Lifting Primordial Non-Gaussianity Above the Noise

Yvette Welling[♣], Drian van der Woude[♠] and Enrico Pajer[♠]

[♣] *Leiden Observatory, Universiteit Leiden, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands*
[♠] *Institute for Theoretical Physics, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands*

Abstract

Primordial non-Gaussianity (PNG) in Large Scale Structures is obfuscated by the many additional sources of non-linearity. Within the Effective Field Theory approach to Standard Perturbation Theory, we show that matter non-linearities in the bispectrum can be modeled sufficiently well to strengthen current bounds with near future surveys, such as Euclid. We find that the EFT corrections are crucial to this improvement in sensitivity. Yet, our understanding of non-linearities is still insufficient to reach important theoretical benchmarks for equilateral PNG, while, for local PNG, our forecast is more optimistic. We consistently account for the theoretical error intrinsic to the perturbative approach and discuss the details of its implementation in Fisher forecasts.

Contents

1 Introduction and summary

Primordial deviations from Gaussianity are key to understand inflation and will soon be tested via a number of ambitious Large Scale Structure (LSS) surveys. It is therefore imperative to understand how late-time LSS observations can be related to the parameters that characterize *primordial non-Gaussianity* (PNG). This relation is complicated and non-linear. The degree to

which we will be able to collect further primordial information from LSS survey will eventually be determined by our ability to model this non-linear relation. In this work, we focus on specific source of non-linearities, namely perturbative *matter non-linearities*. These are generated by the sub-horizon gravitational evolution of small initial matter inhomogeneities into larger ones, until the perturbations compete locally with the homogeneous background expansion. For concreteness, we study local, equilateral and quasi-single field non-Gaussianity, since these are well-motivated theoretically and represent signals that are complementary from the point of view of observations. Additional sources of non-linearity are also important, such as for example bias and redshift space distortion. In case of equilateral and quasi single field PNG, these are expected to further worsen our ability to constraint primordial parameters. In this sense, our results can be interpreted as *lower bounds* on the precision of future constraints. For local PNG, it is possible that non-linearities encapsulated in the biasing of tracers, if very well understood, might eventually help us improve on the bounds we find here (see [1] and e.g. [2, 3] for a recent estimate). We will discuss this possibility in subsection 4.1.2.

In our analysis, we will use the Effective Field Theory of Large Scale Structures (EFT of LSS) [4], which builds on Standard Perturbation Theory (SPT) [5], and provides a consistent perturbative approach to describe the evolution of matter distribution. We focus exclusively on the matter bispectrum, since it is a primary probe of PNG that is affected by matter non-linearities. Recent work on PNG and the bispectrum includes [2, 6–11]. Within the EFT approach, the bispectrum generated by the late-time gravitational evolution from otherwise Gaussian initial conditions has been studied in [12, 13]). This contribution plays the role of background noise in PNG searches. The signal, namely the primordial bispectrum, is also affected by gravitational non-linearities. This has been recently studied in [14]. Here, we use these two results and present a Fisher forecast for constraints on PNG. A key element of our forecast is the inclusion of theoretical error, employing and further developing the recent proposal of [15].

For the convenience of the reader, we collect here our major findings with references to where they are discussed in the rest paper.

- When using the EFT of LSS, the perturbative approach to model matter non-linearities will not prevent upcoming LSS surveys to *improve upon the current bounds* from CMB anisotropies [16] (see Table 1).

- Our limited perturbative understanding of matter non-linearities limits the achievable bounds on equilateral non-Gaussianity from planned galaxy surveys to $\sigma(f_{NL}^{eq}) \gtrsim 10$ (see Table 1), *far from the theoretically interesting benchmark* $\sigma(f_{NL}^{eq}) \sim 1$ (see e.g. [17, 18] and references therein). We estimate that this remains true even if one included the full two-loop corrections (see Table 2). Local non-Gaussianity is more promising, and we find e.g. for Euclid $\sigma(f_{NL}^{loc}) \gtrsim 1$.

- The consistent treatment of short-scale effects within the EFT approach allows one to *improve PNG constraints by a factor of about 3* (see Table 3). This relies on two facts. First, the EFT parameters provide a better description of the late-time gravitational non-linearities (the "background" discussed in [12]). Second, for the specification of most upcoming experiments, the EFT parameters are only weakly correlated with PNG, and so marginalizing over them hardly degrades the constraints (see subsection 4.2).

- Both the SPT loops and the EFT corrections to primordial non-Gaussianity (the "signal" discussed in [14]) are small and their inclusion does *not* improve the PNG constraints appreciably (see first and second line of Table 3).

- We discuss several aspects of the method proposed in [15] to *model the theoretical error* inherent to the perturbative approach. We show that a wrong shape for the theoretical error can lead to a biased estimate for f_{NL} . This happens when it partly underestimates the error. Conversely, a conservatively large theoretical error leads to correct unbiased

results. We thoroughly analyze the dependence of the Fisher forecast on the correlation length used in [15], and explain our results with a toy model.

This paper is organized as follows. In section 2, we review the results for the matter bispectrum in the EFT of LSS accounting for PNG. In section 3, we discuss the details of the Fisher forecast with particular emphasis on a consistent treatment of theoretical uncertainties. Section 4 is devoted to a discussion of the results of the Fisher forecast on PNG constraints from LSS surveys. We conclude in section 5. Several appendices contain more technical results. In Appendix A, we summarize all relevant formulae to compute the bispectrum in the EFT of LSS. In Appendix B, we present a detailed discussion of how to consistently account for theoretical errors. Appendix C discusses the issue of binning the data for the Fisher forecast and finally, for the convenience of the reader, we collected all symbols used in this paper and their meaning in Appendix D.

Conventions Redshift z and conformal spatial coordinates $\mathbf{x}$ are used as measures of time and position. We use the following convention for the Fourier transform

$$F(\mathbf{x}) = \int_{\mathbf{k}} \tilde{F}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \text{where we use the shorthand} \quad \int_{\mathbf{k}} \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3}. \tag{1}$$

In particular, this implies that we have the following relation between any N-point equal-time correlation function and its spectrum

$$\langle \delta(\mathbf{k}_1) \dots \delta(\mathbf{k}_N) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \dots + \mathbf{k}_N) S(\mathbf{k}_1, \dots, \mathbf{k}_N), \tag{2}$$

where we suppressed the time dependence.

For the numerical analysis, we compute the linear power spectrum with CAMB [19], where we assume a standard cosmological model with $\Omega_\Lambda^0 = 0.728$, $\Omega_m^0 = 0.272$, $h = 0.704$, $n_s = 0.967$ and $A_\zeta = 2.46 \times 10^{-9}$.

2 Analytical predictions for the bispectrum

In this section, we review the analytical predictions for the late-time matter bispectrum within the Effective Field Theory of Large Scale Structures (EFT of LSS), accounting for non-Gaussian initial conditions. In subsection 2.1 we collect the contributions to the bispectrum up to first order in primordial non-Gaussianity and up to 'one-loop' order in perturbation theory. In subsection 2.2, we specify the types of primordial non-Gaussianity (PNG) we study in this paper. In subsection 2.3, we discuss the theoretical errors, which are intrinsic to the perturbative approach.

