanishes in the squeezed limit by construction, we find they differ from the equilateral shape in still having a component of their signal focused along the flattened configuration. The comparative size of this component correlates with the divergence of the shapes from which they are created, S_{ortho}, S_{local}, and $\mathcal{K}_3$.

There are shapes that do not fall clearly into any of these families: $S_{ortho(2)}$ peaks in both the flattened and equilateral configurations, excited states can peak in squeezed and flattened configurations, and S_{QSFI} shapes interpolate between constant and local properties. Beyond this there are shapes with distinct undulating forms, the $S_{Galileon}$ shape for example, that do not peak at either edges or corners. Moreover, not all shapes within each family are alike. For example, the local and warm shapes both peak in squeezed configurations, but their divergence properties in this region are different, leading to a low cosine between them.

Given the breadth of bispectrum shapes that could be created, and the comparatively loose characteristics on which "families" are formed, there is strong motivation to ask how much information we can discern observationally about bispectra. This will help quantify how well we might determine the underlying non-Gaussian shape, if a detection of non-Gaussianity is made.

2.3 Forecasting constraints on general shapes

In the following analysis, we apply the separable, divergent basis and template classes from the previous section to assess how we can constrain the shape of primordial bispectra with upcoming CMB data. Our goal is to quantify what properties of shapes are measurable, and the respective roles of the experimen-

tal uncertainties and theoretical priors on determining distinguishability.

Motivated by a broad cross-section of models in the literature, we will focus on shapes described by basis functions $\{\mathcal{K}_0 - \mathcal{K}_6\}$ that are nearly scale-invariant and contain terms that are potentially as divergent as x_{sq}^{-2} in the squeezed limit.

We describe the Fisher matrix approach we use assuming a Planck-like CMB experiment in section 2.3.1. In section 2.3.2, we present the results of a principal component analysis for the set of shapes $S_{[-2,r]}$ with different divergence criteria in the squeezed limit imposed. Doing so generates the experiment's preferred orthogonal basis of best to worst measured bispectrum configurations, the principal components (PCs) and their corresponding uncertainties, subject to the theoretical prior. We consider the implications for shape normalization and the best-measured k-configuration in sections 2.3.3 and 2.3.4, respectively, and finish in section 2.3.5 by quantifying our potential ability to determine and distinguish shapes.

2.3.1 Fisher matrix approach

We compute the 7×7 Fisher matrix for the amplitudes of the basis modes, $\mathcal{K}_n$, $\{f_{NL}^n, n = 0, ..., 6\}$ defined in Eq. 2.19 as

$$F(f_{NL}^i, f_{NL}^j) = f_{sky} \sum_{abc,pqr} \sum_{\ell_1\ell_2\ell_3} \frac{\partial B_{\ell_1\ell_2\ell_3}^{abc\ (i)}}{\partial f_{NL}^i} (Cov^{-1})_{\ell_1\ell_2\ell_3}^{abc,xyz} \frac{\partial B_{\ell_1\ell_2\ell_3}^{xyz\ (j)}}{\partial f_{NL}^j}, \tag{2.43}$$

where $\{abc\}$ and $\{xyz\}$ each sum over the 8 possible temperature (T) and polarization (E) combinations of bispectra: $TTT, TTE, TET, ETT, TEE, ETE, EET, EEE$.

Given a general primordial shape expanded on the $\{\mathcal{K}_n\}$ basis as in 2.19, the

corresponding CMB reduced bispectrum is

$$b^{abc}_{\ell_1\ell_2\ell_3} = N \sum_n f^n_{NL} \mathcal{K}^{abc(n)}_{l_1 l_2 l_3}, \tag{2.44}$$

where $\mathcal{K}^{(n)}_{\ell_1\ell_2\ell_3}$ denotes the reduced bispectrum of the basis function $\mathcal{K}_n$ in 2.20–2.21. Here we compute it as

$$\mathcal{K}^{abc(n)}_{\ell_1\ell_2\ell_3} = \frac{1}{N^{(n)}_{perm} k_0^{2(n_s-1)}} \int x^2 dx \left[I^{ap}_{\ell_1}(x) I^{br}_{\ell_2}(x) I^{cs}_{\ell_3}(x) + \{prs\}\text{perms} \right] \tag{2.45}$$

$$I^{ap}_{\ell}(x) \equiv \frac{2}{\pi} \int_{k_{min}}^{k_{max}} dk k^{p'} \Delta^a_\ell(k) j_\ell(kx) \tag{2.46}$$

where p' is defined as in (2.21).

We have modified the CAMB [3] code [80] to numerically evaluate the values of $I^{ap}_{\ell_1}$ and then written code to appropriately combine them to form each $\mathcal{K}^{(n)}_{\ell_1\ell_2\ell_3}$. Specifically, we take $k_{min} = 6.6 \times 10^{-6}$ Mpc^{-1}, $k_{max} = 0.56$ Mpc^{-1}, and $x_{max} = 16.5 \times 10^3$ Mpc.

We include a note of caution that since the integrals over k and x for cases where p is very negative (positive) depend on having accurate transfer functions at small (large) values of k, numerical results for these integrations should be carefully checked for robustness. To verify the numerical robustness of our results, we have checked that $I^{ap}_{\ell}(x)$ obtained numerically for $p < 0$ match the expected analytic result in the Sachs-Wolfe limit. We have also quantified how the Fisher matrix results quoted in the next section are robust or exhibit instabilities to changes in the accuracy boost parameter in CAMB, which allows for fine resolution in the k and x integrals. For our purposes of computing the integrals over k and x accurately, increasing the accuracy boost parameter by, for example, a factor of 2 will compute these integrals with twice as many k or x bins. In particular, for the most divergent $\mathcal{K}_6$ mode, which is a combination

[3] http://camb.info

of the most extreme integrals (with $p = -2$ and 4) and thus we would expect to have the greatest amount of numerical error, we find that the Fisher results quoted in the next section changed by less than 0.01% when the accuracy boost was increased from 1.5 to 2. However, we find that the worst measured eigenmode, in the PCA, is far more sensitive to the integral resolution. We find with an accuracy boost of around 2 we get convergence of a few percent in all but the worst measured mode. This last mode oscillates with a variation of around 15% in the standard deviation. This sensitivity in the worst measured mode (which we will find is the least divergent shape in the squeezed limit), can affect the constraints for shapes which have a component described by this mode. In the following sections, we present our results with these cautions attached when appropriate.

The covariance matrix we use from [109] is

$$\begin{aligned}(Cov^{-1})^{abc,xyz}_{\ell_1\ell_2\ell_3} &= (C^{-1})^{ax}_{\ell_1}\left[(C^{-1})^{by}_{\ell_2}(C^{-1})^{cz}_{\ell_3} + (C^{-1})^{bz}_{\ell_2}(C^{-1})^{cy}_{\ell_3}\right] \\ &+(C^{-1})^{ay}_{\ell_1}\left[(C^{-1})^{bz}_{\ell_2}(C^{-1})^{cx}_{\ell_3} + (C^{-1})^{bx}_{\ell_2}(C^{-1})^{cz}_{\ell_3}\right] \\ &+(C^{-1})^{az}_{\ell_1}\left[(C^{-1})^{bx}_{\ell_2}(C^{-1})^{cy}_{\ell_3} + (C^{-1})^{by}_{\ell_2}(C^{-1})^{cx}_{\ell_3}\right], \end{aligned} \tag{2.47}$$

with

$$(C^{-1})^{ax}_{\ell} = \begin{pmatrix} \hat{C}^{TT}_{\ell} & \hat{C}^{TE}_{\ell} \\ \hat{C}^{TE}_{\ell} & \hat{C}^{EE}_{\ell} \end{pmatrix}^{-1} \tag{2.48}$$

$$\hat{C}^{ax}_{\ell} = C^{ax}_{\ell} + N^{ax}_{\ell}. \tag{2.49}$$

Here f_{sky} is the overall fraction of the sky observed, and we assume $f_{sky} = 0.8$. N^{ax}_{ℓ} is the instrument noise for a correlation between observables a and x. We model CMB noise by considering the three lowest frequency bands of the Planck HFI instrument for temperature and E-mode polarization, as described in the Planck Bluebook [110]. We assume each frequency channel has Gaussian beam profile

of width θ_{FWHM} and isotropic noise with error in $X = T, E$ of σ_X. The noise in each frequency channel c is then given by

$$N^{ax}_{\ell,c} = (\sigma_{x,c}\theta_{fwhm})^2 e^{\ell(\ell+1)\theta^2_{fwhm,c}/8\ln 2}\delta_{ax} \quad (2.50)$$

$$N^{ax}_{\ell} = \left[\sum_c \left(N^{ax}_{\ell,c}\right)^{-1}\right]^{-1}. \quad (2.51)$$

Our fiducial flat ΛCDM cosmology is described by the following parameters, which are consistent with the latest WMAP 9-year constraints [21]: $\Omega_b h^2 = 0.02258$, $\Omega_c h^2 = 0.1109$, $\Delta^2_{\mathcal{R}}(k_0) = 2.43 \times 10^{-9}$, $n_s = 0.963$, and $\tau = 0.088$. As has been done in other recent Fisher forecasts on non-Gaussianity parameters, such as [111], we consider the uncertainties on the non-Gaussian amplitudes independent of the uncertainties in the fundamental cosmological parameters that also affect the power spectrum, as these are comparatively small relative to the uncertainties from the bispectrum shape functions [112]. For this initial analysis, we neglect the effect of imperfect measurements of the lensing signal [113, 114], secondary anisotropies [34], and inhomogeneous sky coverage/noise on the constraints (e.g. [115, 16]).

2.3.2 Fisher matrix results

A general bispectrum can be expanded in terms of either $\mathcal{K}_n$ or the component shapes, $\{S_X\}$, in $S_{[R,r]}$, given in (2.33)-(2.38),

$$\begin{aligned}\frac{B_\Phi(k_1,k_2,k_3)(k_1k_2k_3)^2}{N} &= f_{NL}S \\ &= \sum_n f^n_{NL}\mathcal{K}_n(k_1,k_2,k_3) \\ &= \sum_X f^X_{NL}S_X(k_1,k_2,k_3)\end{aligned} \quad (2.52)$$

While the Fisher matrix we used based on $S_{[R,r]}$ automatically includes the additional priors to constrain the divergence properties, these could also be introduced into the $\mathcal{K}_n$ Fisher analysis by using Lagrange multipliers to systematically impose each divergence constraint. The latter makes no assumption a priori about what linear combinations of the shapes given in Table 2.2 should have their amplitudes constrained. While we use the shape expansion in our discussion below, we investigated both approaches and found they led to consistent conclusions.

We use the Fisher matrix in terms of $\mathcal{K}_n$ to construct Fisher matrices for the component shapes in $S_{[-2,r]}$ for $r = -2, -1, 0, 1$. In Table 2.3 we give the ℓ-space correlation coefficients based on (2.11), but here weighted by the data covariance between pairs of the component shapes, S_X and S_Y,

$$Corr_\ell(S_X, S_Y) \quad = \quad \frac{F_{XY}}{\sqrt{F_{XX}F_{YY}}}. \tag{2.53}$$

This gives a measure of the similarity of the component shapes based on how they are measured by the survey, integrated over all ℓ combinations.

We find the similarity between pairs of the four basis shapes in $S_{[-2,1]}$, each of which vanishes in the squeezed limit, are primarily related to the divergence of the shapes from which they are derived. S_{equil} and $S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$ are very similar to each other, while $S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$ and $2\mathcal{K}_3 - \mathcal{K}_6$ also have a high degree of overlap. Interestingly the $S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$ and $2\mathcal{K}_3 - \mathcal{K}_6$ shapes also have significant similarities with the shapes that diverge as x_{sq}^0. This is derived from their strong signal along the configurations between squeezed and flattened configurations, as discussed in section 2.2.5. The shape with x_{sq}^{-1} divergence constructed from the $R = -2$ modes, $2\mathcal{K}_5 - \mathcal{K}_6$, is highly degenerate with the local template; essentially this implies the two are indistinguishable from one another using the

Divergence , $x_{sq}{}^n$, n =		1				0		-1		-2
n	$Corr_\ell$	S_{equil}	$S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$	$S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$	$2\mathcal{K}_3 - \mathcal{K}_6$	S_{ortho}	$2\mathcal{K}_4 - \mathcal{K}_6$	S_{local}	$2\mathcal{K}_5 - \mathcal{K}_6$	$\mathcal{K}_6$
1	S_{equil}	1	0.97	0.41	0.07	-0.11	0.33	0.23	0.21	0.003
	$S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$		1	0.24	-0.10	-	-	-	-	-
	$S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$			1	0.94	0.78	0.98	-	-	-
	$2\mathcal{K}_3 - \mathcal{K}_6$				1	0.80	0.95	-0.29	-0.20	-0.03
0	S_{ortho}					1	0.72	-0.48	-0.40	-0.06
	$2\mathcal{K}_4 - \mathcal{K}_6$						1	-0.12	-0.04	-0.01
-1	S_{local}							1	0.99	0.66
	$2\mathcal{K}_5 - \mathcal{K}_6$								1	0.68

Table 2.3: Correlation coefficients between shapes that diverge as x_{sq}^n. These shapes are components in the general template classes, $S_{[-2,r]}$, for $r \leq n$.

CMB data.

The unmarginalized errors, $\sigma(f_{NL}^X) = 1/\sqrt{F_{XX}}$, give the uncertainty in the measurement of a specific template if the underlying theory is known to be wholly described by that template. We find these are comparatively insensitive to the integral resolution discussed in section 2.3.1. The covariance matrices obtained from inverting the Fisher matrices give the uncertainties on the amplitudes of the component shapes, $\sigma(f_{NL})$, marginalized over the freedom allowed by each model. The marginalization does make the results precision dependent in the worst measured mode, i.e. the results are accurate to better than 15%.

We summarize the results in Table 2.4. The covariance matrix in each case

can be diagonalized to obtain the orthonormal eigenvectors,

$$\hat{e}_i \;=\; \sum_X c_{iX} S_X, \tag{2.54}$$

and associated eigenvalues, which give the variances $\sigma^2(b_i)$ in the amplitudes of the eigenvectors. These then provide a way to rank the best to worst measured bispectra. Given this orthonormal basis, any general bispectrum may be expanded as

$$f_{NL} S = \sum_i b_i \hat{e}_i. \tag{2.55}$$

The principal components obtained by diagonalizing the covariance matrix are not immediately 'shapes' in the way we considered so far. They have unit norm with respect to the component shape basis, $\sum_X |c_{iX}|^2 = 1$, rather than being normalized at the equilateral configuration, $\sum_X c_{iX} S_X(k_0, k_0, k_0) = 1$.

If we restrict the shapes to those described by the first three modes, marginalization does not significantly alter the constraints from the unmarginalized errors, i.e. the three common templates are essentially the principal components (PC) of the covariance matrix, with the eigenvalues showing that the more divergent the shape, the better it is measured. In contrast, when extended to general shapes, constructed of all seven modes, we find marginalized errors for individual shapes are far larger because of observational similarities between shapes of similar divergence, or similar properties in the flattened limit. It seems that only K_6 is well constrained if any shape from the $S_{[-2,r]}$ type is allowed.

When extended to shapes constructed of seven modes, the correspondence between the PC's and divergence remains. We find that, in general, divergence in the squeezed limit, followed by a second divergence measure, corresponding to the signal near the flattened configurations, can be used as coarse indicators of comparative constraining power with the CMB. For the general shape

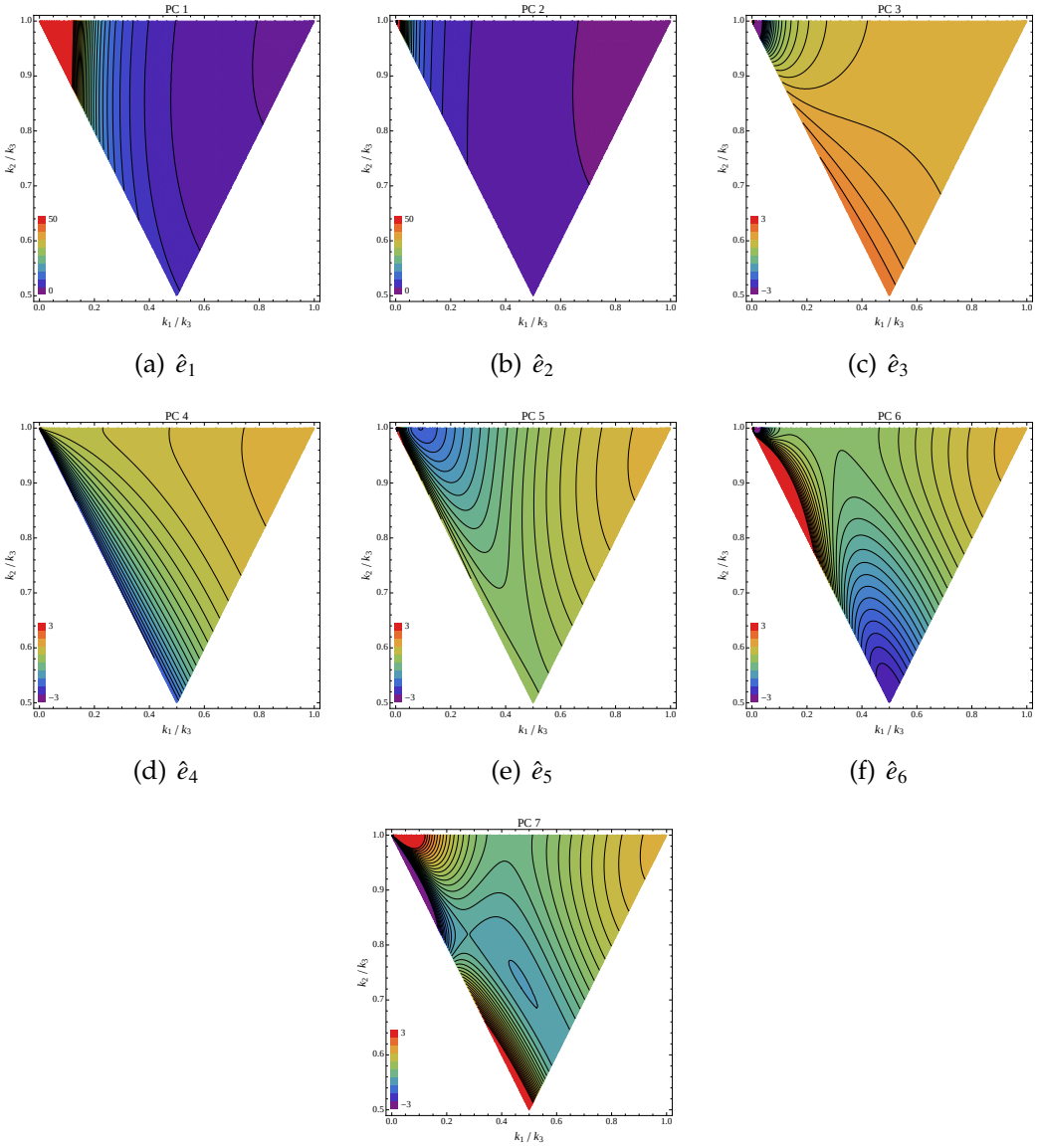
Divergence $x_{sq}{}^n$	Shape	Unmarg. $\sigma(f_{NL})$	$\sigma(f_{NL})$ marg.^{d} over shape					
			$S_{[-1,r]}$		$S_{[-2,r]}$			
			r=0	-1	r=1	0	-1	-2
1	S_{equil}	43	44	45	351	365	387	404
	$S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$	41	-	-	293	-	-	-
	$S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$	32	-	-	920	1064	-	-
	$2\mathcal{K}_3 - \mathcal{K}_6$	18	-	-	468	742	1425	1428
0	S_{ortho}	19	19	22	-	362	364	366
	$2\mathcal{K}_4 - \mathcal{K}_6$	23	-	-	-	1000	1018	1034
-1	S_{local}	3	-	4	-	-	1073	1081
	$2\mathcal{K}_5 - \mathcal{K}_6$	4	-	-	-	-	1074	1082
-2	$\mathcal{K}_6$	0.011	-	-	-	-	-	0.018

Table 2.4: The uncertainties on the amplitudes of the component shapes, in the general template classes $S_{[-2,r]}$, that diverge as x_{sq}^r in the squeezed limit. We give both the unmarginalized errors, assuming the underlying shape is exactly described by the component shape, and the marginalized errors if we allow the shape to be a general linear combinations of components consistent with the prior on the divergence properties.

without any additional divergence constraints, the best measured PC is almost completely composed of the most divergent shape, $\mathcal{K}_6$. The second best measured PC has dominant contributions from S_{local} and $2\mathcal{K}_5 - \mathcal{K}_6$ with which it is very degenerate. If the general shape is restricted to have vanishing divergence in the squeezed limit, then the best measured PC is very similar to a shape like $2\mathcal{K}_3 - \mathcal{K}_6$ which has large signal in the flattened configurations despite vanishing in the squeezed limit. The next best measured PC is then similar to shapes like equilateral or the orthogonal-derived shape $S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$, which has less power on flattened configurations. In both cases, none of the templates look like the two worst measured modes, which exhibit large oscillatory features along flattened configurations.

Shape	$S_{[-2,-2]}$		$S_{[-2,-1]}$		$S_{[-2,0]}$		$S_{[-2,1]}$	
	$\sigma(b_i)$	$\sigma(f_{NL,i})$	$\sigma(b_i)$	$\sigma(f_{NL,i})$	$\sigma(b_i)$	$\sigma(f_{NL,i})$	$\sigma(b_i)$	$\sigma(f_{NL,i})$
$\hat{e}_1$	0.0084	0.016	2.8	4	20	18	31	11
$\hat{e}_2$	4	5	22	18	38	32	138	45
$\hat{e}_3$	24	18	38	32	49	39	1321	26
$\hat{e}_4$	38	32	491	41	1321	25	7518	28
$\hat{e}_5$	522	43	1420	22	9576	16		
$\hat{e}_6$	1505	19	9576	16				
$\hat{e}_7$	9576	16						

Table 2.5: Properties of the principal components for each template class $S_{[-2,r]}$ in terms of their component shapes. The properties in the squeezed limit is determined by the value of r. The table provides uncertainties, for a unit norm eigenvector, $\sigma(b_i)$, and an effective $\sigma(f_{NL}(\hat{e}_i))$, when the eigenvector is normalized consistently at the equilateral configuration.



(a) $\hat{e}_1$ (b) $\hat{e}_2$ (c) $\hat{e}_3$

(d) $\hat{e}_4$ (e) $\hat{e}_5$ (f) $\hat{e}_6$

(g) $\hat{e}_7$

Figure 2.3: Configurations of the principal components for $S_{[-2,-2]}$, a general shape that can be as divergent as x_{sq}^{-2} in the squeezed limit. The plots show the amplitude of the eigenvectors for the best $\hat{e}_1$ to worst $\hat{e}_7$ measured modes as a function of $\frac{k_1}{k_3}$ versus $\frac{k_2}{k_3}$. The principal components are each normalized to be unity at the equilateral configuration.

2.3.3 Drawbacks of normalization at the equilateral configuration

As stated earlier, the PC's as they are originally generated, are not shapes in the usual sense because they are not bispectra normalized at $k_1 = k_2 = k_3 = k_0$. They have a unit norm in terms of the basis shapes. With this normalization, as usual in PCA, their eigenvalues quantify which combinations of the basis shapes are best and worst constrained by data, and the eigenvectors can be combined to create general shapes.

We can convert $\sigma(b_i)$ to an effective $\sigma(f_{NL}(\hat{e}_i))$, corresponding to the amplitude of each eigenvector shape normalized in the conventional way, $\sigma(f_{NL}(\hat{e}_i)) = |\sigma(b_i)\hat{e}_i(k_0, k_0, k_0)|$. Table 2.5 gives the values of $\sigma(b_i)$ and $\sigma(f_{NL}(\hat{e}_i))$. We quote the results when both temperature and polarization data are included. We find that the exclusion of the E-mode polarization from the Fisher analysis does not noticeably change the shape of the principal components, but does increase the eigenvalues by about a factor of $\sim 1 - 3$ across all eigenvectors. The constraints on all but the last eigenvalue under each divergence constraint shown in Table 2.5 are accurate to a few percent. The worst measured eigenmode is measured to $\sim 15\%$ accuracy.

Normalizing our PC's at the arbitrarily chosen equilateral configuration allows us to compare them to other shapes consistently at one point in k-space. $\sigma(f_{NL})$ does not in itself, however, quantify a shape's overall variance across all k. An analogous situation arises in quoting uncertainties on the power spectrum amplitude from two different surveys, say a large-scale CMB survey and a galaxy survey. Both surveys could quote uncertainties at a common arbitrary

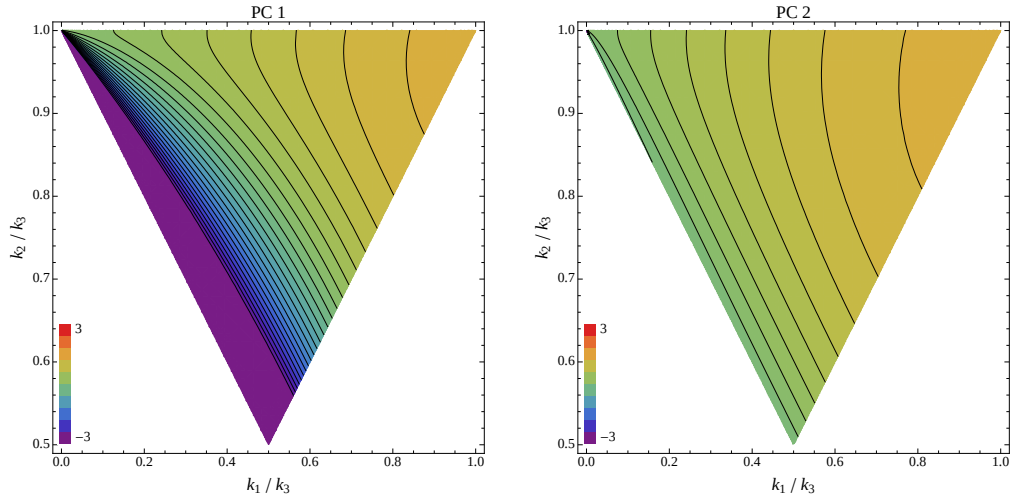
scale, say $k_0 = 0.05h/Mpc$, but while this uncertainty might represent the best measured scale for the galaxy survey, it would grossly overstate the minimum uncertainty in the CMB survey, which is best measured at a much larger scale.

It is entirely possible for a well measured mode to have a significant part of its small variance located in the equilateral configuration, while a poorly measured mode could have its lowest variance in the equilateral configuration but be poorly measured over other regions of k-space. Indeed we find this to be the case, given that the best measured shapes have signal peaked near the squeezed, rather than equilateral, configuration. This means that $\sigma(f_{NL})$ is not a useful measure in itself to assess how well a shape can be measured. This shortcoming of the conventional normalization has been discussed previously in other studies, e.g. [37] and [16], where alternative normalization schemes based on an integrated total amount of non-Gaussianity have been proposed.

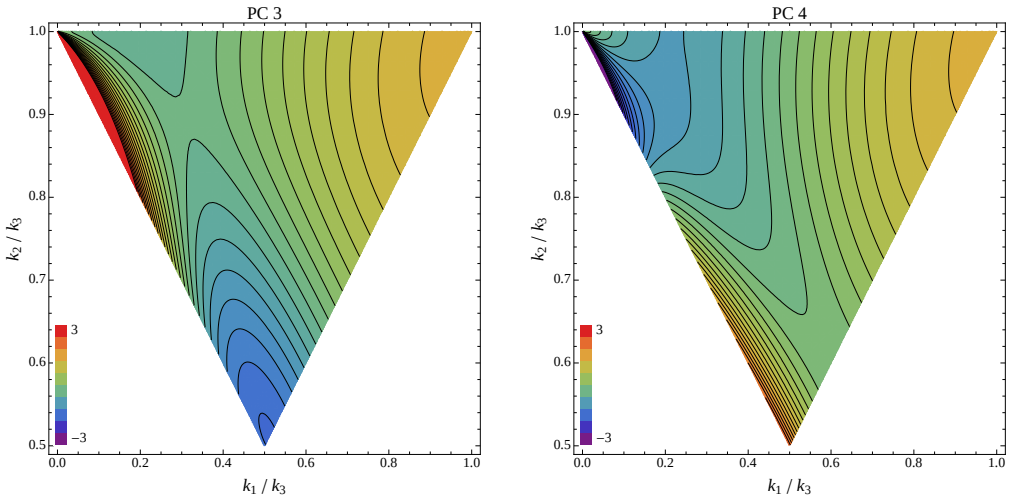
The overall spread in uncertainties from the best to worst eigenvector is much reduced when normalized at the equilateral configuration and can in some cases produce a switch in the ordering of the modes for $\sigma(f_{NL})$ relative to that of $\sigma(b_i)$. This does not present an inconsistency in the analysis, but simply demonstrates the perils of considering a normalization at an arbitrary scale.

Figure 2.3 shows the variety of profiles in the 2-dimensional $(\frac{k_1}{k_3}, \frac{k_2}{k_3})$ space shown in the triangle plots. Given that the power spectrum we consider is not perfectly scale invariant, there is some small dependency of the bispectrum amplitude on the value of k_3, described by p' in (2.21). The spatial profiles, however, in terms of $\frac{k_1}{k_3}$ and $\frac{k_2}{k_3}$ are k_3-independent.

The gradients in the PC configurations reflect the rough ordering from



(a) $\hat{e}_1$ (b) $\hat{e}_2$



(c) $\hat{e}_3$ (d) $\hat{e}_4$

Figure 2.4: As in Figure 2.3, but showing the configurations of the principal components for $S_{[-2,1]}$, a general shape that vanishes in the squeezed limit.

squeezed to flattened to equilateral as the modes span from best to worst. The complementarity of the eigenvectors, reflected by the different directions of gradients of the signals in the configuration space, has implications for the location of the best measured configuration, as we discuss in section 2.3.4.

2.3.4 Best measured k-configurations

In the analysis that follows, we avoid splitting up bispectra into shapes and amplitudes, normalized at an arbitrary configuration. Instead we consider the overall constraints on the bispectrum, $B(k_1, k_2, k_3)$, itself up to the constant normalization, given in (2.2), $f_{NL}S = k_1^2 k_2^2 k_3^2 B(k_1, k_2, k_3)/N$.

The eigenmodes and eigenvalues from the PCA provide a way to compute an error on a general k-space bispectrum. We can calculate the posterior distribution of the uncertainties on $f_{NL}S$ given the data, D, with a theoretical prior given by the eigenvectors $\{\hat{e}_i\}$,

$$p(f_{NL}S\,|D) = \int \prod_{i=0}^{n} db_i p(f_{NL}S\,|b_i) p(b_i|D), \tag{2.56}$$

$$p(f_{NL}S\,|b_i) = \delta(f_{NL}S - \sum_{i=0}^{n} b_i \hat{e}_i(k_1, k_2, k_3)), \tag{2.57}$$

$$p(b_i|D) = \frac{1}{\sqrt{2\pi}\sigma(b_i)} \exp\left(-\frac{b_i^2}{2\sigma^2(b_i)}\right). \tag{2.58}$$

Under this assumption of Gaussian errors this gives the commonly used result,

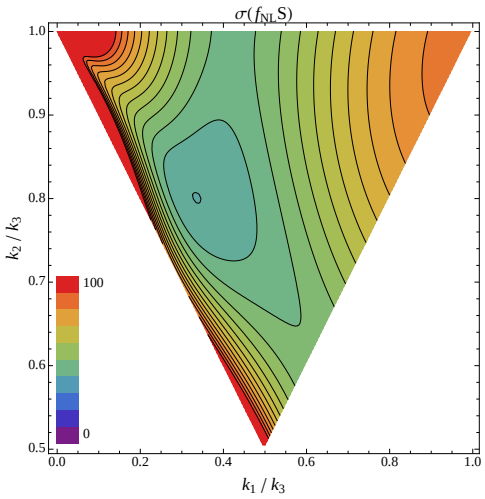
$$\sigma^2\left(f_{NL}S\,(k_1, k_2, k_3)\right) = \sum_i \sigma^2(b_i)\hat{e}_i(k_1, k_2, k_3)^2. \tag{2.59}$$

This equation for computing the error can be applied to each set of PC's generated for each divergence scenario in the previous section. The errors in the $(\frac{k_1}{k_3}, \frac{k_2}{k_3})$ configuration space can be plotted and the best measured k-configuration, and the associated uncertainty, calculated for each scenario.