2.1 The bispectrum in the EFT of LSS

Despite being almost collisionless, cold dark matter on large scales behaves approximately as a fluid. This relies on the fact that the primordial universe is locally in (thermodynamical) equilibrium *and* that, during the age of the universe, dark matter particles move only over a distance that is small compared with the scales of interest. This displacement plays the same role as the mean free path in the more familiar interacting fluids. As long as we consider scales much larger than this displacement, an effective fluid description can be applied [4]. Here we follow the Effective Field Theory approach to Large Scale Structures (EFT of LSS). The dark matter bispectrum induced by gravity was discussed in [12, 13]. Non-Gaussian initial condition were subsequently accounted for in [14]. The EFT of LSS allows to perturbatively compute non-linear correlators of the matter density contrast $\delta(\mathbf{x}, z)$ [20] and velocity $v(\mathbf{x}, z)$ [21], taking into account the effect of short-scale non-perturbative physics on the large-scale dynamics. In practice, one can use the results of Standard Perturbation Theory (SPT) [5], and correct them with additional effective terms, which will be denoted with the subscript 'EFT'. We differentiate between contributions

to the bispectrum coming from primordial non-Gaussianities (superscript "NG") and those coming from the late-time gravitational evolution (superscript "G"). Schematically, the perturbative theoretical prediction for the bispectrum is

$$B^{\rm th} = B^{\rm G}_{\rm SPT} + B^{\rm G}_{\rm EFT} + f_{\rm NL}\left(B^{\rm NG}_{\rm SPT} + B^{\rm NG}_{\rm EFT}\right). \tag{3}$$

As we will see in section 3, for a Fisher forecast we do not need to specify[1] $B^{\rm G}_{\rm SPT}$. The leading order counterterms for Gaussian initial conditions have been computed in [12, 13] and read

$$B^{\rm G}_{\rm EFT} = \xi B^{\rm G}_{\xi} + \sum_{i=1}^{3} \epsilon_i B^{\rm G}_{\epsilon_i}. \tag{4}$$

For non-Gaussian initial conditions, short modes and long modes are already correlated at the initial time. This leads to additional contributions to the matter bispectrum. To leading order, these are given by [14]

$$B^{\rm NG}_{\rm EFT} = \xi B^{\rm NG}_{\xi} + \gamma B^{\rm NG}_{\gamma} + \sum_{i=1}^{2} \gamma_i B^{\rm NG}_{\gamma_i}. \tag{5}$$

For convenience, we have adopted the notation of [14] and collected in appendix A all the explicit expressions for the terms appearing in this subsection.

2.2 Primordial non-Gaussianity

To evaluate the non-Gaussian contributions to (3), we need to specify the primordial bispectrum. In this paper, we study the constraints on three types of primordial non-Gaussianity: local [22], equilateral [23] and quasi-single field [24]. In terms of the primordial potential ϕ, the primordial bispectra are given by the following shapes

$$B^{\rm loc}_{\phi}(k_1,k_2,k_3) = 2 f^{\rm loc}_{\rm NL}\left(P_\phi(k_1)P_\phi(k_2) + {\rm perm}\right), \tag{6a}$$

$$B^{\rm eq}_{\phi}(k_1,k_2,k_3) = 162 f^{\rm eq}_{\rm NL} A^2_\phi \frac{1}{k_1k_2k_3K^3}, \tag{6b}$$

$$B^{\rm qsf}_{\phi}(k_1,k_2,k_3) = 18\sqrt{3} f^{\rm qsf}_{\rm NL} A^2_\phi \frac{1}{k_1k_2k_3K^3}\frac{N_\nu(8\kappa)}{\sqrt{\kappa}N_\nu(8/27)}. \tag{6c}$$

Here we define $K = k_1 + k_2 + k_3$, and $\kappa = k_1k_2k_3/K^3$. Moreover, N_ν is the Neumann function of order ν and we choose $\nu = \frac{1}{2}$. The normalization of the primordial power spectrum[2] is given by $A_\phi = 1.72 \cdot 10^{-8}$. To linearly evolve these to the late time matter bispectrum[3] B_{111}, we use the transfer function $M(k,z)$, defined by

$$\delta_1(k,z) = M(k,z)\phi(k), \quad \text{with} \quad M^2(k,z) \equiv \frac{k^3 P_{11}(k,z)}{A_\phi\left(\frac{k}{k_\star}\right)^{n_s-1}}. \tag{7}$$

Here $k_\star = 0.0028\ h{\rm Mpc}^{-1}$ and $n_s = 0.967$. This means we have

$$B_{111}(k_1,k_2,k_3,z) = M(k_1,z)M(k_2,z)M(k_3,z)B_\phi(k_1,k_2,k_3). \tag{8}$$

We collect all relevant higher order non-Gaussian contributions to the bispectrum in appendix A.

[1] On the other hand, we do need to specify the SPT contributions to the power spectrum to compute the cosmic variance. Assuming it is dominated by the linearly evolved matter power spectrum, we do not have to specify additional 'EFT' parameters.

[2] Note that we define $A_\phi = 2\pi^2 \frac{9}{25} A_\zeta$.

[3] See appendix A for relevant notation.

2.3 Theoretical error

By definition, any results from perturbation theory are approximate - there is always an intrinsic theoretical error, typically estimated within perturbation theory itself. The true bispectrum is therefore given by

$$B^{true} = B^{th} + B^{er}, \tag{9}$$

where B^{th} is the perturbative theoretical prediction given in (3), and B^{er} represents the theoretical error. The strength of a well defined perturbation theory is of course that the error can be estimated within the perturbation theory itself.

In our case, there are in principle two perturbative schemes employed. First, we assume perturbative primordial non-Gaussianity. This means we assume the primordial potential can be schematically expanded as

$$\varphi_p = \varphi_p^{\rm G} + f_{\rm NL}\varphi_p^{\rm G} \star \varphi_p^{\rm G} + \dots . \tag{10}$$

Here φ_p^G is a Gaussian field and $\star$ denotes a convolution in Fourier space. This means we are effectively expanding in $f_{\rm NL}\varphi_p \sim f_{\rm NL}\sqrt{A_\phi}$, which is indeed very small according to current bounds. Hence we will not worry about corrections to this approximation for the rest of the paper.

Second, the EFT of LSS relies on the smallness of density perturbations on large scales, consistently taking into account our ignorance of short scale physics. Effectively, this comes down to an expansion in $k/k_{\rm NL}$ [4]. As argued in [12], the most relevant correction to B^{th} is the two-loop bispectrum. Since we have not computed the full two-loop bispectrum, we are forced to make an educated guess about its size and shape. One way to do this was proposed in [15], and relies on the scaling universe results of [25]. Here we use instead a different estimate. Unless indicated otherwise, we estimate the two-loop bispectrum by adding up the absolute values of the two two-loop diagrams we can compute, namely the so-called reducible diagrams, which we indicate by B_{332}. An explicit expression for B_{332} is again given in appendix A. We compare our estimate to the scaling estimate of [15] in appendix B.5.

The importance of keeping track of the theoretical error for forecasts has recently been stressed in [15], and we build on their approach. Qualitatively, one expects not to be able to learn much about $f_{\rm NL}$ from bispectrum configurations for which B^{er} is larger than the non-Gaussian signal. To get an idea of the configurations for which this is the case, we plot the one-loop expressions for the three types of non-Gaussianities we consider (with $f_{\rm NL} = 1$) and B_{332} as a function of scale for both squeezed and equilateral configurations in Figure 1. For reference we also plotted the one-loop Gaussian contribution to the bispectrum. As expected, for local PNG we can push to smaller scales in the squeezed configuration than for equilateral PNG. Note also that the naive k_{max}, beyond which we do not expect to gain any more signal, is *configuration dependent*. A detailed discussion on how to incorporate this theoretical error in a Fisher analysis is given in section 2.3, which proceeds along the lines of [15]. In appendix B, we present further investigation of the validity of this method of treating the theoretical error.

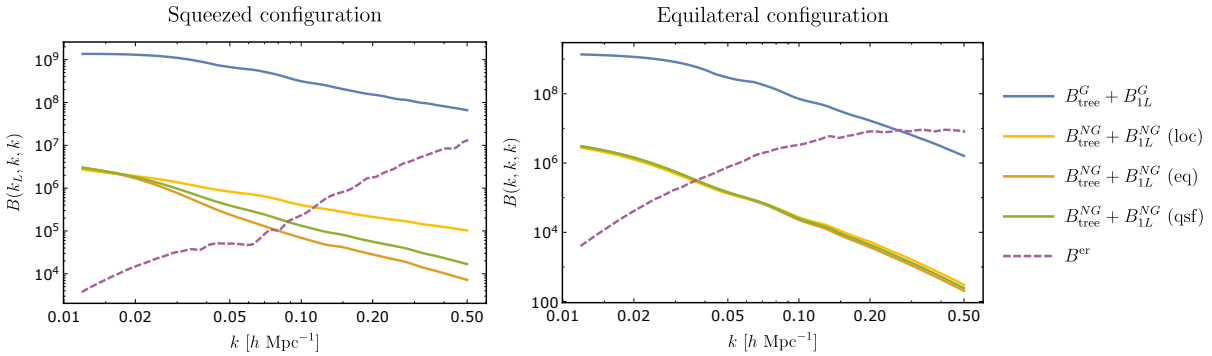


Figure 1: The SPT contributions to the bispectrum in the squeezed (left) and equilateral (right) configurations. The blue solid line denotes the Gaussian tree-level and one-loop contributions. The yellow, orange and green lines denote the one loop non-Gaussian contribution for $f_{\rm NL} = 1$ for local, equilateral and quasi-single-field PNG, respectively. The dashed purple line corresponds to our order-of-magnitude estimate for the Gaussian two-loop correction B_{332}. In the squeezed configurations (left), we chose $k_L = 0.012\,h\text{Mpc}^{-1}$.

3 Fisher analysis

In this section, we outline our method to forecast constraints on primordial non-Gaussianity. We have in mind a Gedankenexperiment that provides us with the matter distribution in space and time up to some maximal redshift. In this highly idealized scenario, we determine to what extent our inability to analytically describe the non-linear gravitational collapse of matter limits the information we can extract on primordial perturbations. We proceed along the lines of [6, 7, 26]. The outcome of the analysis for various surveys is presented in the next section.

3.1 Assumptions and approximations

For the convenience of the reader, we summarize the assumptions and approximations we make in the Fisher analysis.

- We assume we are given an idealized *survey of the late time dark matter density field*, instead of that of some biased tracer. This allows us to answer the question of whether further progress is needed in the modeling of the dark matter distribution to strengthen current bounds on PNG using upcoming LSS surveys.

- The idealized dark matter survey is characterized by a redshift range and the fraction of the sky covered. We divide the survey in redshift bins, to which we assign a fixed time that is equal to the mean redshift of the bin. Hence, we only need to know $z_{\rm bin}$ to predict the power spectrum and bispectrum. Observational redshift errors are neglected.

- We assume that each redshift bin can be approximated by a cube. Then we just need the volume of the bin $V(z_{\rm bin})$ to account for cosmic variance.

- We compute correlation functions only within each bin. This does not seem to be a big drawback in the case of equilateral PNG. Instead, for local PNG, this might cause an unnecessary loss of information. We will discuss this issue elsewhere.

- We include shotnoise in the analysis to correctly remove weight from the higher redshift bins. For this, we use the specifications of specific upcoming surveys. We discuss this in section 4.1.2.

- We assume that the bispectra for different configurations are uncorrelated with each other. This means that we approximate the *bispectrum covariance matrix as diagonal*. In [26] it has been checked that this approximation works fine for the scales $k \leq 0.3h\text{Mpc}^{-1}$ at redshift zero. We assume it holds up to $k \leq 0.4h\text{Mpc}^{-1}$, since for local PNG we still gain

signal up to this scale, as we see in Figure 9. Moreover, we assume that only the linear power spectrum determines the covariance matrix (see subsection 3.3 for more details). Finally, we neglect covariance due to observational effects, such as survey geometry and mask.

Importantly, we parameterize the theoretical error by treating the higher loop corrections to the bispectrum as a source of noise, which we integrate out. This contributes to the effective covariance matrix. Our estimate of the typical size of the two-loop corrections is given by B_{332}, defined in Appendix A.

The time-dependence of the counterterms has been chosen to match the one loop diagrams they are supposed to renormalize [27] (see also [28] for a related discussion). All the time dependence is absorbed in the contributions to the bispectrum, so that all the theoretical parameters become time-independent (see appendix A). This means that we are measuring the same theoretical parameters in each redshift bin.

We need to discretize the bispectrum in order to compute the Fisher matrix. We will use logarithmic bins instead of linear bins, since we do not fully trust linear binning. We refer the reader to the appendix C for more details. Finally, we do not marginalize over the standard cosmological parameters, but fix their values.

3.2 Fisher matrix

In a Fisher forecast, one computes the expected curvature around the maximum of the likelihood. The likelihood is given by

$$\mathcal{L}(\text{data}|\Theta,\ \text{priors}) = \frac{1}{\sqrt{(2\pi)^N \det(C_B)}} P_{\text{prior}}(\Theta) \times \exp\left(-\frac{1}{2}\sum_{\hbar p} \Delta B(\hbar,\Theta) C_B^{-1}(\hbar,q,\Theta)\Delta B(p,\Theta)\right), \tag{11}$$

where Θ denote the set of theoretical parameters and N is the number of datapoints. We suppressed the time dependence. Here we use the Dutch calligraphic lower case symbols $\hbar$ as a shortcut for a triplet of wavenumbers on which the bispectrum depends, i.e. $\hbar = (k_1, k_2, k_3)$. Furthermore, ΔB corresponds to the difference between estimator and theoretical prediction $\Delta B(\hbar,\Theta) = \hat{B}(\hbar) - B(\hbar,\Theta)$ and C_B is the covariance matrix of the bispectrum $C_B = \langle \Delta B \Delta B \rangle$.