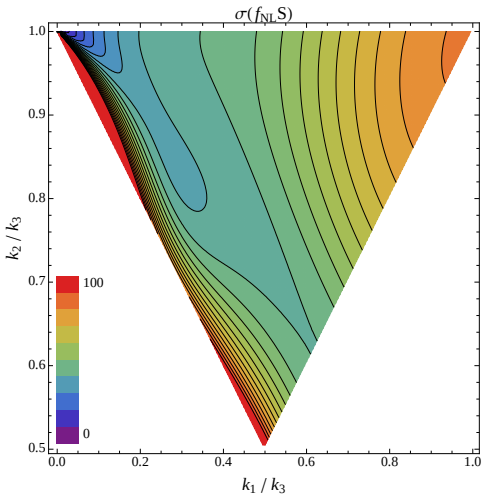
$\sigma(f_{NL}S)$ varies only very weakly across slices in k_3; its functional form can be divided into a dependence on $(\frac{k_1}{k_3}, \frac{k_2}{k_3})$ and a weak dependence on k_3, going as $k_3^{2(n_s-1)}$, for fixed $\frac{k_1}{k_3}$ and $\frac{k_2}{k_3}$. For our choice of theoretical priors on the model, $\sigma(f_{NL}S)$ decreases with increasing k_3. This is because the noise scales as the

signal for the near scale-invariant theoretical prior we impose. An alternative prior would give very different dependencies on k_3. For example if we were to remove the theoretical prior all together and model the bispectrum amplitude as bins in k, the only constraints on the model come from the observational uncertainties, and the noise would diverge exponentially on small scales.

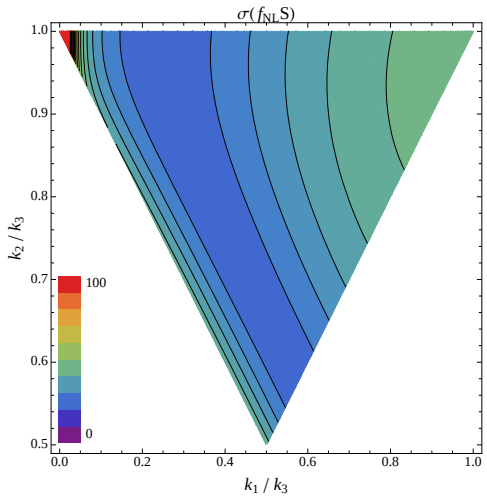
The weak k_3 dependence implies that the uncertainties at one k_3 reasonably reflect the overall uncertainties if one were to marginalize over k_3. Figure 2.5 shows the error on the $k_3 = 0.01$ Mpc^{-1} slice for three different divergence cases. The location of minimum $\sigma(f_{NL}S)$ comes from the sum of the eigenmodes that is weighted by each mode's error, which arises out of the complementarity of the degeneracy directions of the PC's. We find the location of the best measured configuration is consistent for the scenarios that diverge as x_{sq}^{-2} through to a constant in the squeezed limit for $R = -2$. This location is not situated in any one of the corners of the triangle plot associated with squeezed, equilateral, and flattened configurations. Instead it is somewhat centrally located adjacent to the flattened edge. The best-measured configurations are located at $\frac{k_1}{k_3} \approx 0.32$ and $\frac{k_2}{k_3} \approx 0.80$, with minimum $\sigma(f_{NL}S) \approx 37$. For the vanishing divergence prior, the best-measured configuration approaches the squeezed limit, as we have required the noise to scale as the shapes, which go to zero there. We also find that for shapes constructed from the local, equilateral, and orthogonal templates ($R = -1$), the best measured location spans a degeneracy direction also along the flattened edge, with minimum $\sigma(f_{NL}S) \approx 20$.



(a) Diverges as $1/k^2$ in the squeezed limit.



(b) Vanishing divergence in the squeezed limit.



(c) Diverges as $1/k$ in the squeezed limit.

Figure 2.5: Example contour slices of $\sigma(f_{NL}S)$ under different divergence constraints, [top left] $S_{[-2,-2]}$, [top right] $S_{[-2,1]}$, and [bottom] $S_{[-1,-1]}$.

While the best measured region, in which the error is a minimum, is useful in the absence of knowledge about the theory, the signal-to-noise ratio, for a given underlying model can also help determine a survey's ability to distinguish between shapes. We consider this in the following subsection.

2.3.5 Shape determination and distinguishability

We now turn to discussing a central question of the paper: given a detection of non-Gaussianity using a specific template, what can be confidently inferred about the true underlying shape? We have already considered this from one perspective in section 2.3.2 by considering the uncertainties in ascribing a detection using a template to the template's shape itself. If we allow for the possibility that a detection using a specific template could be detecting the component of another shape allowed by the theoretical prior we are considering, then the errors on shape determination can increase significantly, especially for shapes that do not peak in the squeezed or flattened configurations.

In this section we approach the question of shape distinguishability from a second direction, considering the range of possible general shapes, under a divergence prior, that could create the detected template signal and fit the bispectrum data within some confidence range. Such analyses have already been considered in the context of specific models, for example, how well we might disentangle a QSFI model (e.g. in 2.29) from S_{equil} or S_{local} as a function of ν [94, 63]. Here we extend this approach to a more general shape, and consider what implications a detection with one of the common templates has for general models. For specificity we consider a subset of general shapes consistent with

$S_{[-2,1]}$,

$$S_{gen} = (1-\alpha_X-\alpha_Y-\alpha_Z)S_{equil}+\alpha_X S_X+\alpha_Y S_Y+\alpha_Z S_Z, \quad (2.60)$$

where $S_{X,Y,Z}$ can be $\{S_{ortho}+6K_4-6K_3, S_{local}-2K_5+2K_3, 2K_3-K_6\}$. This is investigating a general set of single-field inflation models from which $S_{ortho(2)}$ in 2.39 and $S_{enf(2)}$ in 2.40 are drawn.

How large must a template signal f_{NL}^T be to be confident that the signal is not from a different, more general shape S_{gen}? We set this distinguishable detection threshold to be $\sigma(f_{NL}^T)$, the error on f_{NL}^T for the template, marginalized over f_{NL}^S, the amplitude of the general shape. The marginalized constraint is computed by inverting the 2×2 Fisher matrix for (f_{NL}^T, f_{NL}^S). Thus we are comparing two shapes, where one is a template, and the other is a general shape, in which α_X, α_Y, and α_Z parametrize the deviation from S_{equil}.

In the simplest case, we allow only α_X to be non-zero, such that S_{gen} is a linear combination of two shapes, S_{equil} and S_X, that varies with one parameter. In Figure 2.6, we show $\sigma(f_{NL}^T)$ for the local, equilateral, and orthogonal templates when S_X takes different forms. The minimum value of $\sigma(f_{NL}^T)$ for each template across all values of α recovers the unmarginalized errors of each. A detected value of f_{NL}^{equil} must be larger to produce a 1σ detection of the equilateral shape, as opposed to a more general shape with $\alpha \neq 0$, while f_{NL}^{local} never has to be much larger than the unmarginalized $\sigma(f_{NL}^{local})$ to favor the local model over this general shape, because S_X and S_{local} are weakly correlated.

To illustrate the use of Figure 2.6, for example, a detection of $f_{NL}^{orth}=40$, while greater than the unmarginalized error of 19, would only be sufficient to rule out a false 1σ detection of S_{gen} with $S_X = 2\mathcal{K}_3-\mathcal{K}_6$ for $-5 \lesssim \alpha \lesssim 0.9$. On the other hand, if f_{NL}^{orth} is detected to be larger than 46, then S_{gen} of the 1-parameter form

would be disfavored, as $\sigma(f_{NL}^{orth})$ is smaller than this over all values of α. Models with $S_X = 2\mathcal{K}_3 - \mathcal{K}_6$ are most easily differentiated from S_{equil} because they have the lowest correlation with S_{equil}.

An application of comparing constraints on α_X from two distinct templates is to test whether a given model is consistent with or disfavored by the data. If two template measurements each individually remain consistent with two non-overlapping regions of α-space, then it would be clear that modeling the underlying shape with α alone is not able to produce a viable model. This would be true for dual measurements of $\{f_{NL}^{equil} = 60, f_{NL}^{ortho} = 45\}$ for $S_X = 2\mathcal{K}_3 - \mathcal{K}_6$, since they would imply non-overlapping ranges of α, $-0.7 \leq \alpha \leq 0.3$ versus $1.3 \leq \alpha \leq 3.8$ to each be consistent with the data.

We can extend the same analysis to a comparison between templates and a 2-parameter general shape by allowing both α_X and α_Y to vary simultaneously, while α_Z is fixed to zero. For example, $S_{enf(2)}$ is a specific template for which this is true. In Figure 2.7 we show $\sigma(f_{NL}^{equil})$ and $\sigma(f_{NL}^{orth})$ over different choices of the 2-dimensional space and find that there exist degeneracy directions that are not fully captured by the 1-dimensional projections in Figure 2.6. We find that $\sigma(f_{NL}^{local})$ remains close to the unmarginalized value in this case as well.

In the most general 3-parameter model, we can ask the question of whether there is any area of this space corresponding to a general model that vanishes in the squeezed limit, with a significant enough overlap with the local template to require that a potentially detected f_{NL}^{local} be much greater than the unmarginalized value of 3. If this were the case, then it may be that a local template detection cannot definitively rule out a general shape that satisfies the single-field consistency relation. However, we find that nowhere in the parameter space

does the $\sigma(f_{NL}^{loc})$ become greater than 4.2, showing that a detection of the local template above this threshold would effectively rule out a general shape, vanishing in the squeezed limit, subject to the assumption that it can be written in terms of our basis in $S_{[-2,1]}$. The same distinguishing power is not present for S_{local} if we allow a weaker prior given by $S_{[-2,-1]}$. In this case the significant cosine between S_{local} and $2\mathcal{K}_5 - \mathcal{K}_6$, means we may never be able to confidently attribute a detection with S_{local} to be definitive evidence that the diverging signal is unambiguously S_{local}. A long shot could be to additionally look at the correlation of the bispectrum signal with $2\mathcal{K}_4 - \mathcal{K}_6$ which is mildly negatively correlated with S_{local} and essentially uncorrelated with $2\mathcal{K}_5 - \mathcal{K}_6$.

This last point raises an interesting application of our study: to ask if there are distinct, new templates that we might use to learn about the origins of a detected non-Gaussian signal. In the context of models described by the first three modes, $\mathcal{K}_0$ to $\mathcal{K}_2$, the local, equilateral, and orthogonal templates are almost perfectly aligned with the principal components. If we extend the templates to include $\mathcal{K}_3$ through $\mathcal{K}_6$, however, we find these no longer represent the PC's. For example, what might be the best way to extend the template pool to search for signatures of single-field inflation models with Bunch-Davies vacua? In the context of $r = 1$ shapes, $2\mathcal{K}_3 - \mathcal{K}_6$ is well-aligned with the best measured PC and is only mildly correlated with the existing templates which would make it a reasonable candidate to add as an additional template. We show the resulting constraints on general shapes in Figures 2.6 and 2.7. The figures show that this template probes regions of the allowed α-space which the equilateral and orthogonal templates do not constrain in the same way. Thus it may be possible to combine constraints from the common templates and a small number of new templates, like $2\mathcal{K}_3 - \mathcal{K}_6$, to probe the underlying shape of non-Gaussianity.

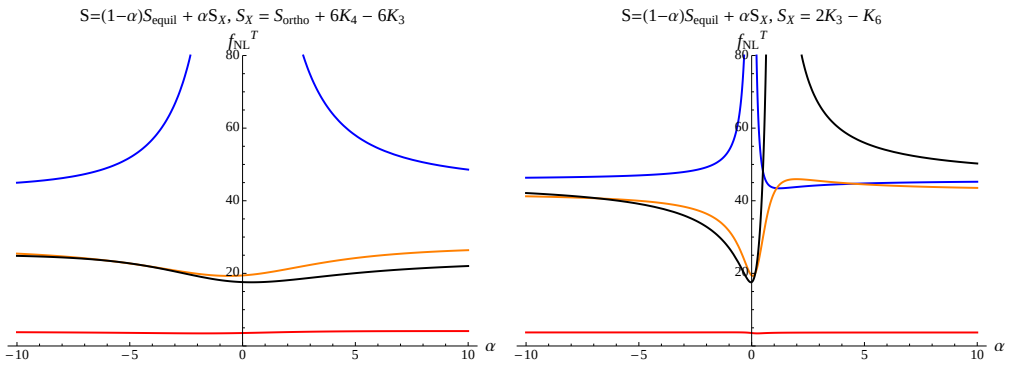


Figure 2.6: Detection thresholds on the amplitude of templates, f_{NL}^{T}, for distinguishing between the template and a general shape, $S = (1-\alpha)S_{equil} + \alpha S_X$, at the 1σ confidence level. [Left] $S_X = S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$, [right] $S_X = 2\mathcal{K}_3 - \mathcal{K}_6$. Blue, orange, red, and black curves denote f_{NL}^{equil}, f_{NL}^{orth}, f_{NL}^{local}, and $f_{NL}^{2\mathcal{K}_3-\mathcal{K}_6}$, respectively. Since $S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$ is very similar to $2\mathcal{K}_3 - \mathcal{K}_6$, the case where $S_X = S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3$ is not shown.

2.4 Conclusion

At the heart of this work is the discussion about how uncertainties quoted on shape detection are inherently dependent on the underlying assumptions made about the shape. While a detection of non-Gaussianity with any template will be extraordinarily transformative in our field, its interpretation, in what it tells us about the underlying shape, has to be considered carefully in terms of our underlying theoretical prior we impose. Even if no detection of non-Gaussianity is made, upper bounds on the deviations from Gaussianity according to templates will have broader impacts for constraints on general shapes.

We have presented an approach for quantifying how well upcoming CMB temperature and E-mode polarization data can determine the shape of primordial non-Gaussianity under minimal assumptions. We proposed a set of polynomial divergent basis functions, $\{K_n\}$, that are well-tuned to describing many

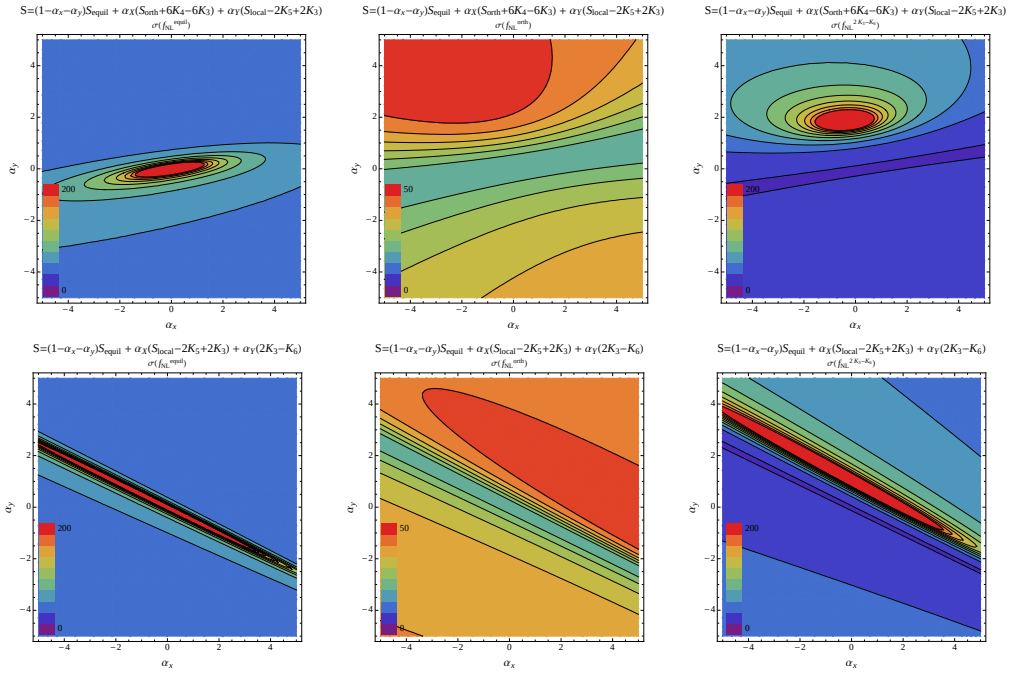


Figure 2.7: Detection thresholds on the amplitude of templates, [left] f_{NL}^{equil}, [center] f_{NL}^{orth}, and [right] $f_{NL}^{2\mathcal{K}_3-\mathcal{K}_6}$, for distinguishing between each template and two forms of a general shape S at the 1σ confidence level. The general shapes considered are [top panels] $S_{gen} = (1-\alpha_X-\alpha_Y)S_{equil} + \alpha_X(S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3) + \alpha_Y(S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3)$ or [bottom panels] $S_{gen} = (1-\alpha_X-\alpha_Y)S_{equil} + \alpha_X(S_{local} - 2\mathcal{K}_5 + 2\mathcal{K}_3) + \alpha_Y(2\mathcal{K}_3 - \mathcal{K}_6)$. Contours for f_{NL}^{loc} are not pictured, because the marginalized $\sigma(f_{NL}^{loc})$ remains close to its unmarginalized value over these 2-dimensional spaces. The case where $S_X = S_{ortho} + 6\mathcal{K}_4 - 6\mathcal{K}_3$ and $S_Y = 2\mathcal{K}_3 - \mathcal{K}_6$ is not pictured because S_X and S_Y are nearly uncorrelated, thus no additional information is revealed beyond that in Figure 2.6.

nearly scale-invariant, smoothly varying, but potentially divergent shapes discussed in the literature. We find we need only three to seven modes to generate matched templates to describe a wide range of physically motivated shapes. In this sense, the divergent basis is more efficient than the polynomial basis used in previous studies (e.g. [16]). Each $\mathcal{K}_n$ in our basis is generally divergent, but linear combinations of the $\mathcal{K}_n$ can be constructed to have cancellations in the squeezed limit, thus creating templates that are less divergent (e.g. equilateral

shape). For example, S_{equil} and $(2\mathcal{K}_3 - \mathcal{K}_6)$ both vanish in the squeezed limit, but still have a low correlation because the latter has more power near the flattened and squeezed configurations.

Using the $\{\mathcal{K}_n\}$ it is straightforward to form template classes, $S_{[R,r]}$, that have specific, common divergence properties in the squeezed limit. Each class is constructed from an irreducible set of shapes, that while constructed out of a basis sets with maximum divergence x_{sq}^R ($R < 0$), through cancellations of divergent terms, have squeezed limit x_{sq}^r ($r > R$). The choice of R controls how many basis modes are used to develop the templates, e.g. $R = -1$ includes $\mathcal{K}_0$ through $\mathcal{K}_2$, while $R = -2$ uses $\mathcal{K}_0$ through $\mathcal{K}_6$. As R becomes more negative it allows templates to be refined and shapes with a broader set of features across configuration space to be modeled.

The classes allow templates to be developed with priors that are well-motivated by theories: $S_{[R,1]}$ represents the class of all single-field models derived from a Bunch-Davies vacuum, $S_{[R,-1]}$ in addition includes all multi-field models that diverge like the local shape in the squeezed limit, and $S_{[R,-2]}$ is the most general class of shapes which includes models from non-Bunch-Davies vacuum initial states.

While the constituent shapes making up each class have the same divergence properties in the squeezed limit, away from this limit they have power weighted differently in the configuration space. For example, we discuss a new shape, $2\mathcal{K}_3 - \mathcal{K}_6$, used in $S_{[R,1]}$, that has the same squeezed limit behavior as the equilateral shape but has ℓ-space cosines with the standard equilateral, orthogonal, and local templates of 0.07, 0.80 and -0.29 respectively. While the divergent terms cancel in the squeezed limit, $2\mathcal{K}_3 - \mathcal{K}_6$ has significant power just away

from the squeezed and flattened configurations that differentiates it from the equilateral shape, and leads to it being most similar (though only mildly) with the orthogonal template.

An added benefit of using the divergent basis and template classes to consider general shapes is that it ties together the methods we use to search for evidence of shapes with CMB data to LSS constraints from a scale-dependent halo bias, which probes the squeezed limits of shapes. It is well-known that templates for physical shapes which work for generating CMB predictions can fail when used for LSS predictions [64], because while CMB constraints represent a weighted average over all k-space configurations, the halo bias traces the squeezed limit region of k-space only. Thus our approach provides a way of generating templates that can potentially be used consistently for both CMB and LSS studies.

We adopt a Fisher matrix approach modeled on a Planck-like survey to estimate uncertainties on the amplitudes of shapes within each shape class, r. As summarized in Table 2.4, we computed the uncertainties on shape attribution under each prior and how these uncertainties on confidently being able to determine that a template is the true shape can change substantially dependent upon the type of prior we impose.

We find that the best measured shapes are those with the strongest divergence and with principal power near squeezed and flattened k-configurations. Though the conventional approach is to quote constraints at the equilateral configuration, $k_1 = k_2 = k_3$, we show, as summarized in Table 2.5, that this convention can mask how well or badly a shape is measured, as doing so has the effect of re-normalizing constraints such that badly measured modes can appear to

have constraints similar to the best measured mode.

Using the PCA results, we map out the k-dependence of the constraints for a general shape given a prior, and show its dependence on the prior. For all but the $r = 1$ case, the best measured location is not in the equilateral configuration where shapes and constraints on f_{NL} are typically normalized, but in a configuration that is neither squeezed, flattened, or equilateral, but somewhere in between. This best measured location at roughly $k_1/k_3 \approx 0.32$ and $k_2/k_3 \approx 0.80$ arises out of the complementary gradients of the power in the PC's. For the $r = 1$ case, the best measured location is weighted more strongly towards the squeezed configuration, reflecting that the signal and the noise, with which it is correlated, both go to zero in this limit.

Given our parametrization of a general shape under a divergence prior, we then ask how well it could be constrained using measurements of amplitudes of common templates, like the local, equilateral, and orthogonal templates. We focus on the class of general shapes that can represent the possible range of single-field models that vanish in the squeezed limit ($r = 1$). We calculate bounds on the subset of shapes that can remain consistent with constraints on the local, equilateral, and orthogonal templates, and find again–consistent with what we found earlier in the analysis–that templates with more power in the squeezed and flattened configurations provide more stringent constraints on this class of general shapes. Thus, the local, equilateral, and orthogonal templates serve different roles in constraining general shapes; the local template, if detected with sufficient amplitude, will rule out any shape of this type, while the equilateral and orthogonal templates serve to put constraints around different regions of the parameter space. In this sense, constraints from different templates can be

complementary.

Furthermore, a general (unknown) shape, will have different overlaps with the templates, creating a possibility that by combining constraints on templates, the overall constraint will shed more light on the underlying theory than any one constraint alone. We find it can also be advantageous to look for signals with a new, distinct template, beyond the three standard ones, that could help constrain models more efficiently; we explored the potential for using $2\mathcal{K}_3 - \mathcal{K}_6$ in this context.

In this initial study we use somewhat idealized assumptions focusing on the effects of cosmic variance and Gaussian noise from a homogeneous sky coverage. We recognize the rich potential for further study to other basis sets, that better characterize sharp or oscillatory features in bispectra, the presence of isocurvature modes, and stronger deviations from scale-invariance. To confidently attribute a primordial source to any measured non-Gaussianity one would also want to fully account for contributions from astrophysical and instrumental sources, including gravitational lensing, inhomogeneous sky coverage, and secondary anisotropies from astrophysical foregrounds. There is also the substantial question of how large-scale structure measurements, with sensitivity to the squeezed limit, can complement the CMB data in constraining these general shapes, as well as whether 4-point statistics and checks of non-Gaussian consistency ansatzes can play a role. We are tackling some of these intriguing issues in work in preparation.

CHAPTER 3

NON-GAUSSIAN SHAPE DISCRIMINATION WITH SPECTROSCOPIC GALAXY SURVEYS

3.1 Introduction

The physical mechanism behind inflation still remains an open question, despite the successes of the inflationary paradigm. An abundance of physically-motivated theories of inflation exist that employ very different physics to drive inflation, ranging from the simplest models with a single slowly-rolling inflaton field, to those with multiple fields, non-canonical kinetic terms, or originating from excited initial states, to broadly name a few. In general, however, each model predicts a distinct 'thumbprint' signature in the statistics of the generated primordial perturbations due to the particular self-interactions of the inflationary field(s) that give rise to a scale-dependence in deviations from Gaussianity in tracers of the dark matter density and Cosmic Microwave Background (CMB) temperature and polarization perturbations.

Recent results from the Planck survey of the CMB have put the strongest constraints yet on the potentially non-Gaussian statistics of primordial perturbations [116, 117] from the imprint on the observed multipole bispectrum. The primordial signal appears as changes in the dark matter halo bias and halo mass function at later cosmological epochs. By exploiting how galaxies are distributed with respect to these halos, we can extract shape constraints from galaxy surveys. Upcoming experiments, such as the Large Synoptic Survey Telescope (LSST) [118], Euclid [119], and DESI [7], will capture the statistics of the evolved late-time perturbations through the distribution of large-scale

structure (LSS). These probe non-Gaussianity, potentially even more strongly than Planck (e.g. [120, 121, 122]). Furthermore, the CMB and LSS observables probe the primordial fluctuations differently, offering the enticing possibility of using them together as complementary data sets to strengthen existing constraints (e.g. [123]).

The simplest way for a distribution to be non-Gaussian is to have a non-zero 3-point correlator, i.e. bispectrum or 'shape'. Typically the amplitude of primordial N-point correlators grows smaller as N increases, so that the bispectrum constraints are the principal N-point non-Gaussian correlation statistic investigated. An exciting realization in inflationary theory is that discerning the scale-dependence, or 'shape', of the bispectrum could provide a direct insight into the inflationary mechanism, through how non-Gaussianity is generated [16, 108]. Taking a specific bispectrum template and comparing it to data is a sensible approach to assess if the theory is consistent, or if it possesses new properties that might be observable with appropriately designed future surveys. Three templates have been studied extensively: the local shape that arises in multi-field inflation [35, 36, 34], and the equilateral [37] and orthogonal shapes [38] that derive from non-trivial kinetic terms in the inflationary action.

Recent theoretical developments have shown there are a far wider population of bispectrum shapes, including those from fast-roll inflation [39, 40, 41, 42, 43], quasi-single field inflation [44, 45], warm inflation [46, 47], and non-Bunch-Davies or excited initial states [39, 49, 50, 51, 124, 125]. There are also hybrids of multi-field and non-slow-roll models [52, 53, 54], the inclusion of isocurvature modes in the non-Gaussian correlations [55, 56, 57], and models motivated by string theory [100, 126, 127]. The large number of alternative theories motivates

looking for distinctive characteristics with which to classify bispectra and reconstruct what we can know about the underlying theory in an unbiased way from the data. One such characteristic in particular has gained prominence in the literature: while the bispectra can have very different shapes, meaning their signal is weighted towards different configurations of the three wavenumbers in (Fourier) k-space, their properties in the 'squeezed limit' when one of the three length scales contributing to the 3-point function becomes much larger than the other two, can signal whether inflation is derived from a single-field or multi-field model or from a non-Bunch-Davies vacuum.

For a preferred model with a particular bispectrum $B_\Phi(k_1, k_2, k_3)$, one can forecast or measure a constraint on the amplitude f_{NL} of the model's shape at a particular pivot value of (k_1, k_2, k_3). Then, under the assumption that this is the right model, we can extend the fixed shape to infer the amplitude of the bispectrum at all other k-configurations. Constraints obtained in this way thus assume a strong theoretical prior, as they do not account for our uncertainty in knowing which model is the right one. In contrast, an alternative and more conservative way of getting constraints on inflation is to assume less theoretical bias, and attempt to use the data itself to reconstruct the general primordial shape and its potentially complicated k-dependence. If there is a detection of a non-zero f_{NL} given one template, e.g. if $f_{NL}^{local} \neq 0$, this could either indicate evidence of local non-Gaussianity as might arise in multifield inflation, or that another inflationary model has generated non-Gaussianity whose signal leaks into, or is partially described by, the local template. These dual questions of distinguishing between model shapes and model-independent constraints on shapes have been considered recently in the context of CMB bispectrum data by WMAP9 [20] and Planck [117] .

In this paper we explore the ability of upcoming galaxy surveys and their measured galaxy power spectra, to obtain general shape constraints and distinguish between shapes, alone and in combination with CMB data. To do this, we consider the effects that different shapes have on dark matter halo statistics through the halo bias and halo mass function, with a particular focus on which k-configurations of the primordial shape are primarily probed by the halo statistics. By implementing the halo model, we forecast Euclid- and DESI-like constraints on f_{NL} for a variety of specific templates, as well as consider the distinguishability and degeneracies between combinations of these shapes. This work builds on prior work considering general shape constraints arising from CMB data [20, 17, 117], and through using basis reconstruction to study general non-Gaussian shapes, provides an alternative approach to other studies of non-Gaussian halo bias constraints in the literature, that used scale dependent modifications to the local template [61, 62] and theoretically motivated two-parameter models [63, 94], to constrain the squeezed limit properties of bispectra.

The paper is organized as follows: Section 3.2 describes the formalism behind our analysis, including an outline of the basis function approach used to describe general non-Gaussian shapes, the halo bias and halo mass function modeling, and the Fisher analysis. Our findings, centered on the potential to use LSS data to constrain shapes beyond the squeezed-limit, are presented in Section 3.3 and in Section 4.5 we summarize our conclusions and their implications.

3.2 Formalism

This section outlines the main theoretical assumptions used in the analysis. Section 3.2.1 outlines the approach we use to describe general bispectrum shapes, and 3.2.2 describes how the halo bias and halo mass function are altered in the presence of primordial non-Gaussianity. Sections 3.2.3 and 3.2.4 respectively summarize how halo statistics are translated into halo and galaxy power spectra within the framework of the halo model. Finally, 3.2.5 outlines the Fisher matrix formalism we use to forecast constraints for upcoming galaxy surveys, such as Euclid and DESI.

3.2.1 Bispectrum shape families and templates

Theoretical predictions for the primordial fluctuations created by different inflationary models are typically given by N-point correlators, such as bispectra, and their specific k-dependence. A primordial bispectrum, B_ζ, is defined through the 3-point correlation function of (comoving) curvature perturbations in Fourier space,

$$\langle \zeta(\vec{k}_1)\zeta(\vec{k}_2)\zeta(\vec{k}_3)\rangle = (2\pi)^3\delta(\vec{k}_1+\vec{k}_2+\vec{k}_3)B_\zeta(k_1,k_2,k_3). \tag{3.1}$$

Each model's bispectrum can then be parametrized in terms of an amplitude, f_{NL}, and shape, $S_\zeta(k_1,k_2,k_3)$, that are related by a normalization factor, N, chosen such that the shape is equal to unity at a pivot scale $\vec{k}_0$,

$$(k_1k_2k_3)^2 B_\zeta(k_1,k_2,k_3) = N f_{NL} S_\zeta(k_1,k_2,k_3), \tag{3.2}$$

where $N \equiv 72\pi^4\Delta_\zeta^4(k_0)/5$, and $\Delta_\zeta^2(k)$ is the dimensionless primordial power spectrum.[1]

Each model's shape is unique, but many shapes exhibit similar features, motivating us to roughly categorize shapes into families of shapes that peak when k_1, k_2, and k_3 form squeezed, equilateral, flattened, etc. triangles. Thus some of the most widely studied shapes are templates that approximate the shapes within each family, e.g. the local, equilateral, orthogonal, and enfolded templates, where the degree of similarity between shapes, or between true shapes and templates, can be quantified by a inner product statistic [16]. These templates are meant to be used as indicators of more complex shapes that arise from models with, for example, multiple fields (local template), higher derivative terms (equilateral and orthogonal templates), or non-trivial vacuum states (enfolded template).

Going beyond shapes characterized by a single template that is chosen to reflect a preferred model, a model-independent way to parametrize a general shape is to choose a basis of shape functions and assume that the underlying true shape is well-approximated by a linear combination of the basis, such as in [16, 17]. We utilize the separable $\{\mathcal{K}_n\}$ basis set introduced in [17]. Each basis function is defined as

$$\mathcal{K}_n(k_1,k_2,k_3) \equiv \frac{1}{\mathcal{N}_n k_0^{2(n_s-1)}}\left[k_1^{p'}k_2^{r'}k_3^{s'} + \{prs\}\text{perms}\right], \tag{3.3}$$

where $\mathcal{N}_n$ is the number of distinct permutations of $\{p,r,s\}$, and p', q' and r' are defined as

$$p' \equiv 2 + \frac{(p-2)(4-n_s)}{3}. \tag{3.4}$$

[1] In terms of the fluctuations in the potential in Chapter 2, $\Phi = \frac{3}{5}\zeta$. We note that the normalization, N, is different, depending on whether it corresponds to the bispectrum of ζ or Φ.

To create near scale-invariance, the powers must satisfy $p + r + s = 0$.

Bispectrum separability is not strictly necessary to efficiently compute the resulting halo bias and halo mass function corrections using 3.15, 3.14, and 3.11 as we describe below. It is, however, a useful approach if one wants to consistently consider constraints on general shapes, and mesh together the squeezed limit and broader shape properties, from LSS and the CMB analyses, as separability is essential for the latter. Separability also allows for computationally efficient simulations that verify our theoretical understanding of how different LSS probes depend on non-Gaussianity, e.g. for nonlinear power spectra and halo mass functions [128], halo bias [64], and matter bispectra and higher-order correlators [77].