Neglecting the theoretical error for the moment, the theoretical prediction for the bispectrum is given in equation (3). The 8 parameters we include in the Fisher analysis are therefore $\{f_{\text{NL}},\ \xi,\ \epsilon_1,\ \epsilon_2,\ \epsilon_3,\ \tilde{\gamma},\ \tilde{\gamma}_1,\ \tilde{\gamma}_2\}$[4]. The parameter ξ also appears in the power spectrum and has been measured to be[5] 0.98 $h^{-2}\text{Mpc}^2$ [27]. Therefore, we can put a sharp prior on this parameter. The other parameters are unknown, but we expect them to be of the same order of magnitude (see [12, 14] for naive numerical estimates). Therefore, we can use a fiducial value of zero and a Gaussian prior with variance of 10.

For simplicity, we assume that all priors are Gaussian with covariance matrix C_Θ. Then the Fisher matrix is given by (see e.g. [30])

$$F_{ij} \equiv -\left\langle \frac{\partial^2 \log(\mathcal{L})}{\partial\Theta_i \partial\Theta_j} \right\rangle = \frac{1}{2}\text{Tr}\left[C_B^{-1} C_{B,i} C_B^{-1} C_{B,j}\right] + B_i^T C_B^{-1} B_j + \left(C_\Theta^{-1}\right)_{ij}. \tag{12}$$

[4]Here we denote $\tilde{\gamma} = f_{\text{NL}} \cdot \gamma$ and similarly for γ_i, so that the bispectrum is linear in all parameters. This is convenient for the Fisher analysis, as this makes the result independent of the fiducial values of the parameters. On the other hand, we effectively assume that the one-loop non-Gaussian counterterms have amplitudes independent of f_{NL}. Later in this paper we will find that these counterterms are negligibly small, therefore, this will not affect our results.

[5]Previous measurement gave (1.62 ± 0.03) $h^{-2}\text{Mpc}^2$ [29] and (1.5 ± 0.03) $h^{-2}\text{Mpc}^2$ [12], but neglected two-loop corrections.

As we will see in a moment, our approximation for the covariance matrix does not depend on the theoretical parameters. Writing out the time dependence explicitly, the Fisher matrix simplifies to

$$F_{ij}(z) = \sum_{\hbar,\hbar'} B_i^T(\hbar, z) C_B^{-1}(\hbar, \hbar', z) B_j(\hbar', z) + \left(C_\Theta^{-1}\right)_{ij}(z) \tag{13}$$

for each redshift bin. Since the bispectrum is linear in all parameters - taking into account that ξ has been measured - the Fisher matrix is independent of the fiducial value of the theoretical parameters[6], which is very convenient for the analysis. To combine the results from the different redshift bins, we use that the parameters are the same in each bin, since we have fixed their time dependence. This time dependence is chosen to match the time dependence of the divergences they are supposed to cancel, motivated by [27]. The explicit expression can be found in Appendix A. Therefore, we can compute the constraints on $f_{\rm NL}$ by summing the Fisher matrices and then marginalizing over the EFT parameters i.e.,

$$\sigma(f_{\rm NL}) = \sqrt{\left(\sum_z F_{ij}(z)\right)^{-1}_{11}}, \tag{14}$$

where we assumed that the entry of the Fisher matrix belonging to $f_{\rm NL}$ is the first. Note that we have not included cross correlations between bins, which means we might be throwing away valuable information. This effect will be studied elsewhere [31].

3.3 Covariance matrix

To evaluate the Fisher matrix, we need to know the covariance matrix. Let us shortly review the derivation of the bispectrum covariance matrix. The estimator of the bispectrum is given by [32]

$$\hat{B}(k_1, k_2, k_3, z) = \frac{1}{V(z) V_{123}} \int_{\mathbf{q}_1} \int_{\mathbf{q}_2} \int_{\mathbf{q}_3} \delta(\mathbf{q}_1, z)\delta(\mathbf{q}_2, z)\delta(\mathbf{q}_3, z)\delta_D^3(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3), \tag{15}$$

with $V(z)$ the volume of the bin. The integration is over logarithmic bins centered around the given wavenumbers, i.e. $\ln(|\mathbf{q}_i|) \in [\ln(k_i) - \frac{1}{2}\Delta\ln(k), \ln(k_i) + \frac{1}{2}\Delta\ln(k)]$. Moreover, V_{123} corresponds to the following k-space volume squared

$$V_{123} = \int_{\mathbf{q}_1} \int_{\mathbf{q}_2} \int_{\mathbf{q}_3} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \approx \frac{8\pi^2}{(2\pi)^9} k_1^2 k_2^2 k_3^2 \sinh^3(\Delta \ln k). \tag{16}$$

This approximation becomes exact when we consider 'internal' bins, but it fails on the 'edge' bins. In the numerical analysis we compute the exact value of V_{123} for each bin, see appendix C for more details. This allows us to compute the bispectrum covariance matrix, namely

$$C_B(\hbar, \hbar', z) = \left\langle \left(\hat{B}(\hbar, z) - B(\hbar, z)\right)\left(\hat{B}(\hbar', z) - B(\hbar', z)\right)\right\rangle \tag{17}$$

$$\approx \frac{s_{123}}{(2\pi)^3 V(z) V_{123}} P(k_1, z) P(k_2, z) P(k_3, z) \delta_{\hbar,\hbar'}. \tag{18}$$

There is a factor s_{123} which counts the number of non-vanishing contractions when computing $\langle \hat{B}(\hbar)\hat{B}(\hbar)\rangle$, which depends on the type of triangle that the triplet $\hbar$ forms. As each contraction comes with a delta function, this counting factor equals 6, 2 or 1 for equilateral, isosceles and scalene triangles respectively. If we include shotnoise in the covariance matrix, we replace $P(k_i, z)$

[6] To be more precise, the Fisher matrix is independent of the fiducial values of the parameters *to good approximation*. We choose a fiducial value for ξ of zero and in Section 4 we either specify a prior with $\sigma = 1$ for ξ or no prior at all. Therefore, the Fisher matrix has some dependence on the choice of ξ, but it will come exclusively from the non-Gaussian counterterm proportional to ξ. Again, since the non-Gaussian counterterms turn out to be extremely small, we expect this not to affect the results. Moreover, we have checked this explicitly by changing its fiducial value to 1.

with $P(k_i,z)+\frac{1}{\bar{n}}$ in equation (18), where $\bar{n}$ is the effective number density for the density contrast. This will be explained more when we include shotnoise in Section 4.1.