The $\{\mathcal{K}_n\}$ basis can be used to describe many nearly scale-invariant shapes in the literature precisely using only a relatively small set of 7 basis functions,

$$(k_1k_2k_3)^2 B_\zeta(k_1, k_2, k_3) = N S_\zeta(k_1, k_2, k_3) \approx N \sum_{n=0}^{6} \alpha_n \mathcal{K}_n(k_1, k_2, k_3). \tag{3.5}$$

The local, equilateral, orthogonal, and enfolded templates can be expressed exactly in terms of this basis, as shown in Table 3.1. Shapes constructed from the first 3 basis functions, $\{\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2\}$, can alternatively be written in terms of linear combinations of the local, equilateral, and orthogonal templates, which we will refer to as the LEO basis. For example, the enfolded template can be rewritten in terms of the LEO basis as $S_{enf} = \frac{1}{2}(S_{orth} - S_{equil})$.

As mentioned in [17], one advantage of the $\mathcal{K}_n$ basis is its ability to recover the correct squeezed limits of shapes. This is especially pertinent for this work, because the similarity between a shape and its template approximation may be more critical for characterizing the model's predictions for the halo bias and halo mass function, since these probes tend to correspond more to the shape's

squeezed limit, compared to the CMB bispectrum. Motivated by this concern, [38] constructed an improved orthogonal template, called $S_{orth(2)}$, which in addition to having the same overall features as the original S_{orth}, also has the same behavior in the squeezed limit as the primordial shape on which the original orthogonal template was based. Similarly, [103] constructed a template peaking on enfolded triangles with a vanishing squeezed limit, which we call $S_{enf(2)}$, but arises from higher-derivative operators of single-field models. Both $S_{orth(2)}$ and $S_{enf(2)}$ can be written exactly in terms of the $\mathcal{K}_n$ basis, as we show in Table 3.1.

Furthermore, combinations of the basis modes can create a set of irreducible shapes with the same divergence properties in the squeezed limit [17]. For example, we can consider modes that individually diverge by up to $1/k^2$ in the squeezed limit, $\{\mathcal{K}_0 - \mathcal{K}_6\}$, and, construct three irreducible templates, in addition to the equilateral template, S_{equil}: $\{\mathcal{K}_0 + 3\mathcal{K}_3 - 3\mathcal{K}_4, \mathcal{K}_2 + 2\mathcal{K}_3 - 2\mathcal{K}_5, 2\mathcal{K}_3 - \mathcal{K}_6\}$ that each vanish in the squeezed limit, but have significant differences from S_{equil} away from the squeezed limit.

A general template that vanishes in the squeezed limit can be reconstructed out of linear combinations of these four templates. An example of this is the Self-Ordering Scalar Fields (SOSF) model [2], which while vanishing in the squeezed limit, like the equilateral shape, has principal power in the 'aligned' configuration where $k_1 \sim 2k_2 \sim 2k_3$. In [2], a fitting function for S_{SOSF} is provided that fits the numerical results to a few percent. We reconstruct a template, included in Table 3.1, that has the maximal cosine (> 0.99) with the fitting function in [2], and vanishes in the squeezed limit. We use the fitting function for the LSS analysis, but use the separable template in Table 3.1 to allow efficient computation of the CMB constraints.

Shape	Template in terms of the $\mathcal{K}_n$ basis
S_{local}	$\mathcal{K}_2$
S_{equil}	$-2\mathcal{K}_0 + 6\mathcal{K}_1 - 3\mathcal{K}_2$
S_{orth}	$-8\mathcal{K}_0 + 18\mathcal{K}_1 - 9\mathcal{K}_2$
S_{enf}	$-3\mathcal{K}_0 + 6\mathcal{K}_1 - 3\mathcal{K}_2$
$S_{orth(2)}$	$(1+p)S_{equil} - p\left(\frac{2}{9}\mathcal{K}_0 + \frac{8}{3}\mathcal{K}_1 - 2\mathcal{K}_2 + \frac{20}{9}\mathcal{K}_3 - \frac{10}{3}\mathcal{K}_4 + \frac{4}{3}\mathcal{K}_5 - \frac{1}{9}\mathcal{K}_6\right)$
$S_{enf(2)}$	$(1+\alpha)S_{equil} - \alpha\left(\frac{6}{5}\mathcal{K}_0 + \frac{16}{5}\mathcal{K}_3 - \frac{18}{5}\mathcal{K}_4 + \frac{1}{5}\mathcal{K}_6\right)$
S_{SOSF}	$22.6\,S_{equil} - 5.98\,(\mathcal{K}_0 + 3\mathcal{K}_3 - 3\mathcal{K}_4) - 29.5\,(\mathcal{K}_2 + 2\mathcal{K}_3 - 2\mathcal{K}_5) + 13.9\,(2\mathcal{K}_3 - \mathcal{K}_6)$
S_1	$2\mathcal{K}_4 - \mathcal{K}_6$

Table 3.1: The $\mathcal{K}_n$ expansion for a variety of templates discussed in the literature. The variables p and α, used in $S_{orth(2)}$ and $S_{enf(2)}$, are chosen to maximize the template's fit to the physical shape. The separable template for S_{SOSF} is used for the CMB analysis, and is constructed from combinations of the basis functions that vanish in the squeezed limits, with coefficients that maximize the cosine with the fitting function given in [2], giving a cosine of > 0.99.

We also consider non-Gaussian shapes that can arise from models with large-scale magnetic fields (which could be generated by vector perturbations during inflation), vector fields that are coupled to the inflaton, and the solid (or elastic) inflation model [129]. The basis describing these, $\{S_0, S_1, S_2\}$ in [129], corresponds exactly to a linear combination of $\mathcal{K}_n$ functions, and represents a theoretically-motivated generalization of the local shape, analogous to studies of scale-dependent f_{NL} [62]. In this analysis we include consideration of the S_1 template used in describing these models, as an example of a shape with an anisotropic squeezed limit, in which their properties differ depending on from which direction, in k-configuration space, the squeezed limit is approached.

3.2.2 Non-Gaussian halo statistics

In general, primordial non-Gaussianity changes the power spectrum of dark matter halos by coupling the local power spectrum to the local long-wavelength matter overdensity. This effect was first discovered in N-body simulations of local non-Gaussianity [58], and since then there has been both theoretical and numerical work towards understanding the effect more precisely and how it extends to other types of non-Gaussianity (e.g. [128, 64, 8, 130, 131, 132]). We adopt the framework based on the peak-background split approach [132] that has shown good agreement with N-body simulations over the range of length scales and halo masses relevant to this work [128, 64]. Here we review how general bispectrum shapes can change the statistics of halos through its effect on the halo bias and halo mass function within this analytic model.

Given the primordial gravitational potential, $\Phi(\vec{k})$, where $\Phi = 3\zeta/5$, the dark matter density contrast $\delta(\vec{k})$ at redshift z is described by

$$\delta(\vec{k}, z) \quad = \quad M(k, z)\Phi(\vec{k}), \tag{3.6}$$

where

$$M(k, z) \quad \equiv \quad \frac{2D(z)k^2T(k)}{3\Omega_{m,0}H_0^2}, \tag{3.7}$$

$D(z)$ is the growth factor normalized to 1 at $z = 0$, and $T(k)$ is the matter transfer function. $M(k, z)$ is straightforward to compute using publicly available codes, such as CAMB[2]. For simpler notation, from now on we will drop the explicit z-dependence in our equations.

In the presence of non-Gaussianity, the matter fluctuations acquire a non-zero skewness, S_3, which is an integrated measure of the smoothed matter bis-

[2]http://camb.info

pectrum,

$$S_3(m) \equiv \frac{\langle \delta_R^3 \rangle}{\sigma_R^4}, \tag{3.8}$$

where σ_R^2 is the variance of the smoothed density contrast,

$$\sigma_R^2 \equiv \int \frac{d^3k}{(2\pi)^3} M(k)^2 W_R(k)^2 P_\Phi(k), \tag{3.9}$$

and the third moment is

$$\begin{aligned} \langle \delta_R^3 \rangle &\equiv \int d^3k\, d^3q\, M(k)M(q)M(|\vec{k}-\vec{q}|)W_R(k)W_R(q)W_R(|\vec{k}-\vec{q}|)B_\Phi(k,q,|\vec{k}-\vec{q}|) \\ &= \frac{2}{(2\pi)^4} \int d\ln k\, k^3 M(k) W_R(k) \int d\ln q\, q^3 M(q) W_R(q) \\ &\qquad \times \int d\mu M(k_\mu) W_R(k_\mu) B_\Phi(k,q,k_\mu). \end{aligned} \tag{3.10}$$

$W_R(k)$ is the Fourier transform of a spherical top-hat function of radius R, normalized such that $\int dx W_R(x) = 1$, where R corresponds to an effective radius for the halo mass under consideration, with $R = (3M_{halo}/4\pi\rho_m)^{1/3}$, and ρ_m is the matter density today. In what follows, we choose to work with the redshift-independent reduced skewness, $\sigma_R S_3(m)$.

The skewness induces changes in the halo mass function, dn/dm, that gives the number density, n, of halos of mass m, and we parametrize the deviation from what the halo mass function would be in the Gaussian case through the factor R_{NG} [59], defined as

$$R_{NG} \equiv \frac{dn_{NG}/dm}{dn_G/dm} = 1 + \frac{1}{6}\sigma_R S_3 \left(\nu^3 - 3\nu\right) - \frac{1}{6}\frac{d\sigma_R S_3}{d\ln\nu}\left(\nu - \frac{1}{\nu}\right), \tag{3.11}$$

where $\nu \equiv \delta_c/\sigma_R$, and we assume $\delta_c = 1.686$ is the linear threshold for spherical collapse.

In addition to changes in the halo mass function, non-Gaussianity also alters how the halos are distributed with respect to the underlying dark matter fluctuations. At the linear level, the halo density δ_h traces the matter fluctuations with

a bias factor that depends on the halo mass m and wavenumber k,

$$\delta_h(m,\vec{k}) \simeq b_{halo}(m,k)\delta(\vec{k}). \tag{3.12}$$

The linear halo bias can then be split up into three contributions: a Gaussian scale-independent term, a non-Gaussian scale-independent term, and a non-Gaussian scale-dependent term,

$$b_{halo}(m,k) = b_G(m) + \Delta b_{si}(m,f_{NL}) + \Delta b_{sd}(m,k,f_{NL}). \tag{3.13}$$

The scale-independent bias correction is derived from R_{NG} as [133]

$$\begin{aligned}\Delta b_{si}(m,f_{NL}) &\equiv -\frac{1}{\sigma_R}\frac{\partial \ln R_{NG}}{\partial \nu} \\ &= -\frac{1}{6\sigma_R}\frac{1}{R_{NG}} \\ &\quad\times\left[3\sigma_R S_3\left(\nu^2-1\right)+\frac{d\sigma_R S_3}{d\ln\nu}\left(\nu^2-4-\frac{1}{\nu^2}\right)-\frac{d^2\sigma_R S_3}{d\ln\nu^2}\left(1-\frac{1}{\nu^2}\right)\right].\end{aligned} \tag{3.14}$$

The scale-dependent term $\Delta b_{sd}(m,k)$ is given by [132]

$$\Delta b_{sd}(m,k,f_{NL}) = \frac{\delta_c[b_G(m)+\Delta b_{si}-1]}{2M(k)}\frac{I_{21}(m,k)}{\sigma_R^2} + \frac{1}{M(k)}\partial_{\ln\sigma_R^2}\left[\frac{I_{21}(m,k)}{\sigma_R^2}\right]. \tag{3.15}$$

I_{21} is defined and calculated as

$$\begin{aligned}I_{21}(m,k) &\equiv \tfrac{1}{P_\Phi(k)}\int\tfrac{d^3q}{(2\pi)^3}\,M(q)W_R(q)M(|\vec{k}-\vec{q}|)W_R(|\vec{k}-\vec{q}|)B_\Phi(k,q,|\vec{k}-\vec{q}|) \\ &= \tfrac{1}{(2\pi)^2P_\Phi(k)}\int d\ln q\; q^3M(q)W_R(q)\int_{-1}^{1}d\mu\; M(k_\mu)W_R(k_\mu)B_\Phi(k,q,k_\mu),\end{aligned} \tag{3.16}$$

where by construction, $\vec{k_\mu} = \vec{k}-\vec{q}$ such that $k_\mu^2 = k^2+q^2-2kq\mu$, and μ is the cosine of the angle between $\vec{q}$ and $\vec{k}$. This multidimensional integral and others in this work are computed using Cuba, a publicly available code for numerical integration of multidimensional integrals[3].

In the local case, I_{21}/σ_R^2, the first term in the scale-dependent bias, tends towards a constant value of $4f_{NL}$ on large scales, such that the contribution from

[3] http://www.feynarts.de/cuba/

the second, derivative term, $\partial_{\ln \sigma_R^2}[I_{21}(m,k)/\sigma_R^2]$ is negligible. However, in general the inclusion of both terms is important: the first and second terms typically have opposing signs and partially cancel each other on scales above a few Mpc. For the templates where the two terms are comparable, such as in the equilateral and orthogonal cases, the large scale, $k \to 0$ scaling behavior and small scale properties are determined by the cancellation. The relative weight of each term changes with the halo mass, and gives rise to mass-dependent biases.

When computing Δb_{sd} on large scales, the integral over μ in 3.16 averages over squeezed configurations. For isotropic shapes, the absence of μ dependence in the leading term in the squeezed limit expansion of the shape in powers of k/q gives a trivial mapping of the large scale scale-dependence of the large-scale halo bias. For shapes for which the squeezed limit is anisotropic in μ, the μ integral can alter the sign and amplitude of the leading k scaling in the bias, or make it vanish. Thus interpreting the halo bias data using just large scale data, and presuming that the primordial shape is isotropic, could potentially lead to incorrect or incomplete conclusions about the shape determination, and miss an important sector of the theory space.

3.2.3 The halo model

In this section, we first briefly summarize our implementation of the halo model for computing non-linear matter and galaxy power spectra (see [134] for a detailed review of the halo model), and then describe our resulting power spectra from different non-Gaussian shapes.

In the halo model, the halo matter power spectrum P_h is the sum of cor-

relations between separated halos, P^{2h}, and within the same halo, P^{1h}, $P_h(k) = P^{1h}(k) + P^{2h}(k)$, with

$$P^{1h}(k) = \int dm \frac{dn}{dm} \left(\frac{m}{\bar{\rho}}\right)^2 |u(k|m)|^2, \tag{3.17}$$

$$P^{2h}(k) = \left[\int dm \frac{dn}{dm} \left(\frac{m}{\bar{\rho}}\right) u(k|m) b_h(m,k)\right]^2 P_{linear}(k). \tag{3.18}$$

Here P_{linear} is the linear dark matter spectrum, dn/dm is the halo mass function, and b_h is the halo bias. In the fiducial Gaussian case, we compute dn^G/dm and b_G using the numerical fits in [135]. We include non-Gaussianity in the halo mass function using the R_{NG} parameter, and in the halo bias, parametrized by Δb_{si} and Δb_{sd}, as described in section 3.2.2. $u(k|m)$ is the Fourier transform of the halo density profile that describes the distribution of matter within each halo. We assume the Navarro-Frenk-White (NFW) profile [136, 137], which has two free parameters that can be chosen as a scale radius, r_s, and a corresponding matter density, ρ_s. However, it is also common to choose the two free parameters to be how the virial radius, R_{vir}, of the halo is defined, and how the concentration of each halo, c, depends on the halo mass and redshift. Here we define R_{vir} such that it contains an average density that is $\Delta_{vir} = 200$ times the average matter density of the universe, and use the concentration from [138]. Our choices of these inputs to the halo model calculation were made to facilitate comparisons between our results and those of previous forecasts. In particular, we have used the same Gaussian halo mass function and halo density profile that were noted by [6] to provide good agreement with simulations.

Implementing the halo model requires two additional constraints to be imposed by hand: (1) that all matter is contained in halos, and (2) that the calculated Gaussian matter power spectrum on large scales matches the linear expectation, $P_{linear}(k)$. The numerical details of enforcing these constraints are

described in [139].

We include a note of caution, that as one proceeds to significantly non-linear scales the halo model needs to be carefully calibrated against N-body simulations, with an increasing number of parameters, including improvements in the concentration parameter fits. However, in this work we confine our analysis to linear and mildly nonlinear scales, $k \lesssim 0.25h/Mpc$, on which the halo model and N-body simulations have been found to agree reasonably well, e.g. [140, 141].

3.2.4 The galaxy bias model

One can compute the galaxy power spectrum, $P_{gal}(k)$, by using theoretical and empirical prescriptions for how galaxies are arranged within halos of a given mass. With assumptions on the distribution of galaxies within the halo, u_{gal}, and the moments of the statistical distribution of the number of galaxies with a halo of a given mass, the Halo Occupation Distribution (HOD), $\left\langle N_{gal}|m \right\rangle$, the galaxy power spectrum can be written:

$$P^{1h}_{gal}(k,z) = \int dm \frac{dn}{dm} \frac{\langle N_{gal}(N_{gal}-1)|m\rangle}{\bar{n}^2_{gal}} |u_{gal}(k|m)|^p, \tag{3.19}$$

$$P^{2h}_{gal}(k,z) = \left[\int dm \frac{dn}{dm} \frac{\langle N_{gal}|m\rangle}{\bar{n}_{gal}} u_{gal}(k|m) b_h(m,k)\right]^2 P_{linear}(k,z), \tag{3.20}$$

where

$$\bar{n}_{gal}(z) = \int dm \frac{dn}{dm} \langle N_{gal}|m\rangle. \tag{3.21}$$

We make the reasonable assumption that galaxies are distributed within halos in a similar way that dark matter particles are, i.e. $u_{gal}(k|m) = u(k|m)$, and, as discussed in [134], set $p = 1$ or 2 depending on whether $\langle N_{gal}(N_{gal}-1)|m\rangle < 1$ or ≥ 1. We note that using the halo model-derived galaxy power spectrum as our

observable does not require measurements of individual halo masses; instead we have adopted a semi-analytic model for the redshift independent moments $\langle N_{gal}|m\rangle$ and $\langle N_{gal}(N_{gal}-1)|m\rangle$ of the halo occupation distribution [142], as used in [121], that assumes a fixed relationship between a host halo's mass and the number of galaxies that reside within the halo.

As the mapping between galaxies and the underlying dark matter, the galaxy bias, $b_{gal}(k,z) = \sqrt{P_{gal}/P_h}$, can be probed in different ways – observationally, theoretically, and with simulations. In the large scale limit, $u_{gal} \approx 1$ and the 2-halo term dominates so that

$$b_{gal}(z) \approx \int dm \frac{dn}{dm} \frac{\langle N_{gal}|m\rangle}{\bar{n}_{gal}} b_h(m, k \ll 1). \tag{3.22}$$

As our fiducial, Gaussian, galaxy bias, we use the values from semi-analytic models of galaxy formation from Orsi et al. [5], which are in broad agreement with the range of fiducial Gaussian bias values that are taken in the literature for different comparable forecasts. Figure 3.1 gives a comparison of the halo model Gaussian galaxy bias evaluated for halos consistent with emission line galaxies (ELGs) and the Orsi et al. paper, along with others assumed in the literature.

The galaxy bias can be measured most directly if the galaxies and matter are observed in overlapping fields, for instance, with galaxy-lensing cross correlations. Such constraints have been analyzed while assuming the Gaussian case [143, 144, 145, 146, 147, 148], as well as for simultaneously constraining templates of non-Gaussianity [149, 150, 6, 151]. In observations without a lensing component, the matter field is less directly inferred from the data, due to the complex processes of galaxy formation. However, we know from theoretical and observational considerations that galaxy formation is strongly correlated with their host halos, which allows for full encapsulation of complex formation

processes into a HOD with a few free parameters describing the number and spatial distribution of galaxies inhabiting a halo of a given mass [152, 153]. The HOD allows galaxy survey measurements to be more easily related to the underlying halos, the statistics of which are more directly predicted by theoretical cosmology-dependent parameters. With the HOD as a tool to bridge the gap between galaxies and halos, other probes beyond galaxy-lensing correlations have been employed to infer the galaxy bias. One such technique is to employ dark matter simulations and a halo finder algorithm, which can be done for both Gaussian and non-Gaussian initial conditions, to see how halo power spectrum and bispectrum statistics are altered by primordial shapes [128, 64, 154]. Given the output of such simulations of halos, the HOD can further these predictions to the level of galaxies, allowing for easier comparison of theories with data. For example, higher-order galaxy correlations, such as the bispectrum of galaxies, can also constraint galaxy bias parameters in the Gaussian case [155, 156, 157], and the bias parameters simultaneously with the f_{NL} parameters in non-Gaussian cosmologies [158]. While a detailed description and implementation of such techniques is outside the scope of this paper, the combination of these studies paints a picture of steady improvements in the theoretical understanding and observational measurements of the galaxy bias, and understanding how to extract non-Gaussian constraints simultaneously.

3.2.5 Fisher matrix approach for LSS and CMB

We conduct a Fisher matrix analysis as a preliminary tool to estimate potential constraints on bispectrum shapes with upcoming surveys. This approach has been used in a number of previous analyses that also aim to extract constraints

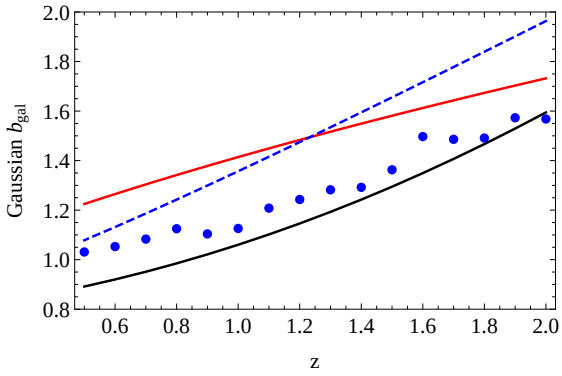


Figure 3.1: Fiducial Gaussian galaxy biases in this work from [5] [blue circles] and computed from the halo model we consider for ELGs [black]. Other bias parametrizations in the literature are shown for comparison: $b^G_{gal}(z) = \sqrt{1+z}$ [red] [6], and ELG bias $b^G_{gal}(z) = 0.84/D(z)$ [blue dashed] [7].

on primordial NG from the halo bias [6, 94, 62, 159, 160]. The analysis we present here assumes a spectroscopic galaxy survey from Euclid-like and DESI-like experiments, where the observed galaxy power spectrum can be used to infer the underlying dark matter distribution, which in turn contains traces of primordial non-Gaussianity. In doing so, however, we must account for the various astrophysical and observational effects that can change how the observed galaxies are related to the halo model's theoretically predicted galaxy power spectrum.

We model the relation between the theoretical galaxy (position) power spectrum we compute from the halo model, P_{gal}, and the one that is observed, P^{obs}_{gal}, as

$$P^{obs}_{gal}(k,\tilde{\mu},z) = \left[1 + \frac{f(z)\tilde{\mu}^2}{b_{gal}(k,z)}\right]^2 P_{gal}(k,z)e^{-\mathcal{D}(k,\tilde{\mu},z)} \tag{3.23}$$

where $\tilde{\mu}$ is the cosine of the angle between the wavenumber k and the line of sight, and should not be confused with μ in the definition of the non-Gaussian shape. The bracketed factor in 3.23 uses the Kaiser formula [161] to model the effect of redshift-space distortions (RSD), and we take $f \equiv \Omega_m(z)^{0.55}$ as the linear

growth rate. On mildly non-linear scales, the power spectrum's form is damped due to three contributions, separately parametrized, in Gaussian form, on the scales of interest: the redshift uncertainty of the survey ($\sigma_z/z = 0.001$), non-linear damping related to the BAO scale, and the finger-of-god effect. The last of these is the dominant one for our analysis. While a Gaussian finger-of-god damping term is commonly used, there are alternative fits in the literature, as discussed in [162, 163].

The set of parameters to be constrained, $\vec{\theta}$, includes the non-Gaussianity parameters we wish to constrain, as well as nuisance parameters that we wish to marginalize over. We allow for the potential for additional, external information about the galaxy bias in the analysis by allowing the Gaussian galaxy bias in each redshift to act as a nuisance parameter, that can be marginalized over with varying priors. For this analysis, our parameter set is $\vec{\theta} = \{f_{NL}; b^G_{gal}(z_1), ..., b^G_{gal}(z_N)\}$, where the Gaussian galaxy bias parameters, b^G_{gal}, for each of the N z-bins are included. We consider scenarios in which they are fully marginalized over, representing no additional information about the bias from external analyses, to perfect knowledge, with no marginalization, and gradations in between through the imposition of independent, Gaussian priors on each.

The Fisher matrix contribution in a given redshift slice, now including the sky fraction, is given by

$$\begin{aligned} F_{\alpha\beta}(z) &= \frac{f_{sky}}{8\pi^2}\int_{\ln k_{min}}^{\ln k_{max}} k^3\, dlnk \\ &\quad \times \int_{-1}^{1} d\tilde{\mu}\frac{\partial \ln P^{obs}_{gal}(k,\tilde{\mu},z)}{\partial\theta_\alpha}\frac{\partial \ln P^{obs}_{gal}(k,\tilde{\mu},z)}{\partial\theta_\beta}V_{eff}(k,\tilde{\mu},z), \end{aligned} \tag{3.24}$$

and evaluated at the fiducial cosmology, for which we use the best-fit cosmological parameters from the WMAP9 results [20], assuming no non-Gaussianity: $\Omega_b h^2 = 0.02264$, $\Omega_c h^2 = 0.1138$, $\Omega_\Lambda = 0.721$, $\Delta^2_{\mathcal{R}} = 2.41 \times 10^9$, $n_s = 0.972$, $\tau = 0.089$.

Previous work studying the potential degeneracies between f_{NL} and the standard cosmological parameters has found minimal impact on f_{NL} constraints due to uncertainties in the underlying cosmology, especially when the LSS constraints are combined with CMB priors from the CMB power spectrum [164, 6]. For this analysis, we consider a fixed cosmology and focus on degeneracies between the non-Gaussian and nuisance parameters.

The effective inverse data covariance is given by

$$V_{eff}(k,\tilde{\mu},z) = V_{survey}(z)\left[\frac{\bar{n}^{obs}_{gal}(z)P^{obs}_{gal}(k,\tilde{\mu},z)}{1+\bar{n}^{obs}_{gal}(z)P^{obs}_{gal}(k,\tilde{\mu},z)}\right]^2 \tag{3.25}$$

$$V_{survey}(z) = \frac{4\pi}{3}\left[\chi(z_{max})^3-\chi(z_{min})^3\right], \tag{3.26}$$

where χ is the comoving distance, $\bar{n}^{obs}_{gal}$ is the predicted number density of galaxies observed for the survey in the redshift bin, and z_{max} and z_{min} are the boundary redshifts for a given z-bin. To combine the constraints from multiple redshift slices and get the total Fisher error, we simply add the Fisher matrices from each redshift bin.

One could use the full 3D information within the survey by including correlations between different redshift bins, which would result in tighter constraints on cosmological parameters (e.g. [165]), chiefly due to the effects of gravitational lensing on the densities and parameters of interest. Though there is potential to glean more information by considering cross-bin corelations, here we adopt a conservative use of the data, and do not include them.

For the Euclid-like forecast, we consider 27 redshift bins from $0.7 < z < 2$, with bin width $\Delta z = 0.05$, where the number densities of galaxies within each bin are taken from [166] and $f_{sky} = 20,000$ deg^2. For the DESI-like forecast, the constraints are calculated separately for the ELG and LRG galaxy populations,

before being combined. The DESI ELG constraint uses 11 redshift bins from $0.6 < z < 1.7$ with $\Delta z = 0.1$, while the DESI LRG constraint uses 6 redshift bins from $0.6 < z < 1.2$ with $\Delta z = 0.1$. For DESI, the estimated number densities of LRG and ELG galaxies is as per Table 2.3 of the DESI Conceptual Design Report [167], and assumes a $14,000$ deg^2 survey.

The k-dependence of the inverse covariance matrix of $\ln P_{gal}^{obs}$ is proportional to $k^3 V_{eff}$ which scales roughly as $\propto k^3$. The mildly non-linear scales, where both the halo mass function and scale-dependent bias are important and sensitive to a broad range of primordial shape configurations, are potentially better measured than larger scales. The additional effect of the bias marginalization has to be taken into account, however, in order to assess the projected constraints.

Where the large scale structure constraints are combined with CMB constraints, we have followed the CMB Fisher calculation described in [17]. We summarize the relevant calculation here, and refer the reader to [17] for further details.

The CMB constraint on set of f_{NL} values, one for each primordial shape that is simultaneously constrained, from CMB bispectrum data can be forecasted by first computing the theoretical CMB bispectrum for each shape, and applying survey-specific modeling to the CMB bispectrum covariance. The full CMB Fisher matrix is then

$$F(f_{NL}^i, f_{NL}^j) = f_{sky} \sum_{abc,pqr} \sum_{\ell_1 \ell_2 \ell_3} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{abc\ (i)}}{\partial f_{NL}^i} (Cov^{-1})_{\ell_1 \ell_2 \ell_3}^{abc,pqr} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{pqr\ (j)}}{\partial f_{NL}^j}, \tag{3.27}$$

where $B_{\ell_1 \ell_2 \ell_3}$ are the angle-averaged CMB bispectra for the, $\{abc\}$ and $\{pqr\}$ run over all possible bispectra from combinations of temperature and E-mode polarization data (i.e. TTT, TTE, TET, etc.), i and j label the primordial shapes, and

f_{sky} is the fractional sky-coverage of the survey. The bispectrum covariance is a function of the CMB power spectra, C_ℓ, and includes modeling of the CMB survey's frequency channels' noise properties and beam profiles. For the particular CMB Fisher constraints discussed in this work, we have modeled the constraints from a Planck-like survey, and keep the fiducial background cosmology fixed and consistent with the LSS forecasts described earlier in this section.

3.3 Findings

In this section we present the major findings. Sections 3.3.1 and 3.3.2 describe, analytically and numerically, how the halo bias and halo mass function probe different regions of the primordial shape. Section 3.3.3 presents the forecasted constraints on fixed shapes, as well as constraints for the distinguishability between shapes and general k-configuration dependent constraints.

3.3.1 Squeezed limit behavior

On large scales where $k \to 0$, the halo bias in 3.15 and 3.16 is a weighted average over squeezed limit configurations of the primordial shape, i.e. configurations with $k/q \ll 1$ with μ a free parameter that is the cosine of the angle between $\vec{q}$ and $\vec{k}$. Table 3.2 shows the squeezed limit forms for the templates discussed in section 3.2.1, and illustrates in particular how the squeezed limits of different shapes may depend on the angle μ.

Many shapes discussed in the literature, such as the local, orthogonal and enfolded shapes, are isotropic, meaning the leading order term in the squeezed

Shape	Squeezed limit form to $O(k/q)$
S_{local}	$\frac{2}{3}\frac{q}{k} - \frac{\mu}{3} + \frac{1}{6}(1+3\mu^2)\frac{k}{q}$
S_{orth}	$-2 + 6(1-\mu^2)\frac{k}{q}$
S_{enf}	$-1 + 2(1-\mu^2)\frac{k}{q}$
S_{equil}	$2(1-\mu^2)\frac{k}{q}$
$S_{ortho(2)}$	$2\left(1-\mu^2-\alpha\left(\frac{5}{27}+\mu^2\right)\right)\frac{k}{q}$
$S_{enf(2)}$	$-2\left(1-\mu^2-\frac{16}{9}\alpha\right)\frac{k}{q}$
S_1	$\frac{2}{3}(1-4\mu^2) + \frac{8}{3}\mu(1-\mu^2)\frac{k}{q}$
S_{SOSF}	$13.3\frac{k}{q}$

Table 3.2: The squeezed limit expansion, in powers of k/q as $k/q \to 0$, for a variety of templates described in section 3.2.1.

limit has no μ dependence as $k/q \to 0$. For an isotropic shape, the squeezed limit is μ-independent, so the calculation of the scale-dependent halo bias on large scales using 3.16 can be simplified by choosing a fixed value for μ when evaluting the integrand. Typically the squeezed isosceles configuration is chosen, equivalent to taking $q = k_\mu$, such that $\mu = k/2q \to 0$, to derive a rule of thumb: for a primordial shape which scales like $\propto 1/k^n$ in the squeezed limit, the scale-dependent halo bias will be $\propto 1/k^{n+1}$. For example, in the local case which diverges as $S_{local} \propto 1/k$ in the squeezed limit, it is straightforward to show that $I_{21}(m)/\sigma_R^2 \to 4f_{NL}$ as $k \to 0$, such that for a fixed halo mass, the halo bias on large-scales is $\Delta b_{sd} \propto 1/M(k)$, resulting in a strongly scale-dependent bias, $\Delta b_{sd} \propto 1/k^2$.