In this expression for the covariance matrix, we completely neglected higher order corrections beyond the power spectrum, making it approximately diagonal. In [26] it has been checked that this approximation works fine for the scales we are considering. The off-diagonal terms become important exactly when the higher order corrections to the power spectrum become important, since they are of the same order. Therefore, in order to be consistent, we only take into account the linear contribution P_{11} to the power spectrum $P(k_i,z)$. In particular, this means that the covariance matrix is independent of the theoretical parameters.

3.4 Theoretical error as nuisance parameters

To account for the theoretical error inherent to the perturbative expansion, we parameterize the bispectrum as

$$B(\hbar) = B^{\rm th}(\hbar) + n(\hbar)B^{\rm er}(\hbar)\,, \tag{19}$$

$$B^{\rm th}(\hbar) = B^{\rm G}_{\rm SPT}(\hbar) + B^{\rm G}_{\rm EFT}(\hbar) + f_{\rm NL}\left[B^{\rm NG}_{\rm SPT}(\hbar) + B^{\rm NG}_{\rm EFT}(\hbar)\right]\,, \tag{20}$$

where $B^{\rm th}$ represents the theoretical prediction up to some order in perturbation theory as before, and $B^{\rm er}$ is the estimate of the theoretical error. Following [15], we introduce a series of nuisance parameters $n(\hbar)$, one per bin in k-space. The reason we implement the theoretical error this way, instead of proposing some k_{max} is that, as discussed in subsection 2.3, k_{max} depends on where the theoretical error and the signal become comparable. This complicates the analysis in two ways. First, k_{max} is configuration dependent, and second, it depends on the fiducial value of $f_{\rm NL}$, which makes finding the error on $f_{\rm NL}$ a recursive problem. In the approach we take, the set of theoretical parameters thus becomes $\Theta = \{n(\hbar)\}_\hbar \cup \{f_{\rm NL},\ \xi,\ \epsilon_1,\ \epsilon_2,\ \epsilon_3,\ \tilde{\gamma},\ \tilde{\gamma}_1,\ \tilde{\gamma}_2\}$. Since the bispectrum remains linear in all parameters, expression (13) for the respective block of the Fisher matrix still applies.

We assume that the true corrections to the bispectrum are of similar size as $B^{\rm er}(\hbar)$. Therefore, we put a Gaussian prior on the parameters $n(\hbar)$, with mean zero and variance one. Moreover, we expect the correction to have a smooth shape, which varies not too rapidly within the contours defined by $B^{\rm er}(\hbar)$. Therefore, the coefficients should have non-negligible cross correlations. Since we would like to have an increasing correlation for nearby points, we include cross-correlations with a typical correlation length as follows

$$N_{\alpha\beta} = \exp\left(-\frac{\sum_i \ln(|k^i_\alpha/k^i_\beta|)}{l}\right). \tag{21}$$

We replaced the label $\hbar_\alpha$ of the nuisance parameter of a given bin with the index α, so that we can reserve latin indices for the other theoretical parameters. Moreover, in order not to confuse this covariance matrix with the covariance matrix of the bispectrum C_B, we denote it as $N_{\alpha\beta}$. Note that we choose $\sigma_\alpha = 1$ for all α's. Here, l denotes the logarithmic correlation length. We could have also chosen a quadratic correlation length, similar to modeling it as a random field [33]. The reason we opted for this form is that here the inverse matrix is very sparse, which is convenient for numerical purposes. Since our final results are quite insensitive to the correlation length (see B.3.3), we do not believe this choice affects the results very much.

Since we introduced a set of new nuisance parameters, we should write down the full Fisher matrix $F_{\mu\nu}$ and invert it

$$F^{-1}_{\mu\nu} = \begin{pmatrix} F_{\alpha\beta} & F_{\alpha j} \\ F_{i\beta} & F_{ij} \end{pmatrix}^{-1} = \begin{pmatrix} \bullet & \bullet \\ \bullet & (F_{ij} - F_{i\gamma}F^{-1}_{\gamma\delta}F_{\delta j})^{-1} \end{pmatrix}, \tag{22}$$

where we use latin indices for the parameters $\{f_{\rm NL},\ \xi,\ \epsilon_1,\ \epsilon_2,\ \epsilon_3,\ \tilde{\gamma},\ \tilde{\gamma}_1,\ \tilde{\gamma}_2\}$ and greek indices from the early alphabeth for the theoretical error parameters $\{n(\hbar_\alpha)\}_\alpha$. We did not write out explicitly

the other entries, since we are only interested in the effective Fisher matrix, after marginalizing over the nuisance parameters coming from the theoretical uncertainty. To compute the effective Fisher matrix, we need to know $F_{\alpha\beta}$ and $F_{\alpha i}$. Since the derivative of the bispectrum with respect to the nuisance parameters Θ_α is only non-zero for the corresponding bin, these contributions to the Fisher matrix are particularly simple. We have

$$F_{\alpha\beta} = B^{\rm er}(k_\alpha) C_B^{-1}(k_\alpha, k_\beta) B^{\rm er}(k_\beta) + N^{-1}_{\alpha\beta} \equiv D_{\alpha\beta} + N^{-1}_{\alpha\beta}\,, \tag{23}$$

and similarly

$$F_{i\beta} = \sum_k B_i(k) C_B^{-1}(k, k_\beta) B^{\rm er}(k_\beta) \quad \text{and} \quad F_{\alpha j} = \sum_k B^{\rm er}(k_\alpha)(k_\alpha) C_B^{-1}(k_\alpha, k) B_j(k). \tag{24}$$

This allows us to compute $F^{\rm eff}_{ij}$ by using (22). After some algebraic manipulations, we can rewrite it in the simple form

$$F^{\rm eff}_{ij} = \sum_{k,p} B_i(k) \left(N^{\rm eff}(k,p) + C_B(k,p) \right)^{-1} B_j(p) + \left(C_\Theta^{-1}\right)_{ij}\,, \tag{25}$$

with $N^{\rm eff}_{\alpha\beta} = B^{\rm er}(k_\alpha) N_{\alpha\beta} B^{\rm er}(k_\beta)$. Again, the time dependence has been suppressed. In Appendix B we present two alternative derivations of(25) and provide further detail.

In the next subsection we show the effectiveness of the current treatment of the theoretical error. However, we believe the interpretation of this method, and its relation to the actual situation, is a subtle matter. In particular, in Appendix B.3 we argue by means of a simple toy model that this way of dealing with the theoretical error is certainly not the right way in the extremes of zero and maximal correlation among the parameters. Namely, on the one hand the theoretical error acts as shot noise per bin for zero correlation length, whereas for maximal correlation it acts as a a single coefficient multiplying a fixed shape, effectively reducing the uncertainty about its shape to one number. Neither of these cases correspond to the way we believe the theoretical error should act. At the same time, Appendix B.3.3 shows that the effect of the correlation length on the results is very mild. This suggests that the main reason our method works so well is that our ansatz for the error is a much steeper function of k than the signal, so that the size of the error is much more important than its shape. Thus, even though our treatment of the theoretical error seems to work for the current case, we recommend a conservative use of the method. In this spirit, we use the correlation lenght that gives the most pessimistic results for the analysis, which we found to be $l \sim 0.5$.

3.5 Testing the effect of the theoretical error

To test the method of integrating out the theoretical error, we study its effect on the constraints on f_{NL} in a χ^2-analysis. To that end, we compare two types of analyses, one which includes the theoretical error as outlined above, and one which does not. We generate a fake dataset with no primordial non-Gaussianity to test the theory. Our datapoints are given by

$$D(k) = B_{112}(k) + E_b(k) + \text{cosmic noise}, \tag{26}$$

where we add some random noise, with variance equal to the cosmic variance, to each point. We consider a survey at redshift $z = 0$ with volume $V = 10\ (h^{-1}\text{Gpc})^3$, and restrict $(k_1,\ k_2,\ k_3)$ to be the central values of the binned range $[0.001, 1]\ h\text{Mpc}^{-1}$, where we take 27 logarithmic sized bins. The additional contribution to the bispectrum is given by

$$E_b(k_1, k_2, k_3) = 3 B_{112}(k_1, k_2, k_3) \left(\frac{k_1 + k_2 + k_3}{3 k_{NL}} \right)^{(3+n)l}\,, \tag{27}$$

with $n = -1.4$, $k_{NL} = 0.45$ and $l = 2$. This is exactly the ansatz for the two-loop contribution to the bispectrum used in [15] and it is based on scaling universes [25]. In appendix B.5, we

compare the ansatz for the higher loop corrections E_b with our ansatz B_{332}. As theoretical model for the bispectrum, we use

$$B^{\rm th}(k) = f_{NL} \cdot B_{111}(k) + B_{112}(k). \tag{28}$$

We now consider two analyses. In the first analysis, we neglect theoretical errors and take only cosmic variance into account. In the second analysis, we use our ansatz for the higher order corrections, namely B_{332}, and we account for both theoretical error and cosmic variance. In order to find the best fit value for f_{NL}, we minimize χ^2_B, which is given by

$$\chi^2_B = \Big(D(k) - B^{\rm th}(k)\Big)\Big(C_B(k,p) + N^{\rm eff}(k,p)\Big)^{-1}\Big(D(p) - B^{\rm th}(p)\Big) + \text{const.}, \tag{29}$$

see (92) in Appendix B. In the first case, we set $N^{\rm eff}$ to zero. Minimizing χ^2_B yields

$$\text{Est}(f_{NL}) = \frac{B_{111}(k)\left(C_B(k,p) + N^{\rm eff}(k,p)\right)^{-1}\left(D(p) - B_{112}(p)\right)}{B_{111}(k)\left(C_B(k,p) + N^{\rm eff}(k,p)\right)^{-1}B_{111}(p)}, \tag{30}$$

and taking another derivative with respect to f_{NL} allows us to compute the standard deviation

$$\sigma(f_{NL}) = \left(B_{111}(k)\Big(C_B(k,p) + N^{\rm eff}(k,p)\Big)^{-1}B_{111}(p)\right)^{-1/2}. \tag{31}$$

With the best-fit value of f_{NL}, we can evaluate $\left(\chi^2_B\right)_{\rm red}$ and the p-value, which are given by

$$\left(\chi^2_B\right)_{\rm red} = \frac{\chi^2_B}{N} \quad \text{and} \quad p\text{-value} = 1 - \text{CDF}_{\chi^2}(N, \chi^2_B), \tag{32}$$

with CDF_{χ^2} the cumulative distribution function of the χ^2-distribution, and $N = N_{\rm bins} - N_{\rm dofs} - 1$ the number of datapoints minus one minus the number of fitting parameters. The p-value takes values between 0 and 1. It gives the probability of finding a higher value for χ^2_B if it was drawn from a χ^2-distribution. Therefore, it should take values around to 0.5. If the p-value is very close to zero, then the proposed theory vector is not a good description of the data. If the p-value is close to one, then either one is overfitting the data, or the estimate for the noise is too pessimistic.

In Figure 2, we plot the estimate for f_{NL} with errorbars, $\left(\chi^2_B\right)_{\rm red}$ and the p-value, as given in equations (30), (31) and (32) respectively, for the two analyses. In the left panel, we show both the results for the analysis in which the higher order corrections are neglected, and the analysis in which we use B_{332} as an ansatz. In the right panel, we use $10 \times B_{332}$ as error estimate to make sure that our ansatz is always bigger than the true value of the higher order corrections. One can check that E_b has a different shape than B_{332}. For instance, in the equilateral configuration, E_b is smaller than B_{332} on small scales (more optimistic). On the other hand, on large scales in the equilateral configuration, and in the squeezed limit, it tends to be larger than B_{332} (more pessimistic). Upon multiplying the latter by a factor 10, we find a robust, conservative estimate (see B.5).

In the left panel of Figure 2, we see that if we neglect the theoretical error (blue lines and contours), we get the wrong value for the best fit value for f_{NL}, because higher order corrections are mistakingly interpreted as signal. Fortunately, the p-value singles out where the theoretical description fails. Taking this into account, we get a reliable estimate for f_{NL}, albeit with larger errorbars, since we have to stop already at a relatively small k_{max}. From the analysis that accounts for the theoretical error (red lines and contours), it seems we can continue the analysis to a higher k_{max}. However, the result we get for f_{NL} is biased, i.e. it is more than 5σ away from the actual value. The problem is that, in certain configurations of the bispectrum, our ansatz takes smaller values than the actual value in the data. This tends to decrease the p-value. At the same time, the p-value increases in the configurations where the theoretical error is overestimated. The interplay of these two effects can lead to a p-value, which is neither too small or too large,