Even though for isotropic shapes the large-scale behavior of Δb_{sd} has a clear implication for the shape's squeezed limit, the same is not true of anisotropic shapes, which may have an unusual variety of μ-dependent limits while still appearing degenerate with other isotropic or anisotropic shapes, due to the integration over μ in 3.16. Local-like anisotropic shapes were considered in [129], and the $\mathcal{K}_n$ basis is capable of describing more general anisotropic shapes that

are not necessarily local-like. Indeed, one can even employ the basis to purposely construct an anisotropic shape such that the averaging over squeezed configurations through the integration over μ in 3.16 produces a halo bias scaling which breaks the rule of thumb. For example, it is possible to construct an anisotropic shape which approaches a μ-dependent constant in the primordial squeezed limit, similarly to S_1, while still yielding a scale-independent large-scale halo bias.

This may be a point of concern if a scale-independent bias is confirmed through future surveys, since the single-field consistency relation states that single-field models of inflation should vanish in the primordial squeezed limit. Under the rule of thumb, a lack of a scale-dependent bias detection implies an inflationary model with a single-field. However, such data may still be consistent with general anisotropic models which do not satisfy the consistency relation, but still yield scale-independent halo biases.

Despite this possible degeneracy between the primordial shape's squeezed limit and the large-scale halo bias, information from the halo bias on smaller scales may help to distinguish otherwise degenerate inflationary models. Once potential μ-dependence is taken into account, the $k \to 0$ limit of 3.15 yields a good power-law approximation to a shape's bias on large scales, but in general as k increases towards smaller scales the bias takes on a more complex scale-dependence.

In this work, since we include both large and small scale power spectra in our Fisher forecasts, we use the full k-dependence of the bias from a numerical calculation of 3.15 and 3.16 for each k. In other analyses, small scale behavior has been approximated by extending the large scale-derived, asymptotic, ex-

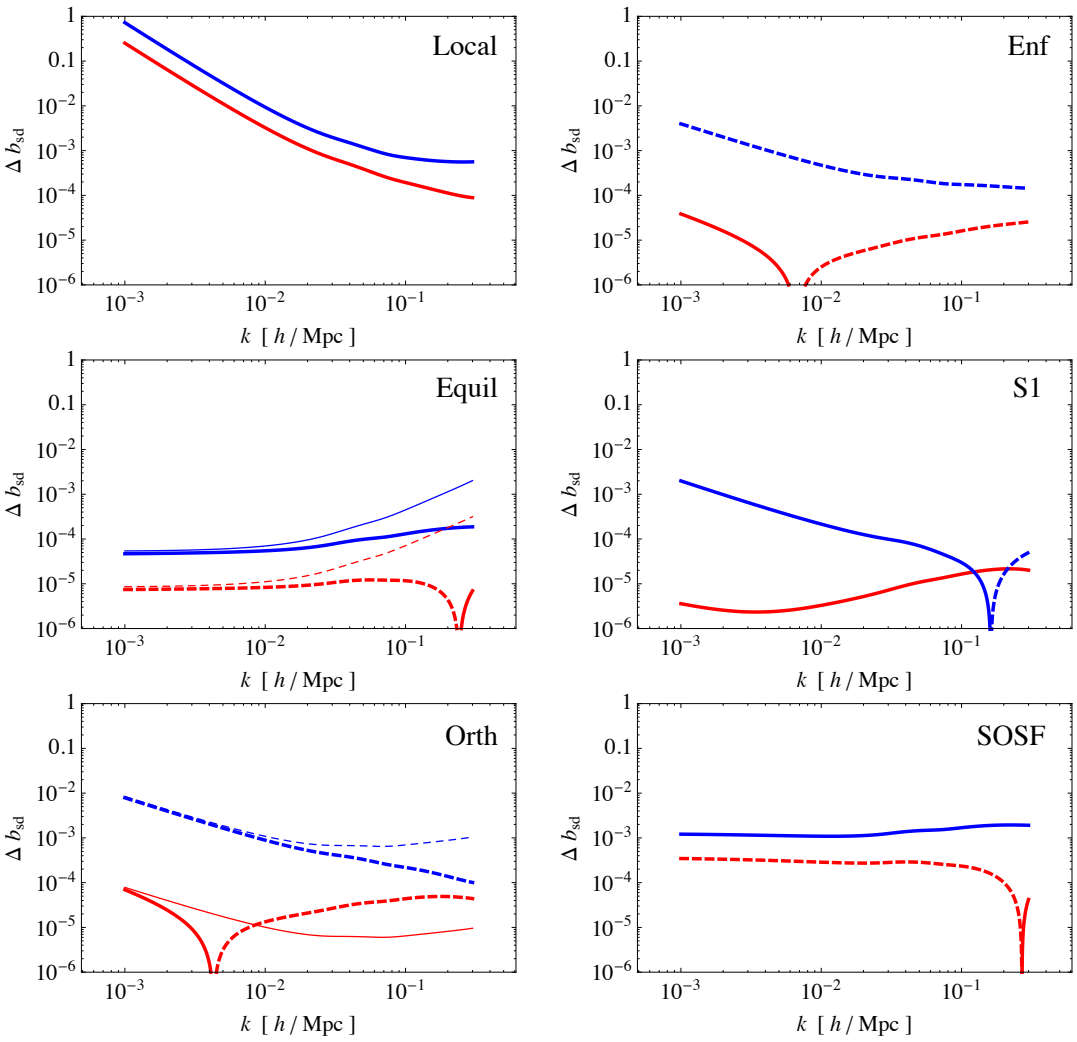


Figure 3.2: The left panels, from top to bottom, shows the scale-dependent halo bias using the analytic derivation [thick], as in (3.15), and the asymptotic form [8] if extrapolated to small scales [thin], for two halo masses, $10^{12.5} M_\odot/h$ [red] and $10^{13.5}\ M_\odot/h$ [blue], for the local, orthogonal, and equilateral shapes at $z = 1$ for $f_{NL} = 1$. The right panels, from top to bottom, show the scale-dependent bias, for the same halo masses, for the enfolded, anisotropic, and SOSF shapes, respectively. Dashed curves denote negative values.

pressions for the halo bias, using an analytically derived $k \to 0$ expression for $I_{21}(m,k)/\sigma_R^2$ and its derivative in evaluating 3.15. Figure 3.2 compares the fully

k-dependent and asymptotic forms of the bias for the local, equilateral, and orthogonal shapes and shows the differences between the two results on small scales for the orthogonal and equilateral templates. The right hand panels show the differences in large scale and small scale properties, and halo mass dependence, for the isotropic shapes, S_{enf} and S_{SOSF}, and anisotropic shape, S_1. The figure also shows that including information on mildly non-linear scales, ~ 10 Mpc, can reveal richer information about the shape than purely considering the large-scale bias: while two shapes can have the same scaling (up to a constant) in the squeezed limit, and thus they both have a scale-dependent bias $\propto 1/k^n$ on large scales, on smaller scales, and at different halo masses, their bias functions can differ.

We note that particularly in the equilateral template constraints, the use of an asymptotic approximation for the scale dependent bias, Δb_{sd}, that extrapolates the large scale behavior to small scales, leads to tighter constraints on f_{NL} [6]. This is due to an artificial increase in the amplitude of the bias on small scales relative to when the full analytic expression for the halo bias is used.

In summary, inferring the squeezed limit behavior of the true primordial bispectrum from observations of the large-scale halo bias becomes less straightforward if we allow for the possibility of general anisotropic shapes. However, exploiting the halo bias on smaller scales, using numerical calculations of the halo bias at all k values (rather than extrapolations of analytic derivations on large scales), may help to break the degeneracy between the large-scale halo bias and the shape's squeezed limit.

3.3.2 Configurations probed by the halo bias and mass function

To explore which configurations of the shape the scale-dependent bias probes best, and allow for the consideration of shapes that are anisotropic in μ as one approaches the squeezed limit, we define a weight function, w_{sd}, that encodes the relative importance to the bias per unit shape amplitude, S, at each $\ln q$ and μ, for a tracer of halo mass m and wavenumber k:

$$\Delta b_{sd}(m,k) = \int d\ln q \int d\mu \; w_{sd}(k, \ln q, \mu, m) S(k, q, k_\mu). \tag{3.28}$$

By finding the triangles, formed by the (k, q, k_μ) wavenumbers, at which the weight is largest, we can infer which triangular configurations of the primordial shape the halo bias is most sensitive to.

Figure 3.3 shows how the precise behavior of the weight function for the scale-dependent bias is dependent upon the halo mass m and the triangular configuration formed by the (k, q, k_μ) wavenumbers. On very large scales, $k \sim 0.001\,h$/Mpc, the bias primarily probes squeezed configurations for which $k/q \lesssim 0.01$, while on small and mildly non-linear scales of $k \sim 0.1\,h$/Mpc, the bias probes a wider range of configurations, including those that are not significantly squeezed. On intermediate scales (not shown), $k \sim 0.01\,h$/Mpc, the weight rises for configurations that are moderately squeezed, $k/q \sim 0.01 - 0.1$, with an overall amplitude comparable to the weight function of the $k \sim 0.1\,h$/Mpc slice, for the same halo mass. In all cases, the weights do not pick out a single, preferred value of μ, but are equally sensitive to all values of μ for a given value of k and q.

To show the mass-dependence in another way, in Figure 3.4 we show just the $\mu = 0$ slice of w_{sd} for $k = 0.001\,h$/Mpc and two halo masses. The figure illustrates

that larger halo masses have a weight function that is broader and larger in amplitude than for less massive halos, suggesting that having data sets specifically targeted to different halo mass bins may yield complementary information by weighting the halo bias to the primordial shape in different ways.

To determine the shape sensitivity of the halo mass function, we can define an analogous weight function, $w_{R_{NG}}$, for R_{NG}:

$$R_{NG}(m) = 1 + \int d\ln k \int d\ln q \int d\mu \; w_{R_{NG}}(k, \ln q, \mu, m) S(k, q, k_\mu). \tag{3.29}$$

We find this weight is highest and has the largest contributions to the total value of R_{NG} on the $k \sim 0.1\,h$/Mpc slice, shown in Figure 3.5. The figure illustrates that the weight has the largest amplitude for triangles that have $q \sim 0.1 - 1\,h$/Mpc. While $w_{R_{NG}}$ does probe squeezed triangles, the weight also extends to include other configurations that are not very squeezed. In this sense, R_{NG} probes similar configurations to $\Delta b_{sd}(k)$ on mildly non-linear scales, but with a different specific triangle configuration-dependence. Since R_{NG} is an integrated measure of the shape (as it integrates over k and has no remaining k-dependence), we might expect one can extract more information from $\Delta b_{sd}(k)$ than from R_{NG}. The non-Gaussian modification to the halo mass function and the scale-independent halo bias by the local, equilateral, and orthogonal templates is shown in Figure 3.6. The figure shows that corrections to R_{NG} and Δb_{si} from non-Gaussanity are more important for the most massive halos, and is typically at the few percent level.

In summary, the scale-dependent bias on large scales probes general squeezed configurations of the primordial shape, while the scale-dependent bias on small scales and the halo mass function are more sensitive to a broader range of configurations.

3.3.3 Prospective constraints from upcoming LSS and CMB surveys

Here we discuss the results of the Fisher matrix analysis, using the approach described in Section 3.2.5, establishing constraints on a variety of non-Gaussian templates in section 3.3.3, and the potential to distinguish between characteristics of different primordial shapes in section 3.3.3.

Constraints on specific templates

Constraints for the different templates discussed in Section 3.2.1 are shown in Table 3.3 for Euclid-like and DESI-like spectroscopic surveys. In general, templates with more strongly k-dependent halo biases on large scales are better measured by LSS, and it is clear that assumptions about the prior knowledge of the galaxy bias in each spectroscopic bin can have a major effect on the constraints. The bias assumption is most key for the least divergent shapes: constraints degrade most when removing any assumptions on galaxy bias for the shapes with the least scale-dependent bias. Below, we describe why this is the case, and what steps might be taken to improve constraints for the least divergent shapes.

For templates that have a scale-independent bias on large-scales, uncertainties in the Gaussian galaxy bias can easily mimic the effects of non-Gaussianity. Scale-dependent changes to the halo-model-derived P_{gal} on small scales, typically smaller than $k \sim 0.1\,h$/Mpc, then provide the sole constraints. If the large scale galaxy bias can be measured then these small scale effects can give rise to f_{NL} constraints that are competitive with the CMB as shown in Table 3.3.

$\sigma(f_{NL})$	CMB	Euclid-like LSS (+CMB) with $b^G_{gal}(z)$ prior							
		marg. b^G_{gal}		10% prior		1% prior		known b^G_{gal}	
S^{local}	3.5	8.4	(3.2)	8.4	(3.2)	8.3	(3.2)	7.1	(3.1)
S^{equil}	43	250	(42)	230	(42)	64	(36)	21	(19)
S^{orth}	19	51	(18)	51	(18)	40	(17)	18	(13)
S^{enf}	34	120	(33)	120	(32)	58	(29)	20	(17)
$S^{orth(2)}$, $\alpha = 8.52$	11	190	(11)	160	(11)	31	(10)	9.8	(7.3)
$S^{enf(2)}$, $\alpha = 0.60$	100	1700	(100)	1300	(100)	220	(93)	71	(58)
S^{SOSF}	10	110	(10)	85	(10)	14	(8.4)	4.6	(4.2)
S_1	23	200	(23)	200	(23)	200	(23)	180	(23)

$\sigma(f_{NL})$	CMB	DESI-like LSS (+CMB) with $b^G_{gal}(z)$ prior							
		marg. b^G_{gal}		10% prior		1% prior		known b^G_{gal}	
S^{local}	3.5	13	(3.4)	13	(3.4)	12	(3.4)	7.4	(3.2)
S^{equil}	43	120	(40)	120	(40)	81	(38)	21	(19)
S^{orth}	19	52	(18)	52	(18)	46	(18)	19	(13)
S^{enf}	34	100	(32)	100	(32)	74	(31)	20	(17)
$S^{orth(2)}$, $\alpha = 8.52$	11	76	(11)	75	(11)	42	(11)	9.8	(7.3)
$S^{enf(2)}$, $\alpha = 0.60$	100	520	(100)	520	(100)	300	(97)	71	(59)
S^{SOSF}	10	32	(9.8)	31	(9.8)	19	(9.1)	4.6	(4.2)
S_1	23	200	(23)	200	(23)	200	(23)	180	(23)

Table 3.3: Fisher constraints on f_{NL} for templates with varied divergences and μ-dependent properties in the squeezed limit. We present the forecasted constraints for a Planck-like CMB experiment, and from the halo properties of galaxies measured by a Euclid-like spectroscopic survey [top] and a DESI-like survey [bottom]. The LSS constraints are shown without CMB and with (in parentheses) and with four different assumptions on the amount of information known about the galaxy bias: assuming no knowledge and marginalizing over the bias in independent redshift bins ('marg.'), assuming a 10% and 1% prior knowledge, and assuming the extreme limit, in which the galaxy bias has been measured precisely by a complementary method.

One way of strengthening constraints on f_{NL} is to improve our understanding of the galaxy bias, through constraints from other probes. Figure 3.7 shows that the redshift evolution and scale dependence of the non-Gaussian galaxy bias are distinct from that of the Gaussian bias predicted using the halo model

and different for both the ELG and LRG galaxy samples at the same redshift. These differences might be utilized to extract out information about the Gaussian component of the galaxy bias, using weak lensing [149, 150, 6, 151] and three-point statistics [158]. These analyses suggest 1-10% constraints using higher order statistics, and comparable or better constraints using weak lensing might be achievable. We have therefore considered constraints from the galaxy clustering power spectrum alone, with the possibility of additional information by imposing 1% and 10% constraints on the Gaussian galaxy bias, and the limiting case, of perfect Gaussian bias determination, in which the Gaussian bias is not marginalized over at all in the analysis. In Table 3.3 we show projected constraints on f_{NL} with a range of assumptions about the Gaussian galaxy bias, including 1% and 10% priors roughly consistent with current constraints from the literature (e.g. [155, 145, 157, 168]).

One interesting feature of the constraints in Table 3.3 is that DESI's marginalized constraints for shapes with scale-independent large-scale halo biases (S^{equil}, $S^{orth(2)}$, $S^{enf(2)}$, and S^{SOSF}) are stronger than those forecasted for Euclid. The principal constraining power in these cases comes from DESI's LRG population, which we attribute to the LRGs having a different HOD that probes more massive halos. At the same time, the ELG population is still better at constraining the more scale-dependent shapes like the local, orthogonal, and enfolded templates. This seems to hint that different tracers are possibly better suited for constraints on different shapes of non-Gaussianity.

In Figure 3.8, we show how the contribution to a Fisher matrix diagonal element at a fixed redshift-slice, e.g. $dF_{f_{NL},f_{NL}}(z)/dk$, varies with k. The magnitude of the Fisher contribution at a particular k scale indicates the amount of informa-

tion, or constraining power, that we can gain by measuring the galaxy power spectrum at k. Figure 3.8 illustrates that, with the exception of the local type of non-Gaussianity at very high redshifts, the principal strength of a template's constraint relies on information coming from small quasi-non-linear scales. To further illustrate the importance of small scales, we computed the LEO constraints in the case where we take more conservative assumptions about how much information we can extract from small scales. We model this conservative case using $k_{max} = 0.15\,h/\mathrm{Mpc}$ and $p_{NL} = 1$. In this case, the LEO marg (unmarg) constraints are: $\sigma(f_{NL}^{local}) = 9.0\,(8.5)$, $\sigma(f_{NL}^{equil}) = 390\,(33)$, and $\sigma(f_{NL}^{orth}) = 79\,(23)$. Compared to the more optimistic, default case, the marginalized fixed equilateral and orthogonal constraints weaken by factors of ~ 1.4 and 1.5, while the local constraint is largely unaffected.

We note that in general the constraints here are weaker than in some previous analyses of constraints from halo bias measurements, e.g. [121, 6]. These arise from updates lowering estimates of galaxy number counts expected from Hα surveys [166, 169] relative to prior expectations [170], the inclusion of non-linear damping effects on small scales, and the use of the full, rather than asymptotic, forms for scale-dependence of the halo bias that alter the small scale constraints in particular for less divergent shapes.

Furthermore, the constraints also differ due to different assumptions about the Gaussian galaxy bias. If instead of using the fiducial galaxy biases from Orsi *et al.* [5], we take the galaxy bias from the halo model as our fiducial, as shown in Figure 3.1, the resulting marginalized (unmarginalized) template constraints on LEO for a Euclid-like survey are weakened to $\sigma(f_{NL}^{local}) = 8.6\,(7.5)$, $\sigma(f_{NL}^{equil}) = 280\,(23)$, $\sigma(f_{NL}^{orth}) = 54\,(20)$. This is consistent with previous forecasts,

which also showed the sensitivity of constraints to the assumed fiducial bias [6].

We have forecasted constraints on fixed shape templates from surveys like DESI and Euclid, and explored the sensitivity of the constraints to different survey specifics and astrophysical uncertainties. For Euclid-like forecasts, we have quantified how resulting constraints have changed with updated and decreased expected number counts of Hα galaxies, while for a DESI-like survey, we have seen that different galaxy populations provide tighter constraints on different templates. We have additionally quantified the effect that different fiducial values and priors on the Gaussian galaxy bias, and conservative vs optimistic treatments of small quasi-linear scales, impacts the resulting constraints.

Discriminating between shapes

In addition to considering single templates, we are also interested in the discriminating power of the data to differentiate between templates. We first consider the joint constraints on the local, equilateral, and orthogonal (LEO) templates.

In Figure 3.9 we show the 2D marginalized constraints on these templates from the CMB, and Euclid-like and DESI-like spectroscopic galaxy surveys. With no prior information about the galaxy bias the constraints on LEO non-Gaussianity from the halo bias are projected to be much weaker than for the CMB. If galaxy bias information is known from a complementary source, such as overlapping weak lensing statistics, then the LSS gives additional constraining power in the LE and, especially, the EO planes, in which the LSS and CMB constraints are more orthogonal. In the case of DESI, the low-z LRG and high-

z ELG galaxies have complementary degeneracy directions in the LEO constraints, coming from the different HODs specific to red vs. blue galaxies.

We have also considered the properties of shapes that are not well described purely by the LEO basis, because of a complex divergence properties, or richer structure, examples include $S_{ortho(2)}, S_{enf(2)}, S_1$, and S_{SOSF}. In some cases these have similar squeezed limits, and therefore large scale halo bias properties, but are dissimilar, or have anisotropic properties, as one moves away from the squeezed limit. These provide distinctive differences at smaller scales. In Figure 3.10 we consider the potential for future surveys to differentiate between pairs of such shapes.

In the case of the orthogonal and enfolded shapes, halo predictions for the two shapes are too similar to distinguish them with the galaxy power spectrum. The degeneracy directions in the LSS and CMB data are somewhat complementary, and with knowledge of the galaxy bias, the overall LSS+CMB constraint is stronger than either data set alone.

The S_1 shape, has the same isosceles squeezed limit (up to a constant) as the orthogonal template but is anisotropic as it approaches the squeezed limit. We find that the LSS constraints on a scenario looking to distinguish these two shapes are quite orthogonal to those of the CMB. Even in the case where the Gaussian bias is marginalized over and the two shapes are highly degenerate with LSS data alone, the LSS+CMB constraint is improved over just the CMB constraint. With prior knowledge of the Gaussian bias, the forecasts suggest that degeneracies that exist between the anisotropic S_1 shape and isotropic S_{orth} shape can be further broken with the inclusion of LSS data.

We can also consider how well one might distinguish between models that all vanish in the squeezed limit. To do this we consider discriminating between the equilateral template and two others: the orthogonal(2) and SOSF shapes. Because of their common large scale scale-independent bias there is a strong degeneracy, and weak constraints, for the LSS-only constraints with no prior knowledge about the Gaussian galaxy bias. If the Gaussian bias is known well, however, the degeneracy directions between the CMB vs LSS are very complementary.

Finally, beyond distinguishing pairs of bispectrum shapes, one could consider model-independent constraints on a general shape constructed from a linear combination of LEO models. Figure 3.11 shows k-dependent constraints that quantify how well the product $f_{NL}S(k_1, k_2, k_3)$ may be constrained for different triangle configurations, under the assumption that the underlying primordial shape is any linear combination of the three LEO templates. For the calculation of these errors, we have used the principal component analysis method described in [17], where $\sigma^2(f_{NL}S(k_1, k_2, k_3)) = \sum_i \sigma^2(b_i)\hat{e}_i(k_1, k_2, k_3)^2$, $\sigma^2(b_i)$ are the eigenvalues of the Fisher matrix for the f_{NL}^{local}, f_{NL}^{equil}, and f_{NL}^{orth} parameters, and $\hat{e}_i(k_1, k_2, k_3)$ are the shapes corresponding to the eigenvalues, each of which are linear combinations of the LEO templates, that have zero covariance for the modeled survey. These configuration-dependent constraints in Figure 3.11 are strongest on triangles that are scalene and somewhat flattened, not equilateral triangles, where shape amplitudes are typically normalized. Prior knowledge of the Gaussian bias improves the constraints but does not significantly alter the morphology of the best measured region.

3.4 Discussion & Conclusions

The large number of viable inflationary theories currently in the literature warrant a careful consideration of how well upcoming galaxy surveys will be able to observationally discriminate between contending models. In this work, we focused on forecasting the ability of forthcoming galaxy power spectrum data from the spectroscopic portions of future surveys, such as Euclid and DESI, to constrain amplitudes of a variety of non-Gaussian shapes, as well as to distinguish between shapes, and put (k_1, k_2, k_3)-dependent constraints on a general primordial shape. In particular, the galaxy power spectrum is affected by changes to the halo bias and halo mass function that arise from non-zero primordial shapes.

Although the halo bias on large-scales generally reflects the primordial shape's behavior in the squeezed limit (i.e. if the squeezed limit of the isotropic primordial shape scales as $\propto 1/k^{\alpha}$, then the halo bias generally goes as $\propto 1/k^{\alpha+1}$), we find this simple correlation between late-time observables and the primordial shape is less straightforward when shapes which are anisotropic in the squeezed limit, such as S_1, are considered.

We consider the effects of non-Gaussianity beyond the large scale, scale dependent halo bias, including the scale independent corrections and the effect on the halo mass function. By defining two weighting functions, we found that while the halo bias on large scales is a weighted average of the shape that is strongly peaked on very squeezed configurations (as expected from analytic arguments in Section 3.3.1), the halo mass function and halo bias on small scales probes more general configurations that are not very squeezed.

A significant hurdle is the degeneracy between the signature induced by non-Gaussianity in the scale-independent halo bias and the Gaussian galaxy bias. The halo model bridges the gap between halos and their resident galaxies and, to a large extent, the galaxy power spectrum preserves the effect of non-Gaussianity on the halos: the large-scale behavior of P_{gal} depends largely on the scale-dependent halo bias, while on smaller scales the k-dependence of P_{gal} is a combination of all effects, from the halo bias, halo mass function and the halo occupation distribution. In theory the use of different galaxy samples, with different HODs, and redshift distributions as tracers, which map the halo information to the galaxy bias in different ways, could allow the Gaussian and scale-independent non-Gaussian contributions to be disentangled.

Understanding the galaxy power spectrum on small quasi-non-linear scales and measuring it precisely can enrich our understanding of the primordial shape: rather than using the survey data to only constrain the squeezed limit of the primordial shape, we can gain information about other regions of the shape as well. We include the potential to differentiate between these two effects, for example through the additional information from weak gravitational lensing, in the Fisher analysis through considering constraints in the presence of a prior on the Gaussian galaxy bias. The forecasted constraints modeled on Euclid and DESI for a range of templates, show that for shapes with a significant scale-dependent bias, assumptions about the scale-independent bias are not critical. In the cases where there is not significant scale dependence, DESI's LRGs are a more powerful probe of f_{NL} in the absence of any information about the Gaussian galaxy bias, while knowledge of the Gaussian galaxy bias to a percent or better makes Euclid and DESI's constraints comparable, and both LSS data sets competitive with CMB constraints, such that the combined LSS+CMB

constraint is better than either alone.

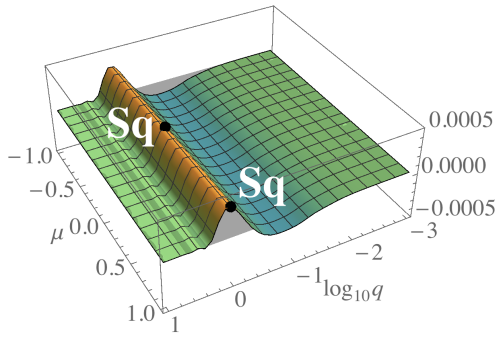
In addition to constraints for fixed shapes we show the forecasts for each survey's ability to distinguish between the local, equilateral, and orthogonal shapes, using LSS-data only, CMB-data only, and the combination of LSS+CMB data. In this case, when no additional information about the Gaussian bias is included, Euclid- and DESI-like constraints are similarly weak in comparison to CMB constraints, though the two LSS surveys have different degeneracy directions in the LEO planes, due to their differing galaxy populations. However, if the galaxy bias is constrained the Euclid-like and DESI-like surveys are comparable, with Euclid's constraints being slightly stronger. Additionally we find that, with strong priors on the galaxy bias, LSS and CMB constraints have complementary degeneracy directions, particularly in the $(f_{NL}^{local}, f_{NL}^{equil})$ and $(f_{NL}^{equil}, f_{NL}^{orth})$ planes.

We explored how well LSS data might distinguish between shapes which have similar scale-dependent biases on large scales, leveraging their quite different forms away from the squeezed limit, or their distinct, anisotropic vs isotropic, squeezed limits. We find that with strong priors on the galaxy bias, LSS and CMB constraints have complementary degeneracy directions and the combination of the data improves upon either data set alone, but the LSS data alone with no other information about the Gaussian galaxy bias is typically not going to noticeably improve on the CMB-only constraints.

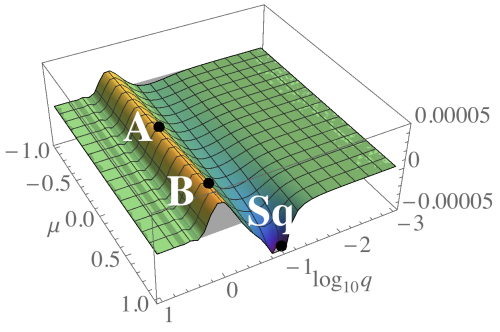
Finally, we computed the k-configuration-dependent constraints on general shapes via a principal component analysis. The principal components illustrate what features of the primordial shape the galaxy power spectrum are best and worst measured, and the k-dependent constraints show how well we can

measure different configurations of a general shape. Without any additional knowledge about the galaxy bias, the LSS constraints do not improve on the CMB constraints. With percent or better priors on the galaxy bias, the LSS+CMB constraints are improved accordingly, though the overall morphology of the k-dependent constraints on the general shape are largely unchanged.

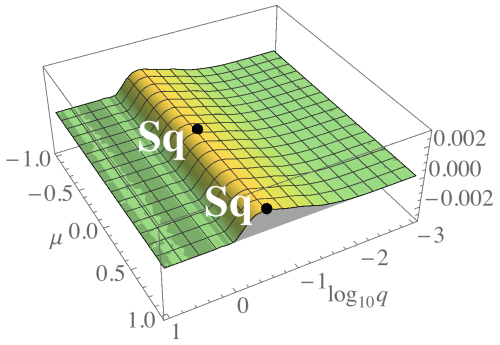
In conclusion, we have quantitatively demonstrated that the galaxy power spectrum can provide a valuable complementary probe of non-Gaussianity to the CMB, and can probe bispectrum configurations beyond the asymptotic squeezed limit. A key requirement to realize the full utility of the galaxy cluster measurements is knowledge of the galaxy bias. This is a rich area we plan to investigate in future work: to quantify the ability for upcoming weak lensing measurements and higher order galaxy clustering statistics to constrain the galaxy bias, as well as the potential for utilizing other tracers of the halo distribution to provide additional constraints.



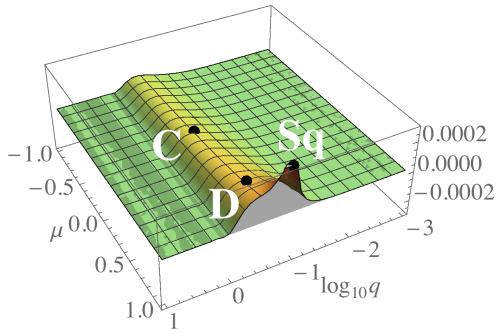
(a) $m = 10^{12.5}\, M_\odot/h$, $k = 10^{-3}\, h/\mathrm{Mpc}$



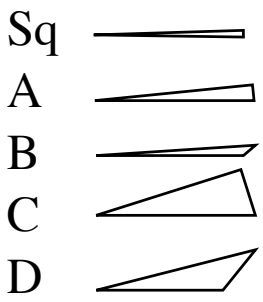
(b) $m = 10^{12.5}\, M_\odot/h$, $k = 0.1\, h/\mathrm{Mpc}$



(c) $m = 10^{13.5}\, M_\odot/h$, $k = 10^{-3}\, h/\mathrm{Mpc}$



(d) $m = 10^{13.5}\, M_\odot/h$, $k = 0.1\, h/\mathrm{Mpc}$



(e)

Figure 3.3: The halo bias weight function, w_{sd}, for halo masses $m = 10^{12.5} M_\odot/h$ [top] and $m = 10^{13.5} M_\odot/h$ [middle], and scales $k = 10^{-3}\, h/\mathrm{Mpc}$ [left] and $k = 0.1\, h/\mathrm{Mpc}$ [right] as a function of the wavenumber q and angle μ used to described the bispectrum shape. We have labeled specific configurations at which each weight shown peaks. Some points are simply labeled as 'Sq' signifying that the triangular configuration is significantly squeezed and not easily illustrated to scale. For weights at $k = 0.1\, h/\mathrm{Mpc}$, we label other configurations and illustrate the corresponding triangles, which are much less squeezed than those labeled 'Sq', showing that Δb_{sd} on small scales can probe configurations beyond very squeezed triangles.