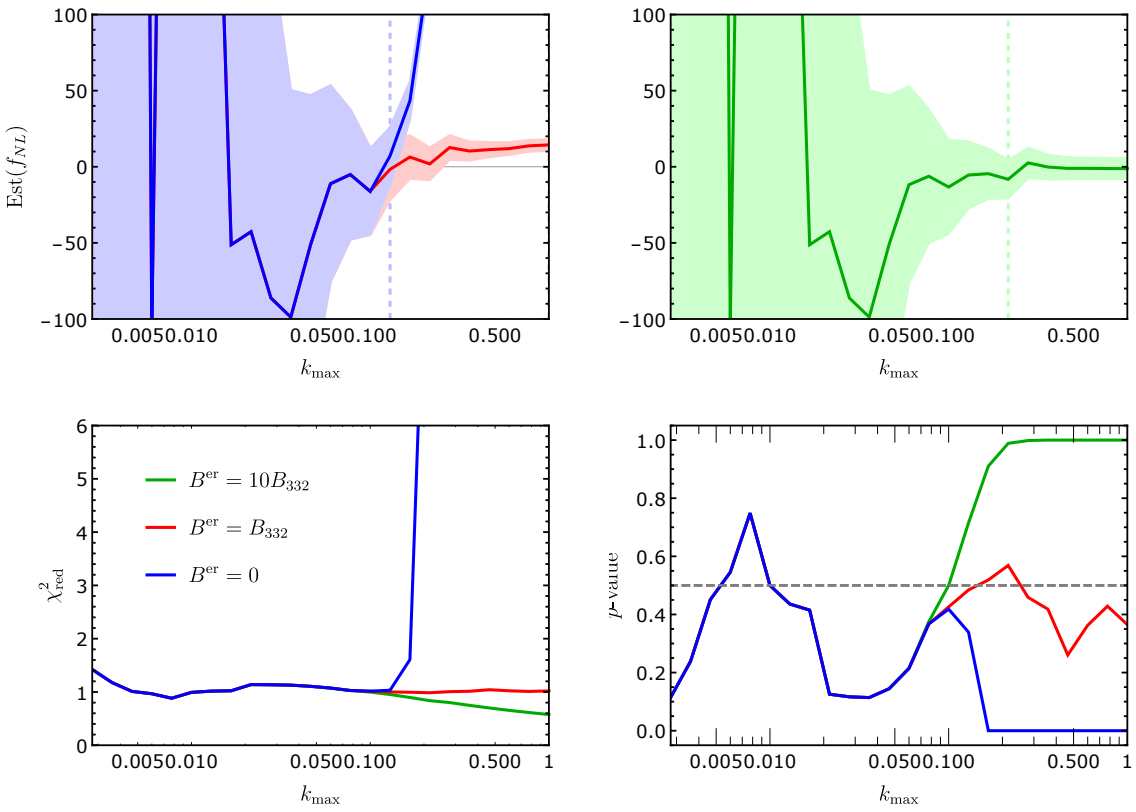


Figure 2: The figure shows the results from the χ-squared analysis for the data and theory given in equations (26) and (28). In the left panel, we show the results for both the analysis in which the higher order corrections are neglected (blue) and the one in which we use B_{332} as ansatz (red). In the right panel, we use $10B_{332}$ as ansatz (green) instead. In the upper panels, we show the best fit value for f_{NL} (solid line) as function of k_{max} and the lighter coloured regions correspond to the 2σ errorbars. The dashed vertical line corresponds to the largest value for k_{max} where the p-value is still between 0.01 and 0.99. The second and third row show $\left(\chi^2_B\right)_{\rm red}$ and the p-value as function of k_{max}.

and this gives rise to a biased estimate. Hence, if one wants to use the p-value as diagnostic for k_{max} and avoid biased results, it is important to have a fairly good understanding of the form of the theoretical noise. Alternatively, one can work with an ansatz which is consistently *underestimating* the theoretical error. In this case, the p-value should go to zero rapidly, as soon as the theoretical error kicks in. As a double check, we did the analysis using $0.1E_b$ as ansatz instead, which indeed gives unbiased results similarly to the case where we neglect the theoretical error altogether. In general, when performing a datafit, if one has insufficient information about the higher order corrections, it is therefore safer not to integrate out the theoretical error at all. Summarizing, *assuming the wrong shape for the theoretical error might lead to a false detection of primordial non-Gaussianity.*

In the right panel (green lines and contours), we show the same results, where now the ansatz is always more pessimistic than the actual theoretical error ($10 \times B_{332} > E_b$). In this case, the estimate for f_{NL} is equal to the real value within 2σ. The p-value now is very *large* and it would naively tells us to stop at some smaller k_{max}. However, since we are obviously not overfitting the data, this reflects the fact that our ansatz for the theoretical error is too pessimistic. Therefore, we can safely evaluate the estimate for f_{NL} and the corresponding errorbar at the highest value of k_{max}, where the errorbar is frozen to a finite value. As expected, we find better errorbars than in the case we neglect the theoretical error. This shows that integrating out the theoretical error helps constraining f_{NL}, as long as one is careful to take a conservative enough ansatz. In the

next section, we take B_{332} as ansatz for the theoretical error[7].

4 Results

In this section, we present the main results of our analysis. First, we give $\sigma(f_{\rm NL})$ for various surveys, comparing our results to [15] and [3]. Next, we study the correlations among the EFT parameters for relevant surveys. Furthermore, we address the question of how much better the constraints would be, if we were able to compute the two-loop bispectrum. Finally, we show that the EFT of LSS clearly outperforms SPT in the constraints on $f_{\rm NL}$, where we assume the EFT contributions to the bispectrum are part of the theoretical error in SPT.

4.1 Constraints as function of $z_{\rm max}$

In this subsection, we compute the constraints on f_{NL} as function of maximum redshift for surveys similar to the ones studied in [15] and [3]. This allows us to study the effects of shotnoise and to compare our results with theirs. Furthermore, we show the effect of marginalizing over the EFT parameters for these surveys.

4.1.1 A large redshift survey (comparison with Baldauf et al. 2016)

First, to compare with [15], we focus on local and equilateral PNG. In the following, we list the specifications of the survey and the particular assumptions we make (in addition to general assumptions for the Fisher analysis discussed in subsection 3.1).

- We assume a survey with maximum redshift of $z = 5$. We divide the survey in redshift bins of equal volume, where the first bin runs from $z = 0$ to $z = 1$. We assume the survey covers 20000 deg^2, which means that, with our choice of cosmological parameters, the volume of each bin is given by $V = 26.5(h^{-1}{\rm Gpc})^3$. We approximate each bin to be a cube, so that $k_{\rm min} \approx 0.002\ h{\rm Mpc}^{-1}$. The maximum redshifts of all the redshift bins are given by[8]

$$\{1.00, 1.39, 1.71, 2.02, 2.31, 2.60, 2.89, 3.19, 3.49, 3.80, 4.11, 4.44, 4.78, 5.13\}.$$

- We restrict $(k_1,\ k_2,\ k_3)$ to the values in the binned range $[0.002, 1]\ h{\rm Mpc}^{-1}$, where we use 15 logarithmic bins per decade[9].

- At high redshift, the late-time non-Gaussian background is particularly small and so the PNG signal becomes comparatively more pronounced. One the other hand, at high redshift, there are also much fewer tracers and this degrades our ability to measure the distribution of matter. To be able to capture this fact, we introduce a shotnoise that mimics what happens for example in galaxy surveys. For the purpose of comparison, we adopt the same convention for shotnoise as in [15], namely

$$\bar{n} = b_1^2 n_0 (1+z)^\alpha, \tag{33}$$

  where we correct for the galaxy bias $b_1 = 2$, since the shotnoise in [15] applies to galaxies, while here it has been translated to the dark matter density field. We choose $n_0 = 10^{-3} h^3 {\rm Mpc}^{-3}$ and $\alpha = -1$.

[7] We checked that our results for $\sigma(f_{NL})$ change with less than a factor of 2, when we use $10B_{332}$ instead.