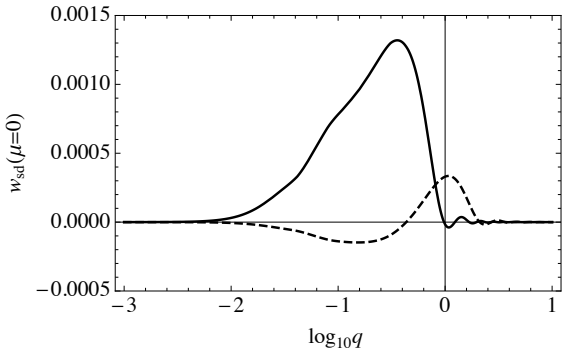
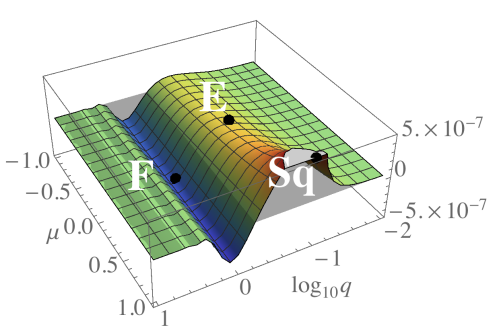
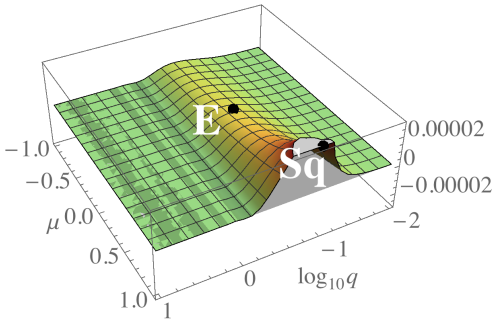


Figure 3.4: The solid and dashed curves respectively show the $\mu = 0$ slice of the weight from the scale-dependent bias for $k = 10^{-3}\,h/\text{Mpc}$ and $M = 10^{13.5} M_\odot/h$ and $M = 10^{12.5} M_\odot/h$ cases.



(a) $m = 10^{12.5}\,M_\odot/h$, $k = 0.1\,h/\text{Mpc}$



(b) $m = 10^{13.5}\,M_\odot/h$, $k = 0.1\,h/\text{Mpc}$

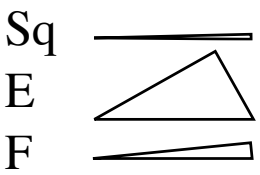


Figure 3.5: The weight function $w_{R_{NG}}$, parameterizing the shape sensitivity in non-Gaussian corrections to the the halo mass function, as defined in 3.29. The function is plotted for a fixed $k = 0.1\,h/\text{Mpc}$ slice, where the weight is peaked, and two halo masses, $m = 10^{12.5} M_\odot/h$ (left) and $m = 10^{13.5} M_\odot/h$ (right). The plots illustrate that there are non-negligible contributions from configurations that are not significantly squeezed, like those labeled E and F.

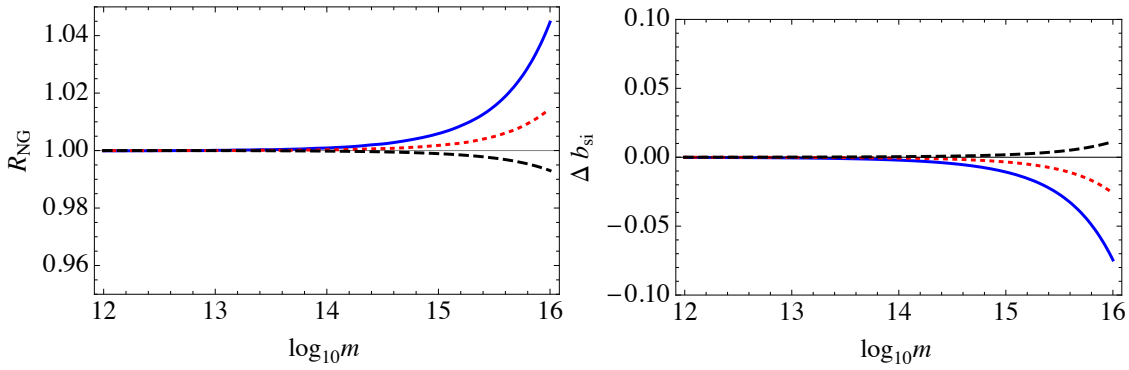


Figure 3.6: The effect of the non-Gaussian corrections to the halo mass function [left panel] and the scale-independent halo bias [right panel] on the halo power spectrum relative to the Gaussian case P_h^G, for the local [full blue], equilateral [dotted red], and orthogonal [dashed black] shapes at $z = 1$ for $f_{NL} = 100$.

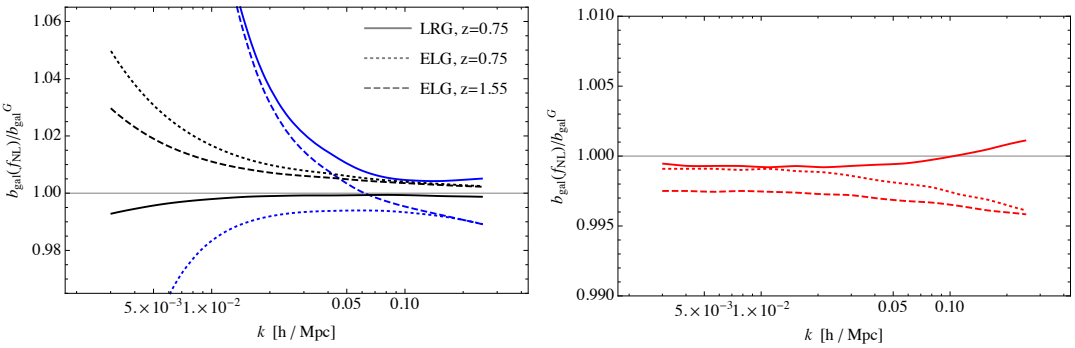


Figure 3.7: The predicted galaxy bias as a function of scale for luminous red galaxies (LRGs) at $z = 0.75$ [solid lines], and emission line galaxies (ELGs) at $z = 0.75$ [dotted] and $z = 1.55$ [dashed], for a variety of non-Gaussian bispectrum templates with $f_{NL} = 100$: local [blue] and orthogonal [black] (left panel), as well as equilateral [red] (right panel).

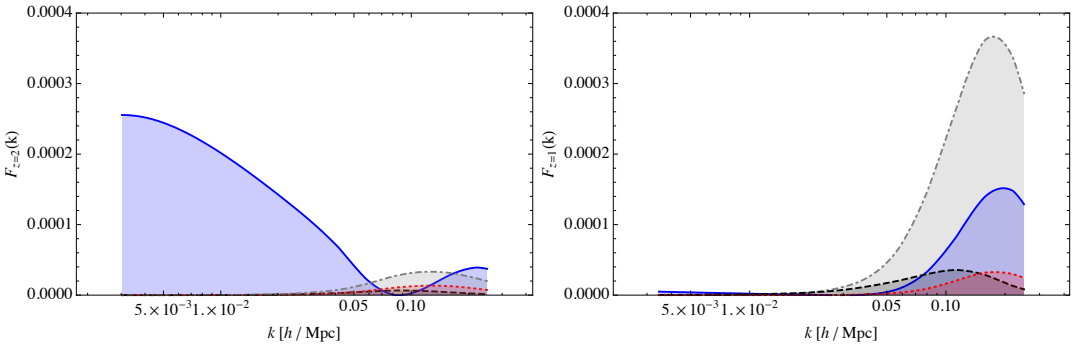


Figure 3.8: Contributions to a Fisher element as a function of k, $F_z(k)$, for (f_{NL}, f_{NL}) in a fixed redshift slice, $z = 2$ [left] or $z = 1$ [right], from the local [blue], equilateral [dotted red], and orthogonal [dashed black] templates. Contributions to the Fisher element for (b^G_{gal}, b^G_{gal}) are also shown [dot dashed gray], multiplied by a scaling factor of 10^{-8} in order to show on the same plot. Unless the primordial shape has a strongly scale-dependent bias on large-scales, most of the resulting constraint on f_{NL} comes from intermediate and small scales, $k \gtrsim 0.05\,h$/Mpc. The constraint on the Gaussian galaxy bias, while it has a larger overall amplitude, its shape is qualitatively similar to the signal on small scales coming from f_{NL}.

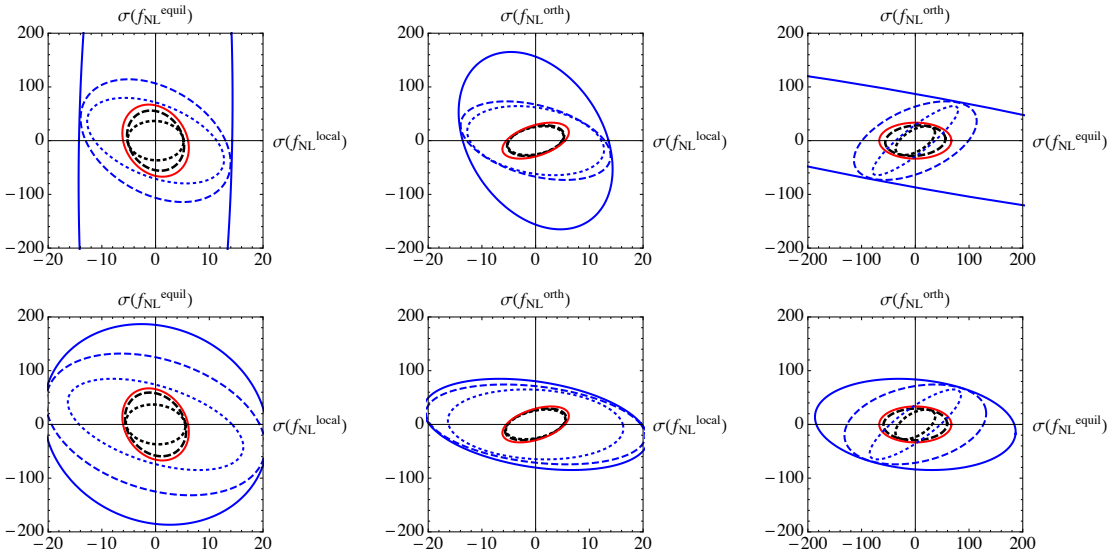


Figure 3.9: Euclid-like [top] and DESI-like [bottom] Fisher matrix projections for the constraints on the amplitudes of contributions to a general shape constructed from local, equilateral and orthogonal (LEO) templates. Constraints from the halo bias only [blue], CMB only [red], and in combination [black] are shown for a range of assumptions on prior knowledge of the galaxy bias: the most conservative scenario [full lines] in which all Gaussian bias information is obtained from this dataset and b^G_{gal} is marginalized over fully, the opposite regime [dotted lines], in which galaxy bias are wholly determined by an alternative technique, such as those discussed in section 3.2.4, and is not marginalized over, and the optimistic scenario [dashed lines] in which the Gaussian bias is well-constrained which we represent using a 1% prior.

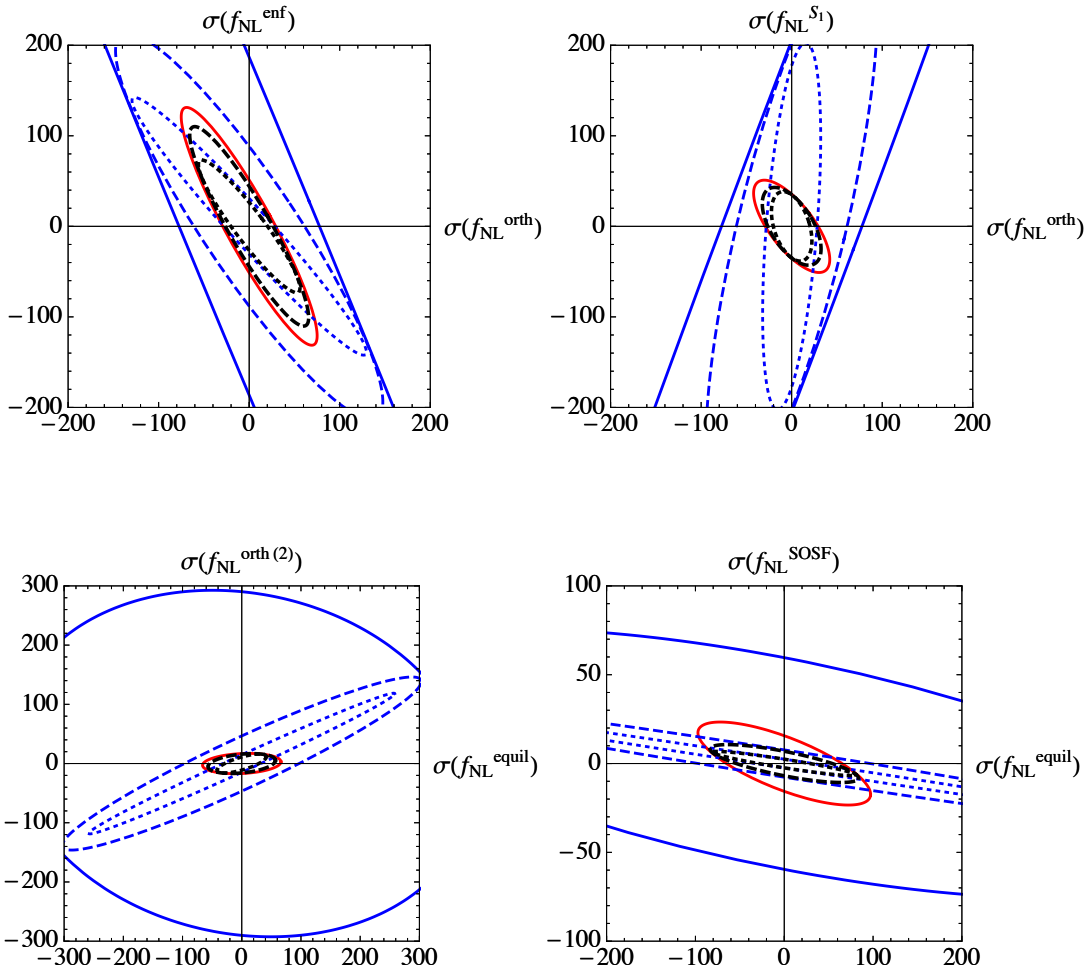


Figure 3.10: Euclid-like Fisher matrix projections for the joint constraints on the amplitudes of a pair of shapes. Constraints from the halo bias only [blue], CMB only [red], and in combination [black] are shown for scenarios in which b_{gal}^{G} is marginalized over fully [full], in which galaxy bias is wholly determined by an alternative technique [dotted lines], and in which the Gaussian bias is well-constrained, which we represent using a 1% prior [dashed lines].

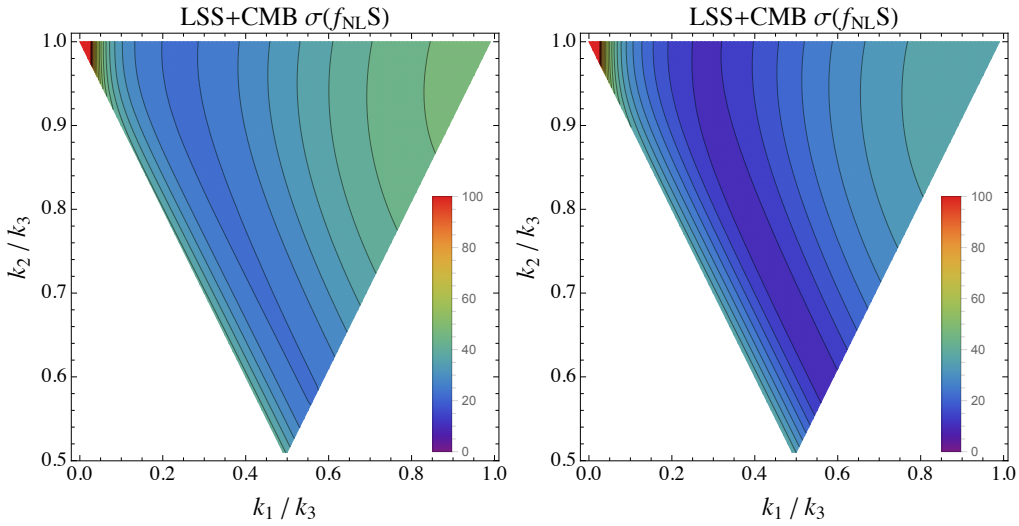


Figure 3.11: Joint LSS+CMB k-dependent constraints on a general shape, $\sigma(f_{NL}S(k_1, k_2, k_3))$, assuming a basis set of local, equilateral, and orthogonal templates assuming no [left] and perfect [right] prior knowledge of the Gaussian bias.

CHAPTER 4

A NEW BASIS FOR NON-SEPARABLE BISPECTRA

4.1 Introduction

We are fortunate to live in a time when cosmological datasets can probe the Universe in exquisite detail. In particular, the cosmic microwave background (CMB) provides a rich source of information about the very early Universe, and is an especially precise probe of the inflationary paradigm. An important question that we are now in a position to probe, more thoroughly than ever, is: what is the initial quantum state of inflationary fluctuations? It is usually taken to be the Bunch-Davies state, but from the point of view of treating inflation as an effective theory it is not unreasonable to consider the choice of state to be open, subject to the conditions that it allow for inflation to occur and that it be consistent with field theoretic precepts. Explicit examples of scenarios that give rise to non-Bunch-Davies initial conditions for inflation can be found in [171, 172, 173, 174, 175, 176]. Assuming that the initial state is more general than the free vacuum, such as a Bogoliubov transform of the Bunch-Davies state or even a mixed state, we can calculate its imprint on cosmological observables like the power spectrum and bispectrum of inflationary perturbations, and in turn those of CMB temperature anisotropies [177, 178, 179, 180, 181, 39, 49, 50, 182, 124, 183, 172, 184, 185, 173, 51, 186, 187]. Whether these effects can actually be observed in cosmological data depends on the extent of departure from a Bunch-Davies state, the number of e-folds of inflation beyond the minimum required, and of course the sensitivity of our experiments [188, 189, 125].

Given these choices, how can we narrow down the possibilities? The initial state of perturbations will leave its imprints on their correlation functions, for example, logarithmic oscillatory modulations in the CMB power spectrum [190]. Higher order correlators, such as the bispectrum [191, 117], are extremely sensitive to deviations from the Bunch-Davies state. The bispectrum carries information both about the amplitude of the correlations, typically encoded in $f_{\rm NL}$, as well as of preferred configurations in momentum space; the three momenta must form a triangle, but the shape of the triangle is sensitive to both the interactions of the inflaton, including the mixing with gravity, as well as the initial state.

The bispectrum for general initial states is highly oscillatory and cannot be written in a separable form, i.e. as a product of separate functions of the three momentum modes. This makes the study of these states via the bispectrum computationally difficult. For such shapes we usually construct a basis of separable functions and rewrite the desired shape as a sum over many such basis functions, for example, using a polynomial basis [16], Fourier basis [73], or divergent basis [17].[1] As long as the original and reconstructed shapes are very similar and have a significant overlap, or in other words have a cosine close to unity, we can look for signatures of the reconstructed shape in the CMB bispectrum and be assured that the result will be an accurate reflection of what we would have obtained for the real shape. Non-Bunch-Davies shapes, however, can be difficult to efficiently describe with existing bases [16, 73, 17, 117]. This leaves open the possibility that signatures of general initial states could still be present, yet undetected, in the CMB data.

[1] In special shape-specific cases, other basis sets can be used: for example, feature or resonant models exhibiting linear or logarithmic oscillations, respectively, can be efficiently reconstructed through a one-dimensional (1D) expansion in the sum of wavenumbers, $k_1 + k_2 + k_3$ [192, 193].

In this paper we describe a new basis of piecewise spline functions and use it to fit non-Bunch-Davies shapes of the bispectrum. The *spline* basis consists of polynomial functions defined locally, between various points called "knots" in 3D space [194, 195]. This makes it particularly suitable for smoothing and interpolating data with complex patterns, and in our case, for capturing localized features of any complicated shape of the bispectrum. Another immediate advantage of defining localized functions is that the basis functions are orthogonal to a good approximation, and there is no need to perform a Gram-Schmidt-like orthogonalization process on the basis functions. Instead, the basis-fitting algorithm we adopt does not rely on having an orthonormal basis. This makes the spline basis easy to implement and to extend, using a large number of mode functions to capture fine features in the bispectrum. We find that the spline basis performs at least as well as the polynomial and Fourier bases in describing most non-Bunch-Davies shapes, and for many shapes offers significant improvements (in the cosines) over existing techniques.

The remaining paper is organized as follows. In section 4.2 we briefly review how we define the initial conditions for the perturbations and obtain the bispectrum for general initial states. We describe the spline basis in section 4.3 and use it to fit non-Bunch-Davies shapes in section 4.4. We conclude in section 4.5 with a summary and discussion on the scope of this work. Appendices A.1 and A.2 contain details on calculations of the correlation functions for general initial states, and appendix A.3 describes our numerical implementation of the spline basis.

4.2 Non-Bunch-Davies shapes

We usually describe primordial correlation functions in terms of the curvature perturbation $\zeta(t, \vec{x})$, since this quantity is conserved outside the horizon [88]. It is defined as the perturbation in the local scale factor $a(t)$; the metric perturbation $h_{ij}(t, \vec{x})$ is then written as $h_{ij} = a^2 e^{2\zeta}\delta_{ij}$. In an effective field theory setting, $\zeta(t, \vec{x})$ is related to the Goldstone mode of time reparameterization symmetry breaking, usually denoted as $\pi(t, \vec{x})$ [26, 196]. In this section we describe how the choice of initial state for the perturbation $\zeta(t, \vec{x})$ (or equivalently $\pi(t, \vec{x})$) affects the primordial bispectrum $B_\zeta(k_1, k_2, k_3)$ defined via $\langle \zeta_{\vec{k}_1}\zeta_{\vec{k}_2}\zeta_{\vec{k}_3}\rangle = (2\pi)^3\delta^3(\vec{k}_1+\vec{k}_2+\vec{k}_3)B_\zeta(k_1, k_2, k_3)$.

Starting with the Einstein-Hilbert action, we can calculate the action for scalar perturbations directly in the ζ-gauge. Writing down the most general Lorentz-invariant scalar-tensor theory with second order equations of motion results in the Horndeski action for the perturbations [197, 198, 199, 200, 42].[2] For $P(X, \phi)$ models of inflation, with $X \equiv -g_{\mu\nu}\partial^\mu\phi\partial^\nu\phi$,[3] the leading order in slow-roll Horndeski action at cubic order in the perturbations is given by

$$S = \int \mathrm{d}^3x\,\mathrm{d}t\,a^3\left\{\frac{\varepsilon}{c_s^2}\left[\dot{\zeta}^2 - \frac{c_s^2}{a^2}(\partial_i\zeta)^2\right] + \Lambda_1\dot{\zeta}^3 + \Lambda_2\zeta\dot{\zeta}^2 + \frac{\Lambda_3}{a^2}\zeta(\partial_i\zeta)^2\right\} + S_{\text{boundary}}\,. \tag{4.1}$$

Here we have set $M_{\rm Pl} = 1$, $\varepsilon \equiv -\dot{H}/H^2$ is the slow-roll parameter, c_s is the constant sound speed for perturbations, and the couplings $\Lambda_1 - \Lambda_3$ are given by

$$\Lambda_1 = \frac{\varepsilon}{Hc_s^4}\left(1 - c_s^2 - 2\frac{\lambda c_s^2}{\Sigma}\right), \tag{4.2}$$

[2] The Horndeski action does not describe ghost inflation, however, which can be included in an effective field theory setting. Another example outside the Horndeski domain is Horăva-Lifshitz gravity, in which Lorentz invariance is explicitly broken.

[3] The action for many single scalar field models of inflation can be written as $S = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\,[R + 2P(X, \phi)]$, with ϕ controlling the dynamics of both the background and perturbations.

$$\Lambda_2 = -3\frac{\varepsilon}{c_s^4}\left(1-c_s^2\right), \tag{4.3}$$

$$\Lambda_3 = \frac{\varepsilon}{c_s^2}\left(1-c_s^2\right), \tag{4.4}$$

with $\lambda \equiv X^2\frac{\partial^2 P}{\partial X^2} + \frac{2}{3}X^3\frac{\partial^3 P}{\partial X^3}$ and $\Sigma \equiv X\frac{\partial P}{\partial X} + 2X^2\frac{\partial^2 P}{\partial X^2} = \frac{H^2\varepsilon}{c_s^2}$. The boundary terms in eq. (4.1) are important and in general do contribute to the bispectrum when the initial state is different from the Bunch-Davies vacuum; here we assume for simplicity that the initial state for $\zeta(t,\vec{x})$ is defined in such a way as to cancel these boundary terms. It is also worth noting that the action in eq. (4.1) is equivalent to that obtained using an effective field theory approach [26, 196] up to boundary terms, and the bispectrum for general initial states agrees between the two actions, as shown in [201].

The usual method to obtain correlation functions of $\zeta(t,\vec{x})$ is to use in-in perturbation theory [202, 203, 204, 205, 206, 207, 208, 209]. The initial state can be input as a density matrix defined at the initial time t_0 [51]; for simplicity we will assume that the initial density matrix is pure and Gaussian.[4] Doubling the fields on the plus and minus branches of the in-in contour, we can calculate the Green's function corresponding to the quadratic (or "free") part of the action in eq. (4.1), including the effect of a general initial state; details of this calculation can be found in the appendix of [51]. The Green's function for $\zeta(t,\vec{x})$ is found to be

$$\mathcal{G}_k^{\zeta}(\eta,\eta') = \frac{c_s^2}{2\varepsilon}\frac{1}{a(\eta)a(\eta')}\begin{pmatrix} G_k^{++}(\eta,\eta') & G_k^{+-}(\eta,\eta') \\ G_k^{-+}(\eta,\eta') & G_k^{--}(\eta,\eta') \end{pmatrix}, \tag{4.5}$$

where η is the conformal time defined as $\eta = \int \mathrm{d}t/a$ and the factor out front comes from rewriting the quadratic action in terms of the canonically rescaled

[4] Here by "pure" we are distinguishing between pure and mixed quantum states. Mathematically speaking, $\mathrm{Tr}\left(\rho^2\right) = 1$ for pure states, while $\mathrm{Tr}\left(\rho^2\right) < 1$ for mixed states, ρ being the density matrix. Relaxing the pure state assumption leads to qualitatively very similar results to what we discuss here. By "Gaussian" we mean that the action describing the initial state is quadratic.

field $\chi(t,\vec{x})$, $\zeta = \frac{1}{\sqrt{2\varepsilon}}\frac{c_s}{a}\chi$. The functions $G_k^{\pm,\pm}(\eta,\eta')$ are given by

$$G_k^{++}(\eta,\eta') = f_k^{>}(\eta)f_k^{<}(\eta')\theta(\eta-\eta') + f_k^{<}(\eta)f_k^{>}(\eta')\theta(\eta'-\eta)\,, \tag{4.6}$$

$$G_k^{+-}(\eta,\eta') = f_k^{<}(\eta)f_k^{>}(\eta')\,, \tag{4.7}$$

$$G_k^{-+}(\eta,\eta') = G_k^{+-*}(\eta,\eta')\,, \tag{4.8}$$

$$G_k^{--}(\eta,\eta') = G_k^{++*}(\eta,\eta')\,, \tag{4.9}$$

and the mode functions $f_k^{\gtrless}(\eta)$ are solutions to the second order differential equation resulting from the Green's function equation,

$$f_k^{\gtrless}(\eta) = \alpha_k^{\gtrless}h_k(\eta) + \beta_k^{\gtrless}h_k^*(\eta) - 2f_k^{\gtrless}(\eta_0)A_k g_k(\eta)\theta(\eta_0-\eta)\,. \tag{4.10}$$

The last term above is an additional contribution from the initial density matrix, A_k being the kernel that multiplies the $\zeta_{\vec{k}}^{+}\zeta_{-\vec{k}}^{+}$ term in the initial state action [51]; it does not, however, contribute to the bispectrum, and so we will ignore it in our discussion. The Bogoliubov coefficients $\alpha_k^{\gtrless}$, $\beta_k^{\gtrless}$ are functions of kernels in the initial density matrix, the mode functions, and time derivatives of the mode functions, all at the initial time, and satisfy $\alpha_k^{<} = \beta_k^{>*}$, $\beta_k^{<} = \alpha_k^{>*}$, $\left|\alpha_k^{>}\right|^2 - \left|\beta_k^{>}\right|^2 = 1$. Finally, the function $h_k(\eta)$ is defined at leading order in slow-roll as

$$h_k(\eta) = -\frac{1}{2}(-\pi\eta)^{1/2}H_{3/2}^{(1)}(-c_s k\eta)\,, \tag{4.11}$$

where $H_{3/2}^{(1)}$ is a Hankel function. The Bunch-Davies choice consists of setting the initial time $\eta_0 \to -\infty$, the initial density matrix to unity, and additionally $\alpha_k^{>} = 1$, $\beta_k^{>} = 0$ so that the mode function $f_k^{>}(\eta)$ picks out the positive frequency solution proportional to $e^{-ic_s k\eta}$.

The time η_0 at which the initial conditions are set can be taken to be a constant time in the past at the onset of infation, or can be considered as a scale-dependent quantity, $\eta_0(k)$. In the latter case, the initial conditions for each k

mode are set at the time when the physical momentum corresponding to this mode $c_s k/a(\eta_0)$ (with $a(\eta_0) = -1/(\eta_0 H)$ during inflation, at leading order) crosses a fixed energy scale Λ of new physics. The Bogoliubov transform in eq. (4.10) is correct for either choice of η_0. In appendix A.1 we show that for a scale-dependent initial time, this solution leads to the well-known oscillations in the late-time power spectrum [180, 181]. We use both choices of initial time in section 4.4 when we apply the spline basis to non-Bunch-Davies shapes of the bispectrum.

Let us now discuss how general initial states modify the bispectrum. Observables, such as the bispectrum, can be calculated using any combination of plus and minus fields on the in-in contour. For Gaussian initial states, we can calculate the three-point function in the perturbations as

$$\left\langle \zeta^+_{\vec{k}_1}\zeta^+_{\vec{k}_2}\zeta^+_{\vec{k}_3}\right\rangle(\eta) \;=\; \left\langle \zeta^+_{\vec{k}_1}(\eta)\zeta^+_{\vec{k}_2}(\eta)\zeta^+_{\vec{k}_3}(\eta)\times \exp\left[i\left(S^{(3)+}-S^{(3)-}\right)\right]\right\rangle_{\rm G}, \tag{4.12}$$

where $S^{(3)}$ is the cubic part of the action in eq. (4.1) written in momentum space, with conformal time derivatives, and with the time integral running from η_0 to η. The subscript "G" indicates that Wick contractions on the right are carried out using the Gaussian theory. At leading order in slow-roll, only the three operators $\dot\zeta^3$, $\zeta\dot\zeta^2$, and $\zeta(\partial_i\zeta)^2$ contribute, and we can write the three-point function at late times $\left\langle \zeta^+_{\vec{k}_1}\zeta^+_{\vec{k}_2}\zeta^+_{\vec{k}_3}\right\rangle(\eta)|_{\eta\to 0^-}$ as a sum of contributions from each of these three operators. To calculate the three-point function in eq. (4.12) at late times we need the function $G_k^{\zeta,++}(0,\eta')$ and its derivative $\partial_{\eta'}G_k^{\zeta,++}(0,\eta')$ for $\eta' \geq \eta_0$; using eqs. (4.6), (4.10) (discarding the $\theta(\eta_0-\eta)$ term), and (4.11) these are given by

$$G_k^{\zeta,++}(0,\eta') \;=\; \frac{H^2}{4\varepsilon c_s k^3}\left[a_k(1-ic_sk\eta')e^{ic_sk\eta'} + \; b_k(1+ic_sk\eta')e^{-ic_sk\eta'}\right] \tag{4.13}$$

and

$$\partial_{\eta'}G_k^{\zeta,++}(0,\eta') \;=\; \frac{H^2c_s\eta'}{4\varepsilon k}\left(a_ke^{ic_sk\eta'} + b_ke^{-ic_sk\eta'}\right), \tag{4.14}$$

where we have defined the functions

$$a_k = (\alpha_k^> - \beta_k^>)\,\alpha_k^{>*}, \tag{4.15}$$

$$b_k = -(\alpha_k^> - \beta_k^>)\,\beta_k^{>*}. \tag{4.16}$$

Using these in eq. (4.12) and performing the time integrals we find that the contributions to the bispectrum from the three operators are given by

$$\begin{aligned}\left\langle \zeta^+_{\vec{k}_1}\zeta^+_{\vec{k}_2}\zeta^+_{\vec{k}_3}\right\rangle_{\dot\zeta^3}(\eta)|_{\eta\to 0^-} &= -\frac{3}{32}(2\pi)^3\,\delta^3\left(\sum \vec{k}_i\right)\\ &\times \Lambda_1\frac{H^5}{\varepsilon^3}\left[\sum_{l,m,n=0}^{1} c^{(l)}_{k_1}\,c^{(m)}_{k_2}\,c^{(n)}_{k_3}\,\mathcal{F}_{\dot\zeta^3}\left((-1)^l k_1,(-1)^m k_2,(-1)^n k_3,\eta_0\right)\right]\\ &+ \text{c.c.}\,,\end{aligned} \tag{4.17}$$

$$\begin{aligned}\left\langle \zeta^+_{\vec{k}_1}\zeta^+_{\vec{k}_2}\zeta^+_{\vec{k}_3}\right\rangle_{\zeta\dot\zeta^2}(\eta)|_{\eta\to 0^-} &= \frac{1}{32}(2\pi)^3\,\delta^3\left(\sum \vec{k}_i\right)\\ &\times \Lambda_2\frac{H^4}{\varepsilon^3}\left[\sum_{l,m,n=0}^{1} c^{(l)}_{k_1}\,c^{(m)}_{k_2}\,c^{(n)}_{k_3}\,\mathcal{F}_{\zeta\dot\zeta^2}\left((-1)^l k_1,(-1)^m k_2,(-1)^n k_3,\eta_0\right)\right]\\ &+ \text{c.c.}\,,\end{aligned} \tag{4.18}$$

and

$$\begin{aligned}\left\langle \zeta^+_{\vec{k}_1}\zeta^+_{\vec{k}_2}\zeta^+_{\vec{k}_3}\right\rangle_{\zeta(\partial\zeta)^2}(\eta)|_{\eta\to 0^-} &= \frac{1}{64}(2\pi)^3\,\delta^3\left(\sum \vec{k}_i\right)\\ &\times \Lambda_3\frac{H^4}{c_s^2\varepsilon^3}\left[\sum_{l,m,n=0}^{1} c^{(l)}_{k_1}\,c^{(m)}_{k_2}\,c^{(n)}_{k_3}\,\mathcal{F}_{\zeta(\partial\zeta)^2}\left((-1)^l k_1,(-1)^m k_2,(-1)^n k_3,\eta_0\right)\right]\\ &+ \text{c.c.}\,,\end{aligned} \tag{4.19}$$

where "c.c." denotes complex conjugate and

$$c_k^{(i)} = \begin{cases} a_k & i = 0\\ b_k & i = 1\,. \end{cases} \tag{4.20}$$

The functions $\mathcal{F}_{\dot\zeta^3}$, $\mathcal{F}_{\zeta\dot\zeta^2}$, and $\mathcal{F}_{\zeta(\partial\zeta)^2}$ are written out explicitly in appendix A.2. The above equations give us the leading order result for the bispectrum for general initial states. The non-Bunch-Davies contributions to the bispectrum are

strongly peaked in the flattened limit $k_1 \approx k_2 + k_3$ (assuming that k_1 is the largest momentum mode) and in the squeezed limit $k_3 \ll k_1 \approx k_2$. (We show these enhancements and discuss apparent divergences in both limits in appendix A.2.) Further, these shapes are highly oscillatory, which makes them even harder to constrain using the CMB bispectrum. In the next section we discuss the spline basis that we use to rewrite these shapes as a sum of separable functions.

4.3 The spline basis

B-splines, short for "basis splines", are a well-established, and conceptually simple, mathematical formalism for curve fitting, using a set of piecewise polynomial functions [194, 195]. For a basis in one dimension, one chooses a set of "knots", $\{x_0, x_1, \ldots, x_N\}$, representing the points at which the polynomial function pieces are joined, and the degree q of the polynomials. For example, a 1D spline basis spanning $0 \leq x \leq 1$ with a set of six piecewise cubic polynomials is shown in Figure 4.1. As an explicit example of the functional form of the basis, one of the basis functions shown in the figure is

$$B_1(x) = \begin{cases} \frac{9}{4}\left(4x - 18x^2 + 21x^3\right) & 0 \leq x < \frac{1}{3} \\ \frac{1}{4}\left(8 - 36x + 54x^2 - 27x^3\right) & \frac{1}{3} \leq x \leq \frac{2}{3}\,. \end{cases} \tag{4.21}$$

The numerical coefficients defining the spline basis for input knots and q are easily generated using existing software libraries in many languages.

A general 1D function, $f(x)$, can be expanded, and approximated, using the spline basis,

$$f'(x) = \sum_{n=0}^{N-1} \alpha_n B_n(x)\,. \tag{4.22}$$

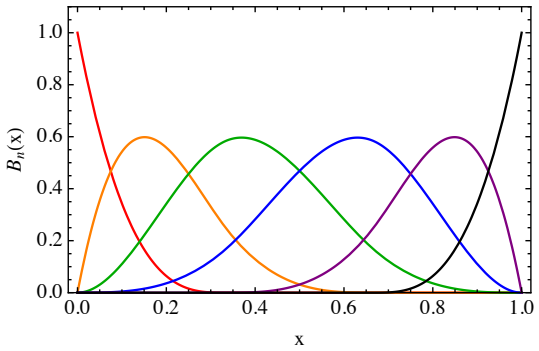


Figure 4.1: Example spline basis generated from knots at $x = \{0, 0, 0, 0, 1/3, 2/3, 1, 1, 1, 1\}$ and polynomials of degree $q = 3$.

The expansion coefficients are computed by first tabulating a sample of (x_i, f_i) values, where $f_i \equiv f(x_i)$, corresponding to M data points, and finding the values of $\{\alpha_n\}$ that minimize the least-squares function,

$$\text{LS} \quad = \quad \sum_{i=1}^{M} \left(f_i - \sum_{n=0}^{N-1} \alpha_n B_n(x_i) \right)^2 . \tag{4.23}$$

In practice, this requires solving the linear system of equations given by $B^T B\vec{\alpha} = B^T \vec{f}$, where B here is an $M \times N$ matrix containing the values of $B_n(x_i)$.

Similarly, we can generate a higher-dimensional basis by multiplying 1D b-splines together. For example, a 3D b-spline fit for a shape[5] can be expressed as

$$S'(k_1, k_2, k_3) \quad = \quad \sum_{k=0}^{N-1} \sum_{j=0}^{k} \sum_{i=0}^{j} A_{ijk} [B_i(k_1) B_j(k_2) B_k(k_3) \; + \; \text{perms}] , \tag{4.24}$$

where "perms" refers to permutations of k_1, k_2, and k_3. We note that the number of relevant (i, j, k) combinations in such a fit is not N^3, because each spline mode is symmetric in the three wavenumbers, and not all (i, j, k) combinations will correspond to a spline mode that is non-zero for wavenumber combinations

[5] A shape function is usually denoted by $S(k_1, k_2, k_3)$, not to be confused with the action S.

that can form a closed triangle. In the 2D and 3D cases, efficient algorithms solving for the expansion coefficients A_{ijk} already exist in the literature [210]. We illustrate how to use the spline basis further through explicit examples and include snippets of our numerical codes in appendix A.3.

In contrast to other bases used to create non-Gaussian templates made of globally varying functions, such as polynomial, Fourier, and divergent functions, the spline basis consists of a set of localized modes, each of which describes only a small region of the allowed k-space. This makes the spline basis very well-suited to describing shapes such as those of non-Bunch-Davies models, that are characterized by highly-peaked features concentrated on very flattened triangles. Because of the localization of spline modes, the modes are by construction nearly orthogonal, and no orthogonalization procedure (such as Gram-Schmidt) is used in our implementation. Avoiding any explicit orthogonalization is an advantage, as orthogonalizing the polynomial/Fourier basis sets via Gram-Schmidt is numerically unstable, and requires very high precision throughout to create a basis with a large number of modes.[6]

4.4 Fitting non-Bunch-Davies shapes

In this section we consider two non-Bunch-Davies shapes included in Planck's analysis that were not well-reconstructed in multipole-space using the polynomial modal expansion, $S_{\rm NBD1}$ and $S_{\rm NBD2}$ [117].[7] We also consider a set of general-

[6]The numerical instability of the classical Gram-Schmidt algorithm can be partially mitigated by instead adopting the modified Gram-Schmidt algorithm, which is what we have implemented in generating the polynomial and Fourier modes.

[7]In the notation of [117], $S_{{\rm NBD}i}$ here is equal to $(k_1k_2k_3)^2 B_\Phi^{{\rm NBD}i}/\left(2A^2 f_{\rm NL}^{{\rm NBD}i}\right)$, where $i = 1\,, 2$.

ized non-Bunch-Davies shapes, $S_{\dot{\zeta}^3}$, $S_{\zeta\dot{\zeta}^2}$, and $S_{\zeta(\partial\zeta)^2}$,[8] with different assumptions for the initial time boundary $c_s\eta_0$ and its potential wavenumber dependence. We consider three different choices for the initial time boundary: $c_s|\eta_0| = 10^3$ Mpc, corresponding to a choice of the largest scale that is amenable to our effective description of inflation, as well as wavenumber-dependent $c_s|\eta_0| = (\Lambda/H)/k_1$ or $(\Lambda/H)/(k_1+k_2+k_3)$, corresponding to alternate ways of defining the early crossing of modes into the effective field theory regime. We allow Λ/H to be 10 or 10^3, representing a physically-motivated range of values. These shapes span a wide variety of possible non-Bunch-Davies features in the bispectrum, and allow us to compare our results with previous analyses.

For any non-separable primordial shape S, and given a choice of basis $\{M_n\}$, one can compute a separable fit S' that approximates S as a linear combination of separable basis functions,

$$S'(k_1,k_2,k_3) \;=\; \sum_n \alpha_n \mathcal{M}_n(k_1,k_2,k_3)\,. \tag{4.25}$$

The similarity between the original shape and its fit is quantified by a cosine,

$$\cos(S,S') \;\equiv\; \frac{\langle S,S'\rangle}{\sqrt{\langle S,S\rangle\,\langle S',S'\rangle}}\,, \tag{4.26}$$

where the inner product is defined in Fourier space with a choice of weighting,

$$\langle S,S'\rangle \;\equiv\; \int \mathrm{d}\mathcal{V}_T\, S(k_1,k_2,k_3)S'(k_1,k_2,k_3)w(k_1,k_2,k_3)\,. \tag{4.27}$$

[8]The shapes $S_{\dot{\zeta}^3}$, $S_{\zeta\dot{\zeta}^2}$, and $S_{\zeta(\partial\zeta)^2}$ here are related to corresponding leading order non-Bunch-Davies corrections to the bispectrum; for example, $S_{\dot{\zeta}^3}(k_1,k_2,k_3) = (k_1k_2k_3)^2\Big[\mathcal{F}_{\dot{\zeta}^3}(-k_1,k_2,k_3,\eta_0) + \mathcal{F}_{\dot{\zeta}^3}(k_1,-k_2,k_3,\eta_0) + \mathcal{F}_{\dot{\zeta}^3}(k_1,k_2,-k_3,\eta_0) + \text{c.c.}\Big]$, where we have set $a_{k_i} = 1$ and $b_{k_i} = 0.01$. In general b_{k_i} can have some scale-dependence as long as it does not spoil constraints from backreaction of the energy density in the initial state. Whatever choice is made for b_{k_i} should be applied consistently in the case of a joint analysis of the power spectrum and bispectrum.

The volume $\mathcal{V}_T$ includes only those combinations of k_1, k_2, and k_3 that can form a closed triangle, with each wavenumber satisfying $k_{\rm min} \leq k_1,\, k_2,\, k_3 \leq k_{\rm max}$, where $k_{\rm min} = 10^{-3}$ Mpc^{-1} and $k_{\rm max} = 0.1$ Mpc^{-1}. The weight $w(k_1, k_2, k_3)$ is typically taken to be either unity or $1/(k_1 + k_2 + k_3)$, where the latter choice is meant to represent a more accurate reflection of the scaling of the covariance of the CMB bispectrum, such that the Fourier-space cosine is closer to the multipole-space cosine between the CMB bispectra corresponding to the shapes S and S'. However, in general we find that both choices result in similar cosines. In our analysis, we use a unit weight for the 3D Fourier basis fits and all 2D fits, and a weight of $1/(k_1 + k_2 + k_3)$ for the remaining fits.

In this section we implement the existing polynomial and Fourier basis methods, and additionally our new basis of piecewise splines, to obtain separable fits to non-Bunch-Davies shapes. In each case, we quantify the performance of the basis by computing the cosine as a function of an increasing number of modes used in the fit.

We use the polynomial basis described in [16] and the Fourier basis described in [73]. In each case, three 1D functions based on either polynomials or sines/cosines of k_i are multiplied together to form 3D separable functions, that are then orthogonalized using a Gram-Schmidt algorithm to create a basis of 3D orthonormalized and separable functions, called $\{\mathcal{R}_n\}$ for polynomials and $\{\mathcal{F}_n\}$ for Fourier modes. Since the basis functions are orthonormal in each case, the expansion coefficients $\{\alpha_n\}$ can be computed through inner products between S and the basis functions, $\alpha_n = \langle S, \mathcal{R}_n \rangle$ or $\alpha_n = \langle S, \mathcal{F}_n \rangle$.

Alternatively, in the spline basis expansion, three 1D piecewise spline functions are multiplied together to form a basis of 3D separable spline functions

$\{B_n\}$. In this case, each mode is highly localized in a region of Fourier space, so any two modes are orthogonal by construction unless they have peaks that overlap. While the lack of strict orthogonality means that the expansion coefficients in the spline basis cannot be computed using simple inner products, existing algorithms can solve for the coefficients efficiently [210] (also see appendix A.3).

In the subsections that follow, we present polynomial, Fourier, and spline basis fits to a variety of non-Bunch-Davies shapes.

4.4.1 $S_{\rm NBD1}$ and $S_{\rm NBD2}$

The cosines for polynomial, Fourier, and spline fits to these shapes are shown in the upper left panel of Figure 4.2. We find that the polynomial expansion produces a better fit than the Fourier basis, and polynomial cosines typically increase slowly beyond about 100 modes for $S_{\rm NBD1}$ and 50 modes for $S_{\rm NBD2}$, indicating that lower order modes contribute most to the fits. While it is possible that increasing the number of polynomial or Fourier modes will increase the cosines further, during our analysis we found that generating large polynomial and Fourier basis sets is computationally very demanding. The separable modes are orthogonalized using a Gram-Schmidt algorithm, which is known to be numerically unstable, and the instabilities become more severe as higher order polynomials are used. On the other hand, the spline basis expansion as we have implemented it does not require orthogonalization, so in comparison a large number of modes can easily be generated and used in the separable fits. In the case of $S_{\rm NBD1}$ and $S_{\rm NBD2}$, we find that the spline expansion performs similarly well as the polynomial expansion when 200 modes are used, and better

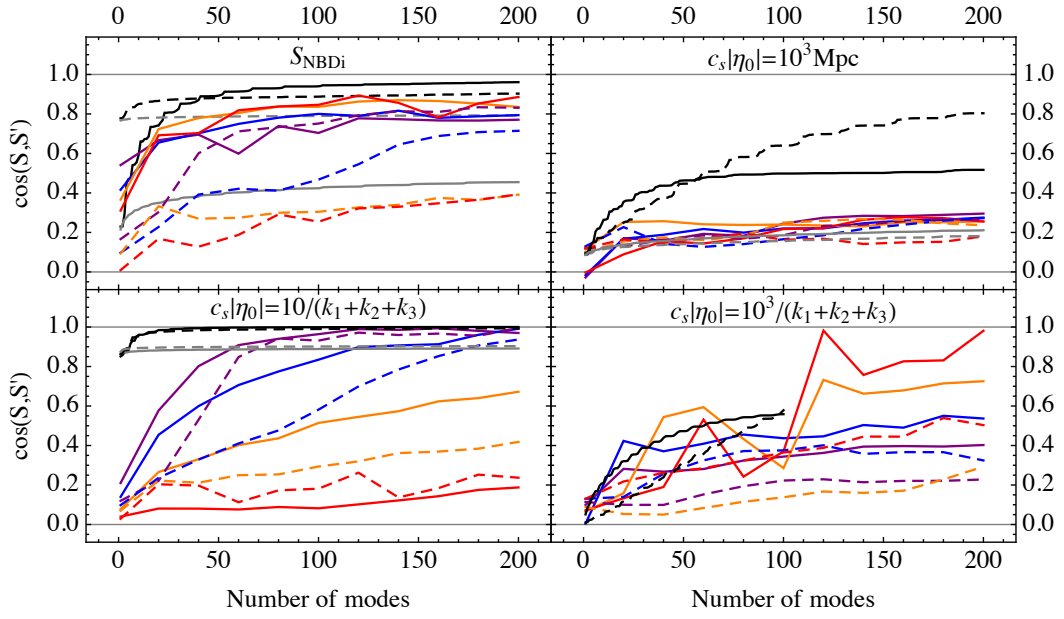


Figure 4.2: Comparisons of shape reconstructions for $S_{\rm NBD1}$ (solid) and $S_{\rm NBD2}$ (dashed) [upper left panel] or $S_{\dot\zeta^3}$ (solid) and $S_{\zeta\dot\zeta^2}$ (dashed) [other three panels] using the polynomial (black), Fourier (gray), and b-spline methods. *Top row and lower left panels:* The 3D spline basis reconstructions correspond to the colored curves, where the purple, blue, orange, and red colors correspond to spline basis sets derived from using 10, 14, 22, and 40 1D spline functions in each dimension. *Lower right panel:* The 2D spline basis reconstructions correspond to the colored curves, where the purple, blue, orange, and red colors correspond to spline basis sets derived from using 50, 100, 200, and 300 1D spline functions in each dimension.

fits can be achieved by using a larger number of modes.

4.4.2 $S_{\dot{\zeta}^3}$, $S_{\zeta\dot{\zeta}^2}$, and $S_{\zeta(\partial\zeta)^2}$ with k-independent $c_s\eta_0$

These shapes have oscillatory features that grow with the size of the triangle (i.e. with k_1) and in the flattened and squeezed limits. With $c_s|\eta_0| = 10^3$ Mpc, the set of 200 polynomial modes produces a cosine of 0.52 for $S_{\dot{\zeta}^3}$ and 0.80 for $S_{\zeta\dot{\zeta}^2}$, as shown in the upper right panel of Figure 4.2.[9] While these modes produce a higher cosine than a set of 200 spline functions, better fits can be generated with a larger set of splines.

4.4.3 $S_{\dot{\zeta}^3}$, $S_{\zeta\dot{\zeta}^2}$, and $S_{\zeta(\partial\zeta)^2}$ with k-dependent $c_s\eta_0$

For small Λ/H, we find that the shapes are easily reconstructed with both the polynomial and spline methods using $\lesssim$ 200 modes, as shown in lower left panel of Figure 4.2. The oscillatory features are of low enough frequency that lower order polynomials are sufficient to capture most of the features of these shapes, and the spline functions also do not need to have a very fine resolution. If the initial conditions are set at $10/k_1$, rather than $10/(k_1 + k_2 + k_3)$, then the oscillatory features are of somewhat higher frequency; for either shape, however, the large global features still allow both the polynomial and spline bases to efficiently reconstruct these shapes.

For cases with k-dependent $c_s\eta_0$ and larger values of Λ/H, the oscillations have a much higher frequency than what can be captured by polynomials, and we find that the spline reconstructions perform similarly poorly. However, the k-dependence of $c_s\eta_0$, whether it be $(\Lambda/H)/(k_1 + k_2 + k_3)$ or $(\Lambda/H)/k_1$, allows

[9] The shape for $S_{\zeta(\partial\zeta)^2}$ is sufficiently similar to $S_{\zeta\dot{\zeta}^2}$ that its fits are not shown separately in the figures.

the shapes to be rewritten as functions of only two free parameters: the ratios $x \equiv k_3/k_1$ and $y \equiv k_2/k_1$. In terms of x and y, the oscillation frequency does not increase drastically throughout the allowed parameter space, which makes it easier to generate good 2D fits. The advantage of rewriting scale-invariant shapes such as these in terms of two free parameters for computing k-space cosines and CMB bispectra has been discussed in earlier works [211].

We generate the 2D fits by defining a 2D analogue of the polynomial/spline basis sets and the cosine. We find, however, that to generate the same number of polynomial modes as in the 3D case (200) requires using higher order 1D polynomials, which exacerbates the numerical issues that we encountered in the 3D case, so for the 2D fits we only use 100 polynomial modes with unit weight. We do not encounter similar numerical issues in the spline fits.

We show the results for the 2D fits in the $c_s|\eta_0| = 10^3/(k_1 + k_2 + k_3)$ case in the lower right panel of Figure 4.2; for $c_s|\eta_0| = 10^3/k_1$ the shapes have very similar features, with the oscillations being slightly more rapid and more difficult to represent in the latter case. 2D spline fits achieve similar cosines as 2D polynomial fits using the same number of modes (~ 100), while also being able to produce higher cosines through the addition of more modes without running into numerical issues, as shown in Figure 4.3.

4.5 Discussion

If the CMB is to be used to its full potential to elucidate the details of inflation, we need to be able to both effectively characterize a wide variety of potential inflationary signatures while also ensuring that this information is accessed in a

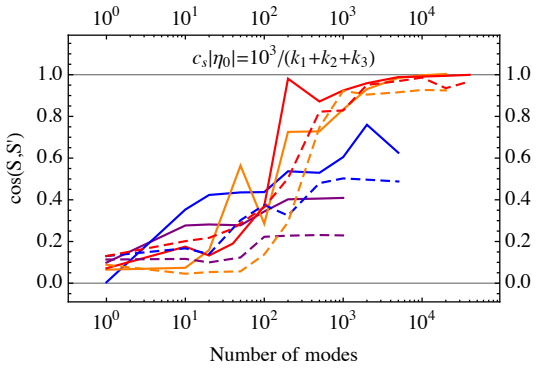


Figure 4.3: 2D b-spline fits to $S_{\dot{\zeta}^3}$ (solid) and $S_{\zeta\dot{\zeta}^2}$ (dashed), where the purple, blue, orange, and red colors correspond to spline basis sets derived from using 50, 100, 200, and 300 1D spline functions in each dimension.

timely manner. For non-Gaussian signatures, present in the CMB bispectrum, a potential bottleneck in this process is the production of separable templates that accurately match the characteristic properties of the underlying shape. This issue is particularly acute for initial states that differ from the Bunch-Davies state, which can generically have highly oscillatory features whose characteristic scale can vary over the bispectrum configuration space.

In this work we have analyzed a variety of functions to generate templates for bispectrum shapes arising in a variety of non-Bunch-Davies scenarios. We have quantified how well different choices of separable basis functions, derived from polynomials, Fourier functions, and b-splines, can reconstruct the original non-Bunch-Davies shapes. The spline expansion method is a new alternative choice of basis, that we implement here for the first time, and can be used at a reasonable computational cost to obtain cosines that can be very close to unity.

We find that the polynomial basis is good for describing some non-Bunch-Davies shapes with large features, and generally performs better than the

Fourier basis. For the rest of the shapes we considered, assuming the low cosines will steadily increase with the addition of more modes, we find that numerical difficulties prevent us from generating enough modes to see higher cosines. For most shapes, the spline basis performs as well as the polynomial one when equal numbers of modes are chosen. The spline basis is both numerically simpler to compute, and does not require orthogonalization due to its localized nature. This allows a larger number of modes to be efficiently calculated to improve the match between the templates and the actual shapes.

The spline basis expansion method is very flexible, and there are other ways to adapt a spline basis to target specific shapes better that we have not explored here. For example, while our spline bases are derived from equally spaced knots, we note that one can create a basis using unequally-spaced knots, such that different k-space regions are sampled more finely than others. Non-Bunch-Davies shapes, where the features are sharply peaked and localized near flattened and squeezed configurations, can potentially be probed more efficiently by optimizing the spline basis in this way.

Finally, while our analysis takes place in primordial k-space, the flexibility and computational simplicity of the spline basis approach translates directly to multipole-space, where it complements existing approaches, such as the polynomial basis. Utilizing a variety of basis expansion techniques ensures that the exquisite CMB data available now, and in the future, can be used efficiently to explore the full theory space of viable inflationary scenarios.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

In the last decade, many theoretical and observational advances have been made towards a fuller understanding of inflation. On the theory side, there are now a wide range of physically-motivated alternatives to simple single-field slow-roll models, each of which has a uniquely non-Gaussian structure in higher-order correlators (e.g. bispectra, or 'shapes'), providing a promising strategy to uncover the key to inflation. Observationally, techniques for extracting constraints on primordial non-Gaussianity from data have improved greatly in order to meet the demands of precision measurements from experiments such as Planck. The overarching goal of this thesis has been to contribute further to these efforts, by better understanding and improving the techniques by which theoretical models are reconstructed and tested with current and future surveys. Looking forward, there are a variety of future directions that can extend the work presented in this thesis.

On the CMB front, Planck's data collection is complete. Two releases have yielded the strongest constraints to date on non-Gaussianity, with no significant detections so far, and a final release is expected in the coming year. One of the major challenges in constraining theories, however, is that the corresponding predicted CMB bispectra are generally not computationally tractable unless the primordial bispectrum can be written as a separable function of three wavenumbers. In order to fully and efficiently probe the allowed space of theoretically motivated models, it is advantageous to have several methods for achieving this. To complement existing methods exploiting separable basis expansions, in Chapter 2 we constructed a separable basis of divergent bispectra, that in con-

trast to basis sets derived from polynomials is able to reconstruct a wide range of primordial shapes, including models with complex squeezed limits (e.g. solid inflation) or wave-like features (e.g. Galileon inflation), with relatively few basis functions. Then, using the divergent basis, we forecasted Planck-like CMB constraints on the bispectrum amplitude of individual models with a fixed k-dependence, including standard templates and more general inflationary models, as well as theoretically-unbiased constraints on the bispectrum that are independent of any particular choice of model. We also quantified constraints on sets of shapes, which can take into account how well competing theories can be distinguished from each other with the data.

Still, it is clear that many theories actively being considered require more precise modeling than what has been achieved so far. For example, realistic bispectra derived from non-Bunch-Davies initial quantum states for the inflaton are very difficult to reconstruct with smooth polynomials because of their strong peaks in both squeezed and flattened triangles, with very high frequency oscillations superimposed over all triangular configurations: Planck constraints on these models relied on a polynomial basis bispectrum reconstruction that achieved only a 60% level of similarity. In Chapter 4, we developed a new basis description built from b-splines that is competitive with, and in select cases, performs better than the polynomial basis used by the Planck analysis, and can reconstruct these shapes to $> 90\%$ fidelity. With these tools in hand, it will be possible to analyze projected constraints on a variety of models that as yet do not have robust separable approximations through other means, starting with generalized non-Bunch-Davies theories. Additionally, feature and resonant models arising from inflaton potentials with sharp features or superimposed oscillations have not yet been fully constrained: most attention on these

models in the literature has been in the limit where oscillations are more rapid than what can be accurately captured by the polynomial basis. The lack of efficient methods for analyzing such shapes motivated other groups to explore yet other bases and strategies for computing CMB bispectra and analyzing CMB data [192, 193], which offers yet another possibility for constraining oscillatory CMB bispectrum signatures. Exploring new methods such as these for testing the full range of inflationary theories will give us the best chances for discovering what drove inflation, and will certainly allow us to make the most out of the exquisite measurements taken by Planck. In the future, when the Planck data is publicly available, analyses and constraints from the data using alternative bases will be able to weigh in on an extended range of models. Future analyses of the Planck data may yield even stronger constraints on the class of oscillatory models: for example, non-Bunch-Davies models generate oscillatory features in not only the bispectrum, but the power spectrum as well, and the amplitude, scale-dependence, and phase of such features should appear consistently in both observables, allowing for stronger discrimination between models than by looking at the power spectrum or bispectrum separately.

To complement CMB constraints on inflation, there are many large-scale structure surveys that are currently in progress or approaching soon: SDSS (BOSS and eBOSS), DES, LSST, DESI, and Euclid. The same primordial, and potentially non-Gaussian, fluctuations that are imprinted on the CMB also leave their signatures on the subsequent distribution and growth of large-scale structure at later times. In the last several years, much effort has been devoted to understanding, both theoretically and verified by N-body simulations, how non-Gaussianity induces changes to the halo mass function and the biasing of dark matter halos with respect to the smooth dark matter fluctuations. With several

new large-scale structure (LSS) surveys coming on-line in the next decade, it is an opportune time for studying how well LSS observables and data will realistically be able to constrain inflationary histories.

In Chapter 3, we forecasted how well DESI and Euclid will be able to extract and distinguish between signatures of general primordial bispectra from forthcoming precision measurements of the galaxy power spectrum. This work improves on existing forecasts by modeling the halo bias more accurately away from the large-scale limit, and taking a more realistic approach to modeling the information that can be extracted from small scales. One of the key results from this study is that the power spectrum down to small quasi-linear scales contains valuable information about the primordial bispectrum—information that is critical to obtaining a strong constraint if the model under consideration does not create a strongly scale-dependent halo bias on large scales (e.g. equilateral type non-Gaussianity). While smaller scales are in theory better measured due to the larger number of modes that fit within the survey volume, in practice there are a number of known astrophysical difficulties with using small scales to constrain physics: non-linear galaxy bias, peculiar velocities, baryonic effects, etc. In preparation for future surveys, improved tools are actively being developed to address these challenges. For example, one strategy is to combine different probes; no observable is without problems, but the strengths of one probe may be able to mitigate the weaknesses of another. Specifically, combining the galaxy power spectrum data with weak lensing data, which is not susceptible to the same systematics, may be able to combine the advantages of both probes, while minimizing their individual weaknesses.

Weak lensing, with its own set of unique systematics, is a wholly differ-

ent observable that is interesting to explore for its potential complementarity to galaxy power spectrum constraints, as well as for its own sake. For example, upcoming surveys with weak lensing components will reach such a high precision that ignoring the effect of baryons on the weak lensing convergence power spectrum can result in biased parameter constraints that are several times larger than the statistical errors of the surveys. However, simulations have shown that the baryonic effects can be recast as changes to the halo mass-concentration relation corresponding to a rearrangement of dark matter within individual halos [212]. Thus jointly constraining the mass-concentration parameters with cosmological parameters of interest is a promising approach to realistically accounting for our imperfect knowledge of baryons, and may even contribute to new knowledge about the galaxy formation process [213]. So far there are no forecasts of weak lensing constraints on non-Gaussianity that include baryonic effects, while forecasts ignoring baryons indicate that anticipated weak lensing data have the potential to strongly constrain primordial bispectra, particularly for those models that are not realistically expected to be strongly constrained by the galaxy power spectrum alone [6]. It may be interesting to explore and obtain more robust forecasts of weak lensing constraints on general inflationary theories. At the same time, the parametrization of baryonic effects clearly has wide-ranging applicability for any constraints that draw their power from small scale structure, such as constraints on massive neutrinos or the growth of structure from modified gravity theories.

Another way that experiments have complementary power when combined, is in the range of scales that they are able to probe effectively. For example, both CMB data and the upcoming galaxy surveys discussed above, probe largely overlapping scales, roughly between $k \sim 0.001 - 1\ h/\mathrm{Mpc}$, but more futuristic

surveys may be able to probe larger and smaller scales than this range. For example, future 21 cm surveys provide a new exciting opportunity to both harness a new cosmological probe for constraining fundamental physics, as well as to see into the reionization epoch and the dark ages, about which we know relatively little. The redshifted 21 cm line from neutral hydrogen can shed light on how the first luminous objects formed, and also act as a tracer of the underlying matter distribution, allowing independent observational constraints on fundamental physics, such as the physics of inflation. For these purposes, 21 cm has strong advantages in theory: LSS, such as the matter power spectrum, can be observed down to smaller scales than with any other probe, with most of the observable range remaining well-described by linear perturbation theory, and the volumes observed will be larger as well. Probing both larger and smaller scales will allow observations of more triangle configurations of the primordial bispectrum than what will be achieved with other CMB and LSS surveys. Though the practical challenges of getting at the cosmological information are immense, current experiments, like PAPER, LOFAR, and MWA, that are aimed at first detecting the 21 cm signal from reionization, are paving the way for future surveys, including CHIME and SKA, that will provide very competitive cosmological constraints on, e.g. dark energy and inflation, assuming the systematics can be sufficiently understood. In particular, a large part of the methods in Chapter 3 for forecasting optical galaxy survey constraints on inflation are directly applicable to studying the constraints that 21 cm galaxy surveys are likely to yield.

APPENDIX A

CHAPTER 1 OF APPENDIX

This appendix provides derivations and details that supplement the discussions in Chapter 4 of bispectra from non-Bunch-Davies initial states. In Appendix A.1, we reproduce the calculation in (cite refs 10 & 11 from the NBD paper), which shows that general initial states for the inflaton, set at a finite scale-dependent time in the past, $\eta(k)$, lead to oscillatory features in the primordial power spectra. In Appendix A.2, we show the full bispectrum expressions that are derived from adopting a general action and allowing general initial states, and focus only on the shapes that are leading-order in the slow-roll parameters: $\mathcal{F}_{\dot{\zeta}^3}$, $\mathcal{F}_{\zeta\dot{\zeta}^2}$, and $\mathcal{F}_{\zeta(\partial\zeta)^2}$. We derive the limits of these bispectra in squeezed and flattened configurations, showing that their amplitudes are strongly enhanced for these triangles. In Appendix A.3, we discuss the b-spline shape-fitting algorithm in detail, with 1D and 2D example fits, and by showing the Mathematica code we have used to calculate the shape reconstructions discussed in Chapter 4.

A.1 Power spectrum for general initial states

In this appendix we show that the result for the Green's function for general initial states in section 4.2 with a scale-dependent initial time $\eta_0(k)$ gives the correct form of the late-time power spectrum in [180, 181]. We start with developing a precise definition for the adiabatic vacuum at $\eta_0 \to -\infty$ or the Bunch-Davies vacuum. Let us write the field $\chi_{\vec{k}}(\eta)$ (defined via $\zeta = \frac{1}{\sqrt{2\varepsilon}}\frac{c_s}{a}\chi$) in terms of annihilation and creation operators at the time η_0, $a_{\vec{k}}(\eta_0)$ and $a^{\dagger}_{-\vec{k}}(\eta_0)$, and the mode

functions $f_k^{\gtrless}(\eta)$, as

$$\chi_{\vec{k}}(\eta) \;=\; a_{\vec{k}}(\eta_0) f_k^{>}(\eta) + a^{\dagger}_{-\vec{k}}(\eta_0) f_k^{<}(\eta)\,. \tag{A1}$$

The conjugate momentum, $\pi_{\vec{k}}(\eta)$ can be obtained from the quadratic Lagrangian for $\chi_{\vec{k}}(\eta)$; at leading order this is given by

$$\pi_{\vec{k}}(\eta) \;=\; \frac{\partial \mathcal{L}^{(2)}}{\partial \dot{\chi}} \;=\; \frac{\mathrm{d}\chi_{\vec{k}}}{\mathrm{d}\eta} - \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}\eta}\chi_{\vec{k}}\,, \tag{A2}$$

We can write $\pi_{\vec{k}}(\eta)$ in terms of its corresponding mode functions $g_k^{\gtrless}(\eta)$ as

$$\pi_{\vec{k}}(\eta) \;=\; -i\left(a_{\vec{k}}(\eta_0) g_k^{>}(\eta) - a^{\dagger}_{-\vec{k}}(\eta_0) g_k^{<}(\eta)\right)\,. \tag{A3}$$

For Bunch-Davies modes we choose the positive frequency solution at early times, and using $a(\eta) = -1/(\eta H)$ at leading order during inflation, the mode functions are given by

$$f_k^{>}(\eta) \;=\; \frac{1}{\sqrt{2c_s k}} e^{-ic_s k\eta}\left(1 - \frac{i}{c_s k\eta}\right)\,, \tag{A4}$$

$$g_k^{>}(\eta) \;=\; \sqrt{\frac{c_s k}{2}}\, e^{-ic_s k\eta}\,. \tag{A5}$$

The choice of Bunch-Davies vacuum can then be expressed as the following relationship between the field and its conjugate momentum in the infinite past,

$$\pi_{\vec{k}}(\eta_0) \;=\; (-ic_s k)\chi_{\vec{k}}(\eta_0)\,. \tag{A6}$$

Note that this does not imply that the position and momentum operators commute at all times; it is merely a statement of how they are related at $\eta_0 \to -\infty$. Equivalently, in terms of the mode function $f_k^{>}(\eta)$ we can write

$$\left.\frac{\mathrm{d}f_k^{>}}{\mathrm{d}\eta}\right|_{\eta=\eta_0} - \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}\eta} f_k^{>}(\eta_0) \;=\; (-ic_s k) f_k^{>}(\eta_0)\,, \tag{A7}$$

for $\eta_0 \to -\infty$. The above condition is the definition of the adiabatic vacuum in the infinite past, or equivalently the Bunch-Davies vacuum. As shown in [214, 180] this choice corresponds to a minimum uncertainty state.

The prescription to choose an adiabatic vacuum at a finite initial time is to enforce the same condition in eq. (A7) at a given η_0. This corresponds to a state which minimizes the uncertainty at $\eta = \eta_0$. Choosing the initial density matrix to be unity, but still having a non-Bunch-Davies initial state by allowing $\beta_k^> \neq 0$, i.e. with $f_k^>(\eta)$ given by eq. (4.10) with $A_k = 0$, this leads to the following relation between the Bogoliubov coefficients,

$$\frac{\beta_k^>}{\alpha_k^>} = \left(\frac{i}{2c_s k\eta_0 + i}\right) e^{-2ic_s k\eta_0} . \tag{A8}$$

Combining this with the usual condition $\left|\alpha_k^>\right|^2 - \left|\beta_k^>\right|^2 = 1$ we find that

$$\left|\alpha_k^>\right|^2 = \frac{4c_s^2k^2\eta_0^2 + 1}{4c_s^2k^2\eta_0^2} . \tag{A9}$$

Let us now choose η_0 to be a function of k such that the physical momentum $c_s k/a(\eta_0)$ crosses some fixed high energy scale Λ of new physics at η_0, then

$$\eta_0 = -\frac{\Lambda}{Hc_s k} . \tag{A10}$$

With the above equations we can write the following solution for $\alpha_k^>$ and $\beta_k^>$ (note that we are free to choose any overall phase),

$$\alpha_k^> = \left(\frac{2\Lambda/H - i}{2\Lambda/H}\right) e^{-i\Lambda/H} , \tag{A11}$$

$$\beta_k^> = -\left(\frac{i}{2\Lambda/H}\right) e^{i\Lambda/H} . \tag{A12}$$

For $\Lambda/H \gg 1$ we can now write the late-time power spectrum as

$$\begin{aligned} \mathcal{P}_\zeta(k) &= \frac{k^3}{2\pi^2}\frac{c_s^2}{2\varepsilon}\frac{1}{a^2(\eta)} f_k^>(\eta) f_k^<(\eta)\Big|_{\eta\to 0^-} \\ &= \frac{H^2}{8\pi^2\varepsilon c_s}\left[1 - \frac{H}{\Lambda}\sin\left(\frac{2\Lambda}{H}\right)\right]\Bigg|_{c_s k = aH} , \end{aligned} \tag{A13}$$

which includes a scale-dependent oscillatory term. The argument of the oscillations is usually written as being proportional to $\ln(k/k_p)$, where k_p is some fixed

pivot scale. This can be seen by expanding H around the pivot scale so that $H(k) \approx H(k_p)\left[1-\varepsilon(N-N_p)+\ldots\right] \approx H(k_p)\left[1-\varepsilon \ln \frac{k}{k_p}+\ldots\right]$, leading to logarithmic oscillations in the power spectrum. Note that this expansion only holds for small ε and a reasonable range of scales.

A.2 Bispectrum for general initial states

The functions $\mathcal{F}_{\dot{\zeta}^3}$, $\mathcal{F}_{\zeta\dot{\zeta}^2}$, and $\mathcal{F}_{\zeta(\partial\zeta)^2}$ in eqs. (4.17) - (4.19) are given by

$$\mathcal{F}_{\dot{\zeta}^3}(p_1,p_2,p_3,\eta_0) = \frac{1}{k_1k_2k_3}\left[-\frac{2}{K_1^3}+\frac{e^{ic_sK_1\eta_0}}{K_1}\left(\frac{2}{K_1^2}-\frac{2ic_s\eta_0}{K_1}-c_s^2\eta_0^2\right)\right], \tag{B1}$$

$$\begin{aligned}\mathcal{F}_{\zeta\dot{\zeta}^2}(p_1,p_2,p_3,\eta_0) = \frac{1}{k_1k_2k_3}\Bigg[&\frac{-2K_1^3K_3^3+K_1^2K_2^4+K_1K_2^2K_3^3}{K_1^3K_3^6}\\ &+\frac{e^{ic_sK_1\eta_0}}{K_1^3K_3^6}\left(2K_1^3K_3^3-K_1^2K_2^4-K_1K_2^2K_3^3+ic_sK_1^2K_2^2K_3^3\eta_0\right)\Bigg],\end{aligned} \tag{B2}$$

and

$$\begin{aligned}\mathcal{F}_{\zeta(\partial\zeta)^2}(p_1,p_2,p_3,\eta_0) = \frac{1}{(k_1k_2k_3)^3}\Bigg[&\frac{K_1^6-3K_1^4K_2^2-K_1^3K_3^3+2K_1^2K_2^4+2K_1K_2^2K_3^3}{K_1^3}\\ &+\frac{e^{ic_sK_1\eta_0}}{c_sK_1^3\eta_0}\Big\{i\left(K_1^5-2K_1^3K_2^2\right)+c_s\left(K_1^4K_2^2+K_1^3K_3^3-2K_1^2K_2^4-2K_1K_2^2K_3^3\right)\eta_0\\ &-ic_s^2\left(K_1^4K_3^3-2K_1^2K_2^2K_3^3\right)\eta_0^2\Big\}\Bigg],\end{aligned} \tag{B3}$$

where for brevity of notation we have suppressed the explicit momentum dependence of the functions K_1, K_2, and K_3,

$$K_1(p_1,p_2,p_3) = p_1+p_2+p_3\,, \tag{B4}$$

$$K_2(p_1,p_2,p_3) = (p_1p_2+p_2p_3+p_3p_1)^{1/2}\,, \tag{B5}$$

$$K_3(p_1,p_2,p_3) = (p_1p_2p_3)^{1/3}\,. \tag{B6}$$

In the next two subsections we show how one obtains the flattened and squeezed enhancements from the functions $\mathcal{F}$ (also see [65]). We work with the simplest function $\mathcal{F}_{\dot{\zeta}^3}$, though the results are similar for the other two functions (or at least for their appropriate sum) as well.

A.2.1 Flattened limit

The $b_{k_1}a_{k_2}a_{k_3}$ term leads to an enhanced flattened limit ($k_1 \approx k_2 + k_3$, k_1 being the largest momentum mode) bispectrum. Let us first assume that b_{k_1} is purely imaginary, so that the bispectrum is proportional to the imaginary part of $\mathcal{F}$. The corresponding $\mathcal{F}$ function we consider is,

$$\mathcal{F}_{\dot{\zeta}^3}(-k_1, k_2, k_3, \eta_0) \;=\; -\frac{2}{\tilde{k}_1^3 k_1 k_2 k_3} \;+\; \frac{e^{ic_s\tilde{k}_1\eta_0}}{\tilde{k}_1 k_1 k_2 k_3}\left(\frac{2}{\tilde{k}_1^2} - \frac{2ic_s\eta_0}{\tilde{k}_1} - c_s^2\eta_0^2\right), \qquad \text{(B7)}$$

where $\tilde{k}_1 \equiv -k_1 + k_2 + k_3$. In the limit of $\tilde{k}_1 \to 0$, the exponential can be expanded as

$$\lim_{\tilde{k}_1\to 0} e^{ic_s\tilde{k}_1\eta_0} \;=\; 1 + ic_s\tilde{k}_1\eta_0 - \frac{c_s^2\tilde{k}_1^2\eta_0^2}{2} - \frac{ic_s^3\tilde{k}_1^3\eta_0^3}{6} + \ldots \qquad \text{(B8)}$$

Using this in eq. (B7), and noticing that any term with $\tilde{k}_1$ in the numerator goes to zero, we find that

$$\lim_{\tilde{k}_1\to 0} \mathcal{F}_{\dot{\zeta}^3}(-k_1, k_2, k_3, \eta_0) \;=\; -\frac{i}{3(k_2+k_3)k_2k_3} c_s^3\eta_0^3\,. \qquad \text{(B9)}$$

For fixed η_0, we can set $c_s|\eta_0| = 1/k_*$. The above limit of the three-point function is therefore enhanced (though not divergent) in the flattened limit. For $\eta_0(k)$ with large Λ/H as well we see an enhancement in the flattened limit. If we instead

assume that b_{k_1} is real, then the bispectrum is still enhanced, though not in the exactly flattened limit but in a near-flattened limit.

A.2.2 Squeezed limit

Let us now look at the $a_{k_1} b_{k_2} a_{k_3}$ piece,

$$\mathcal{F}_{\dot{\zeta}^3}(k_1, -k_2, k_3, \eta_0) \;=\; -\frac{2}{\tilde{k}_2^3 k_1 k_2 k_3} \;+\; \frac{e^{ic_s \tilde{k}_2 \eta_0}}{\tilde{k}_2 k_1 k_2 k_3}\left(\frac{2}{\tilde{k}_2^2} - \frac{2ic_s\eta_0}{\tilde{k}_2} - c_s^2\eta_0^2\right), \quad \text{(B10)}$$

where $\tilde{k}_2 \equiv k_1 - k_2 + k_3$. In the squeezed limit ($k_3 \ll k_1 \approx k_2$) we have $\tilde{k}_2 \to k_{\rm min}$, where $k_{\rm min}$ is the smallest momentum mode observable today. Using this in eq. (B10) we find that

$$\lim_{\tilde{k}_2 \to k_{\rm min}} \mathcal{F}_{\dot{\zeta}^3}(k_1, -k_1, k_3, \eta_0) \;=\; -\frac{2}{k_1^2 k_{\rm min}^4} \;+\; \frac{e^{ic_s k_{\rm min}\eta_0}}{k_1^2 k_{\rm min}^2}\left(\frac{2}{k_{\rm min}^2} - \frac{2ic_s\eta_0}{k_{\rm min}} - c_s^2\eta_0^2\right). \quad \text{(B11)}$$

For fixed η_0 and in the limit of $k_{\rm min} \gg k_*$, the $c_s^2\eta_0^2$ term gives the largest contribution. This term is multiplied with a highly oscillatory function though, and averaging over the large argument of the cosine (real part of the exponential) we expect its contribution to vanish. The leading order contribution is then proportional to $1/(k_1^2 k_{\rm min}^4)$, which shows a strong squeezed limit enhancement. In the limit of $k_{\rm min} \gtrsim k_*$ or for $\eta_0(k)$ this argument no longer holds and we may or may not see enhancements.[1]

A.3 B-splines fitting algorithm

In this appendix, we build on the discussion of section 4.3 to describe in more detail the b-spline fitting algorithms we have used, and illustrate b-spline fitting

[1] In our fits with fixed η_0 in section 4.4 we took k_* to be similar to $k_{\rm min}$ ($\sim 10^{-3}$ Mpc^{-1}) which is what the Planck team had used in [117].

examples for a simple 1D function and a 2D representation of the scale-invariant enfolded template. Snippets of Mathematica 10 codes we have implemented are shown here, and the same algorithms were first constructed and explicitly shown as Matlab code in [210].

Our first example considers a spline fit to a simple 1D function $f(x)$, with $0 \leq x \leq 1$. To generate a fit, we first make two choices: a choice of basis and a choice of data points to fit. The spline basis is determined by a choice of $k+1$ equidistant knots, $\{x_0, x_1, \ldots, x_k\}$, at which each basis function's degree q polynomial pieces will be joined. The b-splines in general do not have to be generated using constant knot intervals, but for simplicity we always start with equidistant knots, and additionally include q extra knots at each of $x = 0$ and $x = 1$ to produce a "clamped" basis. Without these extra knots, the generated basis sets do not have splines with non-zero amplitudes at $x = 0$ and $x = 1$, and it will be difficult to fit functions that are non-zero at the endpoints.

Given these inputs, existing codes, such as Mathematica's `BSplineBasis` function, can recursively generate a basis of $k + q$ b-splines such that each one is spanned by $q + 2$ knots, made up of $q + 1$ polynomial pieces, with derivatives continuous up to order $q - 1$. In addition, the sum of all b-spline amplitudes at any x is unity. For example, the b-splines in Figure 4.1 are easily generated by choosing $q = 3$ and $k = 3$ such that the knots vector is `knots = {0,0,0,0,1/3,2/3,1,1,1,1}`, and executing `BSplineBasis[{q,knots},i,x]`, where i is an integer $0 \leq i \leq k + q - 1$ identifying each particular b-spline. In this work, we vary the number and widths of the b-splines by varying k, but always keep the degree fixed to $q = 3$.

Next, the choice of data points depends on how finely we wish to sample

$f(x)$. We would like to choose a data set consisting of M data pairs, (x_i, f_i), where $f_i \equiv f(x_i)$, such that our data resolves any potentially fine features in the function we would like to fit, without including so many extra data points that our numerical calculation becomes intractable. The final fit should ultimately be insensitive to the sampling we have chosen. Again, for simplicity, we always use equidistant sampling points x_i in our analysis, but vary the density of sampling points by changing M.

The spline fit then approximates the original function as

$$f'(x) \;=\; \sum_{n=0}^{N-1} \alpha_n B_n(x)\,, \tag{C1}$$

where the expansion coefficients $\{\alpha_n\}$ are solved for by minimizing the least-squares function,

$$\mathrm{LS} \;=\; \sum_{i=1}^{M} \left(f_i - \sum_{n=0}^{N-1} \alpha_n B_n(x_i) \right)^2 . \tag{C2}$$

This requires solving the linear system of equations given by $B^T B\vec{\alpha} = B^T \vec{f}$, where B is an $M \times N$ matrix containing the values of $B_n(x_i)$, and is easily performed with algorithms such as Mathematica's `LinearSolve`.

We note one extension of b-splines, called p-splines, that aims to avoid overfitting a set of input data by imposing smoothness on the resulting fit. Short for "penalty b-splines", in the p-spline method, the fit's expansion coefficients are determined by both the choice of basis and the choice of input data, plus a choice of penalty function that generally disfavors fits with large differences between coefficients of neighboring b-splines [215]. In this context one would instead minimize

$$\mathrm{LS} \;=\; \sum_{i=1}^{M} \left(f_i - \sum_{n=0}^{N-1} \alpha_n B_n(x_i) \right)^2 + \lambda \sum_{j=k}^{N-1} \left(\Delta^k \alpha_j \right)^2 , \tag{C3}$$

where λ is a constant that controls the smoothness of the fit and k is the order of the penalty, a typical choice being $k = 2$, such that $\Delta^2\alpha_j = \alpha_j - 2\alpha_{j-1} + \alpha_{j-2}$. The use of a penalty is optional, and its main purpose in the context of data fitting is to avoid fitting any noisy features in the data. Further, if there are not sufficiently many data points sampling $f(x)$, with many more splines than data points, then without a penalty the fits may display spurious features, as we will deliberately try to show in the 1D example that follows.

In Figure A.1 we show fits to $f(x) = \sin(10x)/(10x)$ with different choices of data points and smoothing parameter λ. We have fixed the knots at $\{0, 0, 0, 0, 0.1, 0.2, \ldots, 0.8, 0.9, 1, 1, 1, 1\}$, yielding a basis of 13 b-splines. We find, as illustrated in the figure, that we can achieve good fits without introducing extra smoothing through a non-zero value of λ, as long as we fit to enough data points. So we now continue to an example of fitting a primordial shape in two dimensions, without a penalty.

To illustrate the b-spline fitting algorithm in two dimensions, we construct a basis of 2D b-splines and use it to fit the scale-invariant enfolded template,

$$S_{\rm enf}(x, y) = \frac{1}{xy}(1 - x - y - x^2 - y^2 + x^3 + y^3 - x^2y - xy^2 + 3xy), \qquad \text{(C4)}$$

where $x \equiv k_3/k_1$ and $y \equiv k_2/k_1$. As in the 1D case, the inputs to the fitting algorithms are made up of a choice of basis and a set of data points. The basis is specified by a choice of polynomial degree and a sequence of knots in each of the two dimensions, x and y. In our particular application, since we are aiming to fit shape functions that are symmetric in their wavenumber arguments, we only specify the knots and degree in one dimension, and use the same b-spline basis for the additional second dimension. The 2D b-splines are then made of products of any two 1D splines, for example, $B_n(x)B_m(y)$. The data are given by

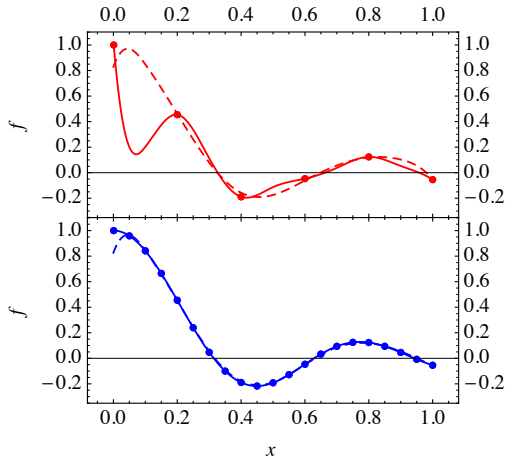


Figure A.1: Example 1D b-spline fits to $f(x) = \sin(10x)/(10x)$. All fits shown have used a basis of 13 splines. In the upper panel, the fits have been computed using 6 data points, sampled at $x_i = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Attempting to construct such a fit without smoothing [solid red] causes the splines to produce spurious features, especially at smaller values of x, while introducing a penalty and a small amount of smoothing [dashed red], $\lambda = 0.1$, restores the fit to a reasonable representation of the true function. The lower panel is the same fit, except with 21 data points, sampled at $x_i = \{0, 0.05, 0.1, \ldots, 0.9, 0.95, 1\}$. In this case, the fits with [dashed blue] and without [solid blue] smoothing are very similar.

$(x_i, y_j, S(x_i, y_j))$ and stored in $Y_{ij} = S(x_i, y_j)$, where again some care must be taken in the choice of sampling, which must be dense enough to capture any small features such as oscillations that we would like to capture in the resulting fit.

The 2D analogue of the least-squares function in eq. C2 is

$$\mathrm{LS} = \sum_{i=1}^{M}\sum_{j=1}^{M}\left(Y_{ij} - \sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\alpha_{mn}B_m(x_i)B_n(y_j)\right)^2 . \tag{C5}$$

To turn the problem of solving for the expansion coefficients α_{mn} into a linear system, we create a regression basis C from the $M \times N$ b-spline basis matrices in

each dimension, B_1 and B_2, which in our case are equal. We define

$$C \quad = \quad (B_2 \otimes e_L^T) \odot (e_K^T \otimes B_1) \equiv B_1 \square B_2 \,, \tag{C6}$$

where $\otimes$ is the Kronecker product, $\odot$ is an element-by-element multiplication, the second equality defines the $\square$ operation, and e_L and e_K are vectors of 1's with length L each. In Mathematica, we define the $\square$ operation as `box`:

```
box[B1_,B2_]:=Module[{K,L,eK,eL},
K=Length[B1[[1,All]]];
L=Length[B2[[1,All]]];
eK=ConstantArray[1,{K}];
eL=ConstantArray[1,{L}];
KroneckerProduct[B2,{eL}]*KroneckerProduct[{eK},B1]
]
```

Then by stacking the columns of the coefficients array α_{mn} and the data array Y_{ij} to get vectors $\vec{\beta}$ and $\vec{y}$ respectively, the task of finding a solution for the coefficients is once again reduced to solving a linear system of equations given by $C^T W C\vec{\beta} = C^T W\vec{y}$. Here W is a matrix containing weights, which may be different for each data point, but for simplicity we restrict ourselves to using a weight of unity for all of our data.

For modest amounts of data and numbers of b-splines, one can quickly solve for the coefficients in this straightforward way. For large data sets and numbers of b-splines, however, this approach becomes computationally cumbersome due to the large size of C. While it is still possible to numerically solve for the coefficients α_{mn} using a low-level language like C(++) or Fortran, we have instead used algorithms developed for higher level languages, such as Matlab

in [210], using only vector and matrix operations. This has the benefit of being easier to implement, while still being able to sidestep much of the memory storage and speed issues typical of a more brute-force approach in a high-level language. Instead of starting with a calculation of C in the brute-force approach, the algorithm from [210] that we have adopted computes $C^T WC$ and $C^T W\vec{y}$ using only B. We refer the reader to [210] for a detailed discussion of how the method itself is devised and constructed, or to see the equivalent Matlab code, and present here an implementation of the b-spline fitting algorithms in Mathematica.

The normal equations can be efficiently constructed and solved, given an input of data in Y and information about the data sampling and b-spline basis in B:

```
get2dfit[Y_,B_]:=Module[{m,n,W,R,r,F,a,A},
m=Length[Y[[1,All]]];
n=Length[B[[1,All]]];
W=ConstantArray[1,{m,m}];
R=Transpose[B].(W*Y).B;
r=ArrayReshape[R,{n*n,1}];
F=Transpose[box[B,B]].W.box[B.B];
F=ArrayReshape[F,{n,n,n,n}];
F=TensorTranspose[F,Cycles{{3,2}}]];
F=ArrayReshape[F,{n*n,n*n}];
a=LinearSolve[F,r];
A=ArrayReshape[a,{n,n}]
]
```

After the matrix of coefficients is solved for, we construct the final fit through two steps. First, we map the coefficients output as A from `get2dfit` to a new set of coefficients that corresponds to a 2D basis of splines which is symmetric in its two arguments, so that each 2D basis mode is a sum of up to two terms: $B_i(x)B_j(y) + B_j(x)B_i(y)$. Second, in building up the fit, mode by mode, we start with the modes that contribute most to the fit. Since the b-splines in a choice of basis have similar shapes and amplitudes, we use the magnitude of the A_{ij} coefficient as a proxy for gauging how much any particular b-spline contributes to a fit's overall cosine with the original shape. This motivates building up a fit by adding in modes, starting with those that have the largest $|A_{ij}|$. The cosine is then computed in the usual way, through an inner product over (x, y)-space between the original shape S_{enf} and the fit.

As an example, we use a basis of 10 splines in each dimension constructed by choosing $k = 7$ and $q = 3$, and use as our data set a grid of uniformly spaced (x, y) values from taking 50 samples in each dimension, to compute the matrix of coefficients A_{ij}, using the algorithms `box` and `get2dfit` above. The total number of symmetric modes is then 55, and the modes are ordered by their corresponding values of largest to smallest $|A_{ij}|$ to produce the cosines in Figure A.2. A visual comparison of the full fit using 55 modes and the original enfolded template is given in Figure A.3.

For our 3D fits, the fitting method is the same: a choice of b-spline basis and data set make up the inputs to the algorithm, which returns an array A_{mnp} containing the expansion coefficients that approximate the input shape as $S'(k_1, k_2, k_3) = \sum_{mnp} A_{mnp} B_m(k_1)B_n(k_2)B_p(k_3)$. However, due to the higher dimensionality of the problem, we must introduce a new function, `rho`, to generalize

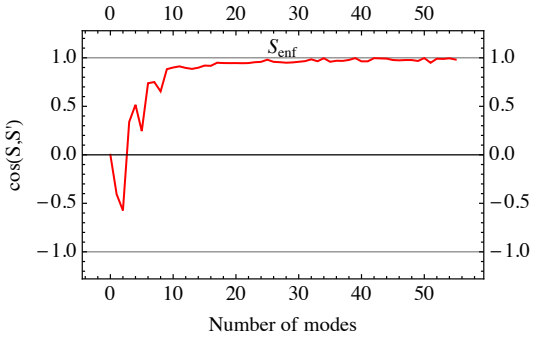


Figure A.2: Cosines between the enfolded shape and the 2D b-spline fits generated by a basis with 10 splines per dimension.

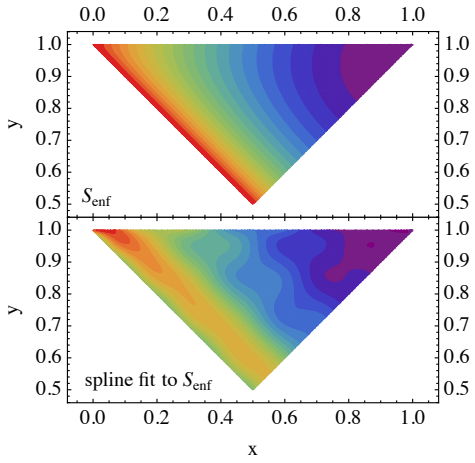


Figure A.3: The original template $S_{enf}(x, y)$ [upper panel] and its b-spline fit using a set of 55 modes [lower panel]. The amplitudes range from 0 (violet) to 1 (red) in both panels.

the matrix product to the product of a matrix and a 3D array. Below, we show Mathematica code for `rho[A,B,p]`, which computes the normal matrix product between rows of A and the p^{th} column of B, resulting in a product, C, which has the same dimensions as B:

```
rho[A_,B_,p_]:=Module[{sa,sb,n,ip,cycles,sbip,prodsbip,
  tempB,C},
```

```
sa=Dimensions[A];
sb=Dimensions[B];
n=Length[sb];
ip=Join[Range[p+1,n],Range[1,p-1]];
Which[
  p==1,cycles=Cycles[{}],
  p==2,cycles=Cycles[{{1,3,2}}],
  p==3,cycles=Cycles[{{1,2,3}}]];
tempB=TensorTranspose[B,cycles];
sbip=sb[[ip]];
prodsbip=Product[sbip[[i]],{i,1,Length[sbip]}];
tempB=ArrayReshape[tempB,{sb[[p]],prodsbip}];
C=Transpose[A].tempB;
C=ArrayReshape[C,Join[{sa[[2]]},sb[[ip]]]];
C=TensorTranspose[C,InversePermutation[cycles]]
]
```

Given this definition of `rho`, the 3D b-spline fit coefficients are calculated using `get3dfit`:

```
get3dfit[Y_,B_]:=Module[{m,n,W,F,R,A},
m=Length[Y[[1,1,All]]];
n=Length[B[[1,All]]];
W=ConstantArray[1,{m,m,m}];
F=rho[box[B,B],W,1];
F=rho[box[B,B],F,2];
F=rho[box[B,B],F,3];
```

```
R=rho[B,Y*W,1];
R=rho[B,R,2];
R=rho[B,R,3];
F=ArrayReshape[F,{n,n,n,n,n,n}];
F=TensorTranspose[F,Cycles[{{3,2,4,5}}]];
F=ArrayReshape[F,{n^3,n^3}];
A=LinearSolve[F,ArrayReshape[R,{n^3,1}]];
A=ArrayReshape[A,{n,n,n}]
]
```

BIBLIOGRAPHY

[1] J. Fergusson and E. Shellard, "The shape of primordial non-Gaussianity and the CMB bispectrum," Phys.Rev., vol. D80, p. 043510, 2009.

[2] D. G. Figueroa, R. R. Caldwell, and M. Kamionkowski, "Non-Gaussianity from Self-Ordering Scalar Fields," Phys.Rev., vol. D81, p. 123504, 2010.

[3] D. Baumann, "TASI Lectures on Inflation," 2009.

[4] E. Komatsu, "Hunting for Primordial Non-Gaussianity in the Cosmic Microwave Background," Class.Quant.Grav., vol. 27, p. 124010, 2010.

[5] A. Orsi, C. Baugh, C. Lacey, A. Cimatti, Y. Wang, et al., "Probing dark energy with future redshift surveys: A comparison of emission line and broad band selection in the near infrared," 2009.

[6] T. Giannantonio, C. Porciani, J. Carron, A. Amara, and A. Pillepich, "Constraining primordial non-Gaussianity with future galaxy surveys," Mon.Not.Roy.Astron.Soc., vol. 422, pp. 2854–2877, 2012.

[7] M. Levi et al., "The DESI Experiment, a whitepaper for Snowmass 2013," 2013.

[8] F. Schmidt and M. Kamionkowski, "Halo Clustering with Non-Local Non-Gaussianity," Phys.Rev., vol. D82, p. 103002, 2010.

[9] P. Ade et al., "Planck 2015 results. XIII. Cosmological parameters," 2015.

[10] L. Anderson et al., "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples," Mon.Not.Roy.Astron.Soc., vol. 441, no. 1, pp. 24–62, 2014.

[11] G. Hinshaw et al., "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," Astrophys.J.Suppl., vol. 208, p. 19, 2013.

[12] A. H. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," Phys.Rev., vol. D23, pp. 347–356, 1981.

[13] A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems," Phys.Lett., vol. B108, pp. 389–393, 1982.

[14] A. Albrecht and P. J. Steinhardt, "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking," Phys.Rev.Lett., vol. 48, pp. 1220–1223, 1982.

[15] P. Ade et al., "Planck 2015 results. XVII. Constraints on primordial non-Gaussianity," 2015.

[16] J. Fergusson, M. Liguori, and E. Shellard, "General CMB and Primordial Bispectrum Estimation I: Mode Expansion, Map-Making and Measures of fNL," Phys.Rev., vol. D82, p. 023502, 2010.

[17] J. Byun and R. Bean, "Non-Gaussian Shape Recognition," JCAP, vol. 1309, p. 026, 2013.

[18] J. Byun and R. Bean, "Non-Gaussian Shape Discrimination with Spectroscopic Galaxy Surveys," JCAP, vol. 1503, no. 03, p. 019, 2015.

[19] J. Byun, N. Agarwal, R. Bean, and R. Holman, "Looking for non-Gaussianity in all the right places: A new basis for non-separable bispectra," 2015.

[20] C. Bennett, D. Larson, J. Weiland, N. Jarosik, G. Hinshaw, et al., "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results," 2012.

[21] G. Hinshaw, D. Larson, E. Komatsu, D. Spergel, C. Bennett, et al., "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," 2012.

[22] K. Story, C. Reichardt, Z. Hou, R. Keisler, K. Aird, et al., "A Measurement of the Cosmic Microwave Background Damping Tail from the 2500-square-degree SPT-SZ survey," 2012.

[23] J. L. Sievers, R. A. Hlozek, M. R. Nolta, V. Acquaviva, G. E. Addison, et al., "The Atacama Cosmology Telescope: Cosmological parameters from three seasons of data," 2013.

[24] A. G. Sanchez, C. Scoccola, A. Ross, W. Percival, M. Manera, et al., "The

clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the large-scale two-point correlation function," 2012.

[25] A. J. Ross, W. J. Percival, A. Carnero, G.-b. Zhao, M. Manera, et al., "The Clustering of Galaxies in SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Constraints on Primordial Non-Gaussianity," 2012.

[26] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," JHEP, vol. 0803, p. 014, 2008.

[27] S. Weinberg, "Effective Field Theory for Inflation," Phys.Rev., vol. D77, p. 123541, 2008.

[28] L. Senatore and M. Zaldarriaga, "The Effective Field Theory of Multifield Inflation," JHEP, vol. 1204, p. 024, 2012.

[29] N. Bartolo, M. Fasiello, S. Matarrese, and A. Riotto, "Large non-Gaussianities in the Effective Field Theory Approach to Single-Field Inflation: the Bispectrum," JCAP, vol. 1008, p. 008, 2010.

[30] J. Garriga and V. F. Mukhanov, "Perturbations in k-inflation," Phys.Lett., vol. B458, pp. 219–225, 1999.

[31] R. Bean, S. E. Shandera, S. Henry Tye, and J. Xu, "Comparing brane inflation to WMAP," JCAP, vol. 0705, p. 004, 2007.

[32] D. Polarski and A. A. Starobinsky, "Isocurvature perturbations in multiple inflationary models," Phys.Rev., vol. D50, pp. 6123–6129, 1994.

[33] J. Garcia-Bellido and D. Wands, "Metric perturbations in two field inflation," Phys.Rev., vol. D53, pp. 5437–5445, 1996.

[34] E. Komatsu and D. N. Spergel, "Acoustic signatures in the primary microwave background bispectrum," Phys.Rev., vol. D63, p. 063002, 2001.

[35] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, "The Three point correlation function of the cosmic microwave background in inflationary models," Astrophys.J., vol. 430, pp. 447–457, 1994.

[36] L. Verde, L.-M. Wang, A. Heavens, and M. Kamionkowski, "Large

scale structure, the cosmic microwave background, and primordial non-gaussianity," Mon.Not.Roy.Astron.Soc., vol. 313, pp. L141–L147, 2000.

[37] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, "Limits on non-gaussianities from wmap data," JCAP, vol. 0605, p. 004, 2006.

[38] L. Senatore, K. M. Smith, and M. Zaldarriaga, "Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data," JCAP, vol. 1001, p. 028, 2010.

[39] X. Chen, M.-x. Huang, S. Kachru, and G. Shiu, "Observational signatures and non-Gaussianities of general single field inflation," JCAP, vol. 0701, p. 002, 2007.

[40] J. Khoury and F. Piazza, "Rapidly-Varying Speed of Sound, Scale Invariance and Non-Gaussian Signatures," JCAP, vol. 0907, p. 026, 2009.

[41] J. Noller and J. Magueijo, "Non-Gaussianity in single field models without slow-roll," Phys.Rev., vol. D83, p. 103511, 2011.

[42] R. H. Ribeiro, "Inflationary signatures of single-field models beyond slow-roll," JCAP, vol. 1205, p. 037, 2012.

[43] J. Noller, "Constraining fast-roll inflation," 2012.

[44] X. Chen and Y. Wang, "Large non-Gaussianities with Intermediate Shapes from Quasi-Single Field Inflation," Phys.Rev., vol. D81, p. 063511, 2010.

[45] X. Chen and Y. Wang, "Quasi-Single Field Inflation and Non-Gaussianities," JCAP, vol. 1004, p. 027, 2010.

[46] S. Gupta, A. Berera, A. Heavens, and S. Matarrese, "Non-Gaussian signatures in the cosmic background radiation from warm inflation," Phys.Rev., vol. D66, p. 043510, 2002.

[47] I. G. Moss and C. Xiong, "Non-Gaussianity in fluctuations from warm inflation," JCAP, vol. 0704, p. 007, 2007.

[48] D. Lopez Nacir, R. A. Porto, L. Senatore, and M. Zaldarriaga, "Dissipative effects in the Effective Field Theory of Inflation," JHEP, vol. 1201, p. 075, 2012.

[49] R. Holman and A. J. Tolley, "Enhanced Non-Gaussianity from Excited Initial States," JCAP, vol. 0805, p. 001, 2008.

[50] P. D. Meerburg, J. P. van der Schaar, and P. S. Corasaniti, "Signatures of Initial State Modifications on Bispectrum Statistics," JCAP, vol. 0905, p. 018, 2009.

[51] N. Agarwal, R. Holman, A. J. Tolley, and J. Lin, "Effective field theory and non-Gaussianity from general inflationary states," 2012.

[52] D. Langlois, S. Renaux-Petel, D. A. Steer, and T. Tanaka, "Primordial perturbations and non-Gaussianities in DBI and general multi-field inflation," Phys.Rev., vol. D78, p. 063523, 2008.

[53] F. Arroja, S. Mizuno, and K. Koyama, "Non-gaussianity from the bispectrum in general multiple field inflation," JCAP, vol. 0808, p. 015, 2008.

[54] S. Renaux-Petel, "Combined local and equilateral non-Gaussianities from multifield DBI inflation," JCAP, vol. 0910, p. 012, 2009.

[55] D. Langlois and A. Lepidi, "General treatment of isocurvature perturbations and non-Gaussianities," JCAP, vol. 1101, p. 008, 2011.

[56] D. Langlois and B. van Tent, "Hunting for Isocurvature Modes in the CMB non-Gaussianities," Class.Quant.Grav., vol. 28, p. 222001, 2011.

[57] D. Langlois and B. van Tent, "Isocurvature modes in the CMB bispectrum," JCAP, vol. 1207, p. 040, 2012.

[58] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, "The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects," Phys.Rev., vol. D77, p. 123514, 2008.

[59] M. LoVerde, A. Miller, S. Shandera, and L. Verde, "Effects of Scale-Dependent Non-Gaussianity on Cosmological Structures," JCAP, vol. 0804, p. 014, 2008.

[60] L. Verde and S. Matarrese, "Detectability of the effect of Inflationary non-Gaussianity on halo bias," Astrophys.J., vol. 706, pp. L91–L95, 2009.

[61] A. Becker, D. Huterer, and K. Kadota, "Scale-Dependent Non-Gaussianity as a Generalization of the Local Model," JCAP, vol. 1101, p. 006, 2011.

[62] A. Becker, D. Huterer, and K. Kadota, "Constraining Scale-Dependent Non-Gaussianity with Future Large-Scale Structure and the CMB," JCAP, vol. 1212, p. 034, 2012.

[63] J. Norena, L. Verde, G. Barenboim, and C. Bosch, "Prospects for constraining the shape of non-Gaussianity with the scale-dependent bias," JCAP, vol. 1208, p. 019, 2012.

[64] C. Wagner and L. Verde, "N-body simulations with generic non-Gaussian initial conditions II: Halo bias," JCAP, vol. 1203, p. 002, 2012.

[65] I. Agullo and S. Shandera, "Large non-Gaussian Halo Bias from Single Field Inflation," JCAP, vol. 1209, p. 007, 2012.

[66] D. Spergel et al., "Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology," Astrophys.J.Suppl., vol. 170, p. 377, 2007.

[67] R. Hlozek, J. Dunkley, G. Addison, J. W. Appel, J. R. Bond, et al., "The Atacama Cosmology Telescope: a measurement of the primordial power spectrum," Astrophys.J., vol. 749, p. 90, 2012.

[68] S. Bridle, A. Lewis, J. Weller, and G. Efstathiou, "Reconstructing the primordial power spectrum," Mon.Not.Roy.Astron.Soc., vol. 342, p. L72, 2003.

[69] E. J. Copeland, E. W. Kolb, A. R. Liddle, and J. E. Lidsey, "Reconstructing the inflaton potential: Perturbative reconstruction to second order," Phys.Rev., vol. D49, pp. 1840–1844, 1994.

[70] W. H. Kinney, E. W. Kolb, A. Melchiorri, and A. Riotto, "WMAPping inflationary physics," Phys.Rev., vol. D69, p. 103516, 2004.

[71] R. Bean, D. J. Chung, and G. Geshnizjani, "Reconstructing a general inflationary action," Phys.Rev., vol. D78, p. 023517, 2008.

[72] P. Adshead and R. Easther, "Constraining Inflation," JCAP, vol. 0810, p. 047, 2008.

[73] P. D. Meerburg, "Oscillations in the Primordial Bispectrum I: Mode Expansion," Phys.Rev., vol. D82, p. 063517, 2010.

[74] D. Regan, P. Mukherjee, and D. Seery, "General CMB bispectrum analysis using wavelets and separable modes," 2013.

[75] D. Regan, E. Shellard, and J. Fergusson, "General CMB and Primordial Trispectrum Estimation," Phys.Rev., vol. D82, p. 023520, 2010.

[76] J. Fergusson, D. Regan, and E. Shellard, "Optimal Trispectrum Estimators and WMAP Constraints," 2010.

[77] J. Fergusson, D. Regan, and E. Shellard, "Rapid Separable Analysis of Higher Order Correlators in Large Scale Structure," Phys.Rev., vol. D86, p. 063511, 2012.

[78] D. Regan, M. Schmittfull, E. Shellard, and J. Fergusson, "Universal Non-Gaussian Initial Conditions for N-body Simulations," Phys.Rev., vol. D86, p. 123524, 2012.

[79] M. Schmittfull, D. Regan, and E. Shellard, "Fast Estimation of Gravitational and Primordial Bispectra in Large Scale Structures," 2012.

[80] A. Lewis, A. Challinor, and A. Lasenby, "Efficient computation of CMB anisotropies in closed FRW models," Astrophys. J., vol. 538, pp. 473–476, 2000.

[81] U. Seljak and M. Zaldarriaga, "A Line of sight integration approach to cosmic microwave background anisotropies," Astrophys.J., vol. 469, pp. 437–444, 1996.

[82] D. Babich, P. Creminelli, and M. Zaldarriaga, "The Shape of non-Gaussianities," JCAP, vol. 0408, p. 009, 2004.

[83] T. Battefeld and J. Grieb, "Anatomy of bispectra in general single-field inflation – modal expansions," JCAP, vol. 1112, p. 003, 2011.

[84] X. Chen, "Primordial Non-Gaussianities from Inflation Models," Adv.Astron., vol. 2010, p. 638979, 2010.

[85] N. Barnaby, "On Features and Nongaussianity from Inflationary Particle Production," Phys.Rev., vol. D82, p. 106009, 2010.

[86] X. Chen, R. Easther, and E. A. Lim, "Large Non-Gaussianities in Single Field Inflation," JCAP, vol. 0706, p. 023, 2007.

[87] X. Chen, R. Easther, and E. A. Lim, "Generation and Characterization of Large Non-Gaussianities in Single Field Inflation," JCAP, vol. 0804, p. 010, 2008.

[88] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," JHEP, vol. 0305, p. 013, 2003.

[89] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, "Second order cosmological perturbations from inflation," Nucl.Phys., vol. B667, pp. 119–148, 2003.

[90] D. Wands, "Multiple field inflation," Lect.Notes Phys., vol. 738, pp. 275–304, 2008.

[91] C. T. Byrnes and K.-Y. Choi, "Review of local non-Gaussianity from multi-field inflation," Adv.Astron., vol. 2010, p. 724525, 2010.

[92] K. Enqvist, A. Jokinen, A. Mazumdar, T. Multamaki, and A. Vaihkonen, "Non-Gaussianity from preheating," Phys.Rev.Lett., vol. 94, p. 161301, 2005.

[93] D. Wands, "Local non-Gaussianity from inflation," Class.Quant.Grav., vol. 27, p. 124002, 2010.

[94] E. Sefusatti, J. R. Fergusson, X. Chen, and E. Shellard, "Effects and Detectability of Quasi-Single Field Inflation in the Large-Scale Structure and Cosmic Microwave Background," JCAP, vol. 1208, p. 033, 2012.

[95] P. Creminelli, "On non-Gaussianities in single-field inflation," JCAP, vol. 0310, p. 003, 2003.

[96] D. Seery and J. E. Lidsey, "Primordial non-Gaussianities in single field inflation," JCAP, vol. 0506, p. 003, 2005.

[97] C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, "k - inflation," Phys.Lett., vol. B458, pp. 209–218, 1999.

[98] M. Li, T. Wang, and Y. Wang, "General Single Field Inflation with Large Positive Non-Gaussianity," JCAP, vol. 0803, p. 028, 2008.

[99] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, "Ghost inflation," JCAP, vol. 0404, p. 001, 2004.

[100] E. Silverstein and D. Tong, "Scalar speed limits and cosmology: Acceleration from D-cceleration," Phys.Rev., vol. D70, p. 103505, 2004.

[101] M. Alishahiha, E. Silverstein, and D. Tong, "DBI in the sky," Phys.Rev., vol. D70, p. 123505, 2004.

[102] J. Magueijo, J. Noller, and F. Piazza, "Bimetric structure formation: non-Gaussian predictions," Phys.Rev., vol. D82, p. 043521, 2010.

[103] P. Creminelli, G. D'Amico, M. Musso, J. Norena, and E. Trincherini, "Galilean symmetry in the effective theory of inflation: new shapes of non-Gaussianity," JCAP, vol. 1102, p. 006, 2011.

[104] A. Nicolis, R. Rattazzi, and E. Trincherini, "The Galileon as a local modification of gravity," Phys.Rev., vol. D79, p. 064036, 2009.

[105] C. Burrage, C. de Rham, D. Seery, and A. J. Tolley, "Galileon inflation," JCAP, vol. 1101, p. 014, 2011.

[106] R. H. Ribeiro and D. Seery, "Decoding the bispectrum of single-field inflation," JCAP, vol. 1110, p. 027, 2011.

[107] A. Berera, I. G. Moss, and R. O. Ramos, "Warm Inflation and its Microphysical Basis," Rept.Prog.Phys., vol. 72, p. 026901, 2009.

[108] M. Liguori, E. Sefusatti, J. Fergusson, and E. Shellard, "Primordial non-Gaussianity and Bispectrum Measurements in the Cosmic Microwave Background and Large-Scale Structure," Adv.Astron., vol. 2010, p. 980523, 2010.

[109] D. Babich and M. Zaldarriaga, "Primordial bispectrum information from CMB polarization," Phys.Rev., vol. D70, p. 083005, 2004.

[110] P. collaboration, "Planck Collaboration Blue Book," 2005.

[111] E. Sefusatti, M. Liguori, A. P. Yadav, M. G. Jackson, and E. Pajer, "Constraining Running Non-Gaussianity," JCAP, vol. 0912, p. 022, 2009.

[112] M. Liguori and A. Riotto, "Impact of Uncertainties in the Cosmological Parameters on the Measurement of Primordial non-Gaussianity," Phys.Rev., vol. D78, p. 123004, 2008.

[113] A. Lewis, A. Challinor, and D. Hanson, "The shape of the CMB lensing bispectrum," JCAP, vol. 1103, p. 018, 2011.

[114] R. Pearson, A. Lewis, and D. Regan, "CMB lensing and primordial squeezed non-Gaussianity," JCAP, vol. 1203, p. 011, 2012.

[115] K. M. Smith and M. Zaldarriaga, "Algorithms for bispectra: Forecasting, optimal analysis, and simulation," Mon.Not.Roy.Astron.Soc., vol. 417, pp. 2–19, 2011.

[116] P. Ade et al., "Planck 2013 results. XXII. Constraints on inflation," 2013.

[117] P. Ade et al., "Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity," 2013.

[118] P. A. Abell et al., "LSST Science Book, Version 2.0," 2009.

[119] A. Refregier, A. Amara, T. Kitching, A. Rassat, R. Scaramella, et al., "Euclid Imaging Consortium Science Book," 2010.

[120] C. Carbone, L. Verde, and S. Matarrese, "Non-Gaussian halo bias and future galaxy surveys," Astrophys.J., vol. 684, pp. L1–L4, 2008.

[121] C. Fedeli, C. Carbone, L. Moscardini, and A. Cimatti, "The clustering of galaxies and galaxy clusters: constraints on primordial non-Gaussianity from future wide-field surveys," Mon.Not.Roy.Astron.Soc., vol. 414, pp. 1545–1559, 2011.

[122] L. Amendola et al., "Cosmology and fundamental physics with the Euclid satellite," 2012.

[123] L. Verde, "Non-Gaussianity from Large-Scale Structure Surveys," Adv.Astron., vol. 2010, p. 768675, 2010.

[124] A. Ashoorioon and G. Shiu, "A Note on Calm Excited States of Inflation," JCAP, vol. 1103, p. 025, 2011.

[125] A. Ashoorioon, K. Dimopoulos, M. Sheikh-Jabbari, and G. Shiu, "Reconciliation of High Energy Scale Models of Inflation with Planck," JCAP, vol. 1402, p. 025, 2014.

[126] X. Chen, "Inflation from warped space," JHEP, vol. 0508, p. 045, 2005.

[127] S. Renaux-Petel, "Orthogonal non-Gaussianities from Dirac-Born-Infeld Galileon inflation," Class.Quant.Grav., vol. 28, p. 182001, 2011.

[128] C. Wagner, L. Verde, and L. Boubekeur, "N-body simulations with generic non-Gaussian initiallo conditions I: Power Spectrum and halo mass function," JCAP, vol. 1010, p. 022, 2010.

[129] M. Shiraishi, E. Komatsu, M. Peloso, and N. Barnaby, "Signatures of anisotropic sources in the squeezed-limit bispectrum of the cosmic microwave background," JCAP, vol. 1305, p. 002, 2013.

[130] V. Desjacques, D. Jeong, and F. Schmidt, "Accurate Predictions for the Scale-Dependent Galaxy Bias from Primordial Non-Gaussianity," Phys.Rev., vol. D84, p. 061301, 2011.

[131] V. Desjacques, D. Jeong, and F. Schmidt, "Non-Gaussian Halo Bias Reexamined: Mass-dependent Amplitude from the Peak-Background Split and Thresholding," Phys.Rev., vol. D84, p. 063512, 2011.

[132] 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., vol. D85, p. 083002, 2012.

[133] V. Desjacques, U. Seljak, and I. Iliev, "Scale-dependent bias induced by local non-Gaussianity: A comparison to N-body simulations," Mon.Not.Roy.Astron.Soc., vol. 396, pp. 85–96, 2009.

[134] A. Cooray and R. K. Sheth, "Halo models of large scale structure," Phys.Rept., vol. 372, pp. 1–129, 2002.

[135] A. Pillepich, C. Porciani, and O. Hahn, "Universal halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions," Mon.Not.Roy.Astron.Soc., vol. 402, pp. 191–206, 2010.

[136] J. F. Navarro, C. S. Frenk, and S. D. M. White, "The Structure of Cold Dark Matter Halos," Astrophys. J., vol. 462, pp. 563–575, 1996.

[137] J. F. Navarro, C. S. Frenk, and S. D. White, "A Universal density profile from hierarchical clustering," Astrophys.J., vol. 490, pp. 493–508, 1997.

[138] J. S. Bullock et al., "Profiles of dark haloes: evolution, scatter, and environment," Mon. Not. Roy. Astron. Soc., vol. 321, pp. 559–575, 2001.

[139] C. Fedeli and L. Moscardini, "Cosmic shear statistics in cosmologies with non-Gaussian initial conditions," Mon.Not.Roy.Astron.Soc., vol. 405, pp. 681–694, 2010.

[140] C. Giocoli, M. Bartelmann, R. K. Sheth, and M. Cacciato, "Halo model description of the non-linear dark matter power spectrum at $k \gg 1$ Mpc^{-1}," Mon.Not.Roy.Astron.Soc., vol. 408, p. 300, 2010.

[141] Y. Kulinich, B. Novosyadlyj, and S. Apunevych, "Nonlinear power spectra of dark and luminous matter in the halo model of structure formation," Phys.Rev., vol. D88, no. 10, p. 103505, 2013.

[142] G. Kauffmann, J. M. Colberg, A. Diaferio, and S. D. White, "Clustering of galaxies in a hierarchical universe: 1. Methods and results at z=0," Mon.Not.Roy.Astron.Soc., 1998.

[143] H. Hoekstra, H. K. Yee, and M. D. Gladders, "Measurement of the bias parameter from weak lensing," Astrophys.J., vol. 558, pp. L11–L14, 2001.

[144] H. Hoekstra, L. Van Waerbeke, M. D. Gladders, Y. Mellier, and H. Yee, "Weak lensing study of galaxy biasing," Astrophys.J., vol. 577, pp. 604–614, 2002.

[145] U. Seljak et al., "SDSS galaxy bias from halo mass-bias relation and its cosmological implications," Phys.Rev., vol. D71, p. 043511, 2005.

[146] M. Cacciato, F. C. van den Bosch, S. More, H. Mo, and X. Yang, "Cosmological Constraints from a Combination of Galaxy Clustering and Lensing – III. Application to SDSS Data," 2012.

[147] E. Jullo, J. Rhodes, A. Kiessling, J. E. Taylor, R. Massey, et al., "COSMOS: Stochastic bias from measurements of weak lensing and galaxy clustering," Astrophys.J., vol. 750, p. 37, 2012.

[148] S. More, H. Miyatake, R. Mandelbaum, M. Takada, D. Spergel, et al., "The Weak Lensing Signal and the Clustering of BOSS Galaxies: Cosmological Constraints," 2014.

[149] D. Jeong, E. Komatsu, and B. Jain, "Galaxy-CMB and galaxy-galaxy lensing on large scales: sensitivity to primordial non-Gaussianity," Phys.Rev., vol. D80, p. 123527, 2009.

[150] Y. Takeuchi, K. Ichiki, and T. Matsubara, "Constraints on primordial non-Gaussianity from Galaxy-CMB lensing cross-correlation," Phys.Rev., vol. D82, p. 023517, 2010.

[151] T. Giannantonio and W. J. Percival, "Using correlations between CMB lensing and large-scale structure to measure primordial non-Gaussianity," Mon.Not.Roy.Astron.Soc., vol. 441, p. L16L20, 2014.

[152] Y. Jing, H. Mo, and G. Borner, "Spatial correlation function and pairwise velocity dispersion of galaxies: CDM models versus the Las Campanas Survey," Astrophys.J., vol. 494, p. 1, 1998.

[153] R. Sheth and W. Saslaw, "Synthesising the Observed Distribution of Galaxies," Astrophys.J., vol. 437, p. 35, 1994.

[154] E. Sefusatti, M. Crocce, and V. Desjacques, "The Halo Bispectrum in N-body Simulations with non-Gaussian Initial Conditions," 2011.

[155] L. Verde, A. F. Heavens, W. J. Percival, S. Matarrese, C. M. Baugh, et al., "The 2dF Galaxy Redshift Survey: The Bias of galaxies and the density of the Universe," Mon.Not.Roy.Astron.Soc., vol. 335, p. 432, 2002.

[156] F. A. Marin et al., "The WiggleZ Dark Energy Survey: constraining galaxy bias and cosmic growth with 3-point correlation functions," 2013.

[157] H. Gil-Marn, J. Norea, L. Verde, W. J. Percival, C. Wagner, et al., "The power spectrum and bispectrum of SDSS DR11 BOSS galaxies I: bias and gravity," 2014.

[158] E. Sefusatti and E. Komatsu, "The bispectrum of galaxies from high-redshift galaxy surveys: Primordial non-Gaussianity and non-linear galaxy bias," Phys.Rev., vol. D76, p. 083004, 2007.

[159] N. Agarwal, S. Ho, and S. Shandera, "Constraining the initial conditions of the Universe using large scale structure," 2013.

[160] S. Shandera, N. Dalal, and D. Huterer, "A generalized local ansatz and its effect on halo bias," JCAP, vol. 1103, p. 017, 2011.

[161] N. Kaiser, "Clustering in real space and in redshift space," Mon. Not. Roy. Astron. Soc., vol. 227, pp. 1–27, 1987.

[162] L. Samushia, W. J. Percival, and A. Raccanelli, "Interpreting large-scale redshift-space distortion measurements," Mon.Not.Roy.Astron.Soc., vol. 420, pp. 2102–2119, 2012.

[163] J. Kwan, G. F. Lewis, and E. V. Linder, "Mapping Growth and Gravity with Robust Redshift Space Distortions," Astrophys.J., vol. 748, p. 78, 2012.

[164] C. Carbone, O. Mena, and L. Verde, "Cosmological Parameters Degeneracies and Non-Gaussian Halo Bias," JCAP, vol. 1007, p. 020, 2010.

[165] E. Di Dio, F. Montanari, R. Durrer, and J. Lesgourgues, "Cosmological Parameter Estimation with Large Scale Structure Observations," JCAP, vol. 1401, p. 042, 2014.

[166] Y. Wang, C.-H. Chuang, and C. M. Hirata, "Toward More Realistic Forecasting of Dark Energy Constraints from Galaxy Redshift Surveys," 2012.

[167] "DESI Conceptual Design Report," http://desi.lbl.gov/wp-content/uploads/2014/04/D 2014.

[168] C. Di Porto, E. Branchini, J. Bel, F. Marulli, M. Bolzonella, et al., "The VIMOS Public Extragalactic Redshift Survey (VIPERS). Measuring non-linear galaxy bias at z 0.8," 2014.

[169] J. W. Colbert, H. Teplitz, H. Atek, A. Bunker, M. Rafelski, et al., "Predicting Future Space Near-IR Grism Surveys using the WFC3 Infrared Spectroscopic Parallels Survey," Astrophys.J., vol. 779, p. 34, 2013.

[170] J. Geach, A. Cimatti, W. Percival, Y. Wang, L. Guzzo, et al., "Empirical H-alpha emitter count predictions for dark energy surveys," 2009.

[171] A. Vilenkin and L. H. Ford, "Gravitational effects upon cosmological phase transitions," Phys.Rev., vol. D26, p. 1231, 1982.

[172] A. Dey and S. Paban, "Non-Gaussianities in the cosmological perturbation spectrum due to primordial anisotropy," JCAP, vol. 1204, p. 039, 2012.

[173] A. Dey, E. Kovetz, and S. Paban, "Non-Gaussianities in the cosmological perturbation spectrum due to primordial anisotropy II," JCAP, vol. 1210, p. 055, 2012.

[174] I. Agullo, A. Ashtekar, and W. Nelson, "The pre-inflationary dynamics of loop quantum cosmology: Confronting quantum gravity with observations," Class.Quant.Grav., vol. 30, p. 085014, 2013.

[175] L. Lello, D. Boyanovsky, and R. Holman, "Pre-slow roll initial conditions: Large scale power suppression and infrared aspects during inflation," Phys.Rev., vol. D89, no. 6, p. 063533, 2014.

[176] K. Sugimura and E. Komatsu, "Bispectrum from open inflation," JCAP, vol. 1311, p. 065, 2013.

[177] C. R. Contaldi, R. Bean, and J. Magueijo, "Photographing the wave function of the Universe," Phys.Lett., vol. B468, pp. 189–194, 1999.

[178] J. Martin and R. H. Brandenberger, "Trans-Planckian problem of inflationary cosmology," Phys.Rev., vol. D63, p. 123501, 2001.

[179] N. Kaloper, M. Kleban, A. E. Lawrence, and S. Shenker, "Signatures of short distance physics in the cosmic microwave background," Phys.Rev., vol. D66, p. 123510, 2002.

[180] U. H. Danielsson, "Note on inflation and trans-Planckian physics," Phys.Rev., vol. D66, p. 023511, 2002.

[181] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, "A generic estimate of trans-planckian modifications to the primordial power spectrum in inflation," Phys. Rev., vol. D66, p. 023518, 2002.

[182] I. Agullo and L. Parker, "Non-Gaussianities and the stimulated creation of quanta in the inflationary Universe," Phys.Rev., vol. D83, p. 063526, 2011.

[183] J. Ganc, "Calculating the local-type $f_{\rm NL}$ for slow-roll inflation with a non-vacuum initial state," Phys. Rev., vol. D84, p. 063514, 2011.

[184] D. Chialva, "Signatures of very high energy physics in the squeezed limit of the bispectrum," JCAP, vol. 1210, p. 037, 2012.

[185] S. Kundu, "Inflation with general initial conditions for scalar perturbations," JCAP, vol. 1202, p. 005, 2012.

[186] S. Brahma, E. Nelson, and S. Shandera, "Fossilized gravitational wave relic and primordial clocks," Phys.Rev., vol. D89, no. 2, p. 023507, 2014.

[187] S. Bahrami and E. E. Flanagan, "Primordial non-Gaussianities in single field inflationary models with non-trivial initial states," JCAP, vol. 1410, no. 10, p. 010, 2014.

[188] R. Flauger, D. Green, and R. A. Porto, "On squeezed limits in single-field inflation. Part I," JCAP, vol. 1308, p. 032, 2013.

[189] A. Aravind, D. Lorshbough, and S. Paban, "Non-Gaussianity from excited initial inflationary states," JHEP, vol. 1307, p. 076, 2013.

[190] P. A. R. Ade et al., "Planck 2015. XX. Constraints on inflation," 2015.

[191] C. Bennett et al., "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results," 2012.

[192] M. Munchmeyer, F. Bouchet, M. G. Jackson, and B. Wandelt, "The Komatsu Spergel Wandelt estimator for oscillations in the cosmic microwave background bispectrum," Astron.Astrophys., vol. 570, p. A94, 2014.

[193] M. Munchmeyer, P. D. Meerburg, and B. D. Wandelt, "Optimal estimator for resonance bispectra in the CMB," Phys.Rev., vol. D91, no. 4, p. 043534, 2015.

[194] C. de Boor, "A practical guide to splines," Springer, 2001.

[195] P. Dierckx, "Curve and surface fitting with splines," Oxford University Press, 1995.

[196] C. Cheung, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "On the consistency relation of the 3-point function in single field inflation," JCAP, vol. 0802, p. 021, 2008.

[197] G. W. Horndeski, "Second-order scalar-tensor field equations in a four-dimensional space," Int.J.Theor.Phys., vol. 10, pp. 363–384, 1974.

[198] X. Gao and D. A. Steer, "Inflation and primordial non-Gaussianities of "generalized galileons"," JCAP, vol. 1112, p. 019, 2011.

[199] A. De Felice and S. Tsujikawa, "Inflationary non-Gaussianities in the most general second-order scalar-tensor theories," Phys.Rev., vol. D84, p. 083504, 2011.

[200] C. Burrage, R. H. Ribeiro, and D. Seery, "Large slow-roll corrections to the bispectrum of noncanonical inflation," JCAP, vol. 1107, p. 032, 2011.

[201] N. Agarwal, R. H. Ribeiro, and R. Holman, "Why does the effective field theory of inflation work?," JCAP, vol. 1406, p. 016, 2014.

[202] J. S. Schwinger, "Brownian motion of a quantum oscillator," J.Math.Phys., vol. 2, pp. 407–432, 1961.

[203] K. T. Mahanthappa, "Multiple production of photons in quantum electrodynamics," Phys.Rev., vol. 126, pp. 329–340, 1962.

[204] P. M. Bakshi and K. T. Mahanthappa, "Expectation value formalism in quantum field theory. 1.," J.Math.Phys., vol. 4, pp. 1–11, 1963.

[205] L. P. Kadanoff and G. Baym, "Quantum Statistical Mechanics," New York, 1962.

[206] P. M. Bakshi and K. T. Mahanthappa, "Expectation value formalism in quantum field theory. 2.," J.Math.Phys., vol. 4, pp. 12–16, 1963.

[207] L. V. Keldysh, "Diagram technique for nonequilibrium processes," Zh.Eksp.Teor.Fiz., vol. 47, pp. 1515–1527, 1964.

[208] R. D. Jordan, "Effective field equations for expectation values," Phys. Rev., vol. D33, pp. 444–454, 1986.

[209] E. Calzetta and B. L. Hu, "Closed time path functional formalism in curved space-time: application to cosmological back reaction problems," Phys.Rev., vol. D35, p. 495, 1987.

[210] P. Eilers, I. Currie, and M. Durban, "Fast and compact smoothing on large multidimensional grids," Computational Statistics and Data Analysis, vol. 50, pp. 61–76, 2006.

[211] J. R. Fergusson and E. P. S. Shellard, "Primordial non-Gaussianity and the CMB bispectrum," Phys.Rev., vol. D76, p. 083523, 2007.

[212] D. H. Rudd, A. R. Zentner, and A. V. Kravtsov, "Effects of Baryons and Dissipation on the Matter Power Spectrum," Astrophys.J., vol. 672, pp. 19–32, 2008.

[213] A. R. Zentner, E. Semboloni, S. Dodelson, T. Eifler, E. Krause, et al., "Accounting for Baryons in Cosmological Constraints from Cosmic Shear," Phys.Rev., vol. D87, no. 4, p. 043509, 2013.

[214] D. Polarski and A. A. Starobinsky, "Semiclassicality and decoherence of cosmological perturbations," Class.Quant.Grav., vol. 13, pp. 377–392, 1996.

[215] P. Eilers and B. Marx, "Flexible Smoothing with B-splines and Penalties," Statistical Science, vol. 11, pp. 89–121, 1996.