[8] Here, and in the next section, we make a particular choice of redshift bins. Larger redshift bins imply that we can include more configurations of the bispectrum in the analysis, in particular more configurations in the squeezed limit. At the same time, we fix the redshift of the bin to be the mean redshift, therefore, larger redshift bins imply a smaller maximum redshift. Therefore, the choice of binning might affect the final result. We will study a bin-independent approach in [31].

[9] This is a little less than 45 bins over the full range, corresponding to $O(3000)$ triangles.

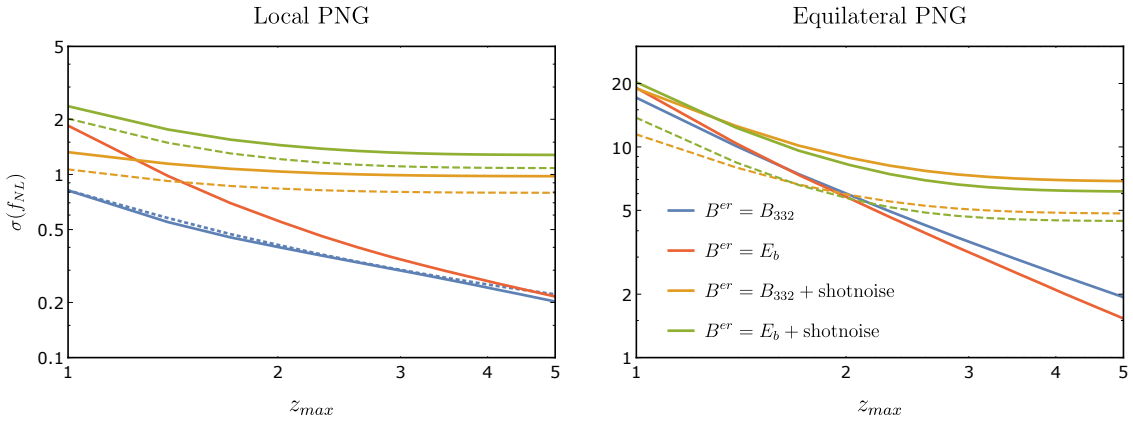


Figure 3: In these plots we show $\sigma(f_{NL})$ as function of maximum redshift $z_{\rm max}$ for local PNG (left) and equilateral PNG (right). We include shotnoise with the same specifications as [15] (orange and green lines). For the green lines we used E_b as ansatz for B^G_{2L} whereas for the other lines we used B_{332}. The blue solid line shows the result without shotnoise. The solid lines show the marginalized results assuming Gaussian priors for the EFT parameters. The dashed lines correspond to the unmarginalized result with the same color. Finally, the dotted blue line in the left panel corresponds to the curve $\sigma(z_1)/\sqrt{N}$ with N the number of redshift bins.

- When we marginalize, we assume for each EFT coefficient a Gaussian prior with $\sigma = 10$, except for the EFT parameter ξ for which we take $\sigma = 1$.
- For $B^{\rm er}$ we consider both B_{332} and the ansatz E_b given in [15] (see (27)).

The results for $\sigma(f_{NL})$ are shown in Figure 3. We show the effects of the ansatz for $B^{\rm er}$, shotnoise, and marginalization over the EFT parameters.

We can compare our results with those found in [15] by looking at the unmarginalized results, using their ansatz E_b for $B^{\rm er}$. Thus, we should compare our dashed green lines with their dotted red lines in the last columns of their Figure 6 and 7. We indeed find a reasonably good agreement, given the fact that our analyses are not identical (different sky coverage, redshift bins and cosmological parameters, and moreover the translation of their shotnoise to ours is not perfect as we only took into account b_1). This check confirms that our code runs as expected.

Let us now study the effect of the EFT parameters. The solid lines in Figure 3 correspond to marginalization over the EFT parameters, where we assume a Gaussian prior with $\sigma = 10$ for each EFT coefficient, except for the EFT parameter ξ, for which we take $\sigma = 1$. We see that *the results for local PNG are almost unaffected by marginalizing over the EFT parameters*. The constraints on equilateral PNG weaken by a factor of about two, however.

Using B_{332} as an ansatz for $B^{\rm er}$, we find slightly more optimistic results for local PNG, and slightly worse results for equilateral PNG, as compared with [15]. This can be understood from the comparison between B_{332} and E_b, shown in appendix B.5. Local PNG peaks in the squeezed limit, and B_{332} is more optimistic than E_b in this configuration. On the other hand, equilateral PNG peaks in the equilateral configuration, and in this configuration E_b is more optimistic.

If we neglect shotnoise, we find that the differences at low redshifts are even bigger for the two ansätze. The difference is largest for local PNG, since B_{332} is an order of magnitude bigger than E_b in the squeezed configuration, whereas the difference in the equilateral configuration is only of the order of a few. However, at higher $z_{\rm max}$ the differences disappear. This might seem a bit strange at first. However, we should stress that what we find here is not the true result in case of zero shotnoise. It turns out that with no shotnoise, we can always gain information in the ultra squeezed limit, and at higher redshift we can actually go to higher k than our choice $k_{\rm max} = 1h{\rm Mpc}^{-1}$. Since $\sigma(f_{NL})$ does not freeze before we reach $k_{\rm max}$, this explains why the red and blue curves can get close for high $z_{\rm max}$. In appendix B.5, we show these statements explicitly.

For comparison, note that if we were to always gain signal up to the same $k_{\rm max}$ in each redshift bin, we would find roughly the same $\sigma(f_{NL})$ for each redshift bin[10]. Therefore, combining all the redshift bins, we should find a $1/\sqrt{N_{\rm bins}}$ behavior if we neglect shotnoise. The dotdashed blue line in the figure corresponds to $\sigma(z_1)/\sqrt{N_{\rm bins}}$, which indeed resembles the blue solid line quite well. This provides another indication that we can go up to higher $k_{\rm max}$. Interestingly, we find that we can also go to much smaller scales for equilateral PNG than suggested by scaling estimates for $k_{\rm max}$ (see for instance [15]). The squeezed limit allows us to extract more information, also for equilateral PNG.

If we include shotnoise, it correctly cuts off the signal before we reach $k_{\rm max}$, so these results are reliable. However, one should keep in mind that, for more optimistic galaxy number densities, we might still extract more information from the ultra-squeezed limit. For the number densities we consider at the higher redshifts, shotnoise is the dominant source of noise. This is why we do include shotnoise in our analysis.

4.1.2 Current and upcoming surveys (comparison with Tellarini et al. 2016)

Next, we compare with [3], using the specifications of the surveys Euclid [34], BOSS [35], eBOSS [36] and DESI [37]. We have to consider the emission line galaxies (ELG), the luminous red galaxies (LRG) and quasars (QSO) separately, as they are measured at different redshifts, and have different number densities and bias coefficients. The specifications and assumptions are as follows.

- For the precise number densities and bias coefficients as function of redshift we refer to appendix D of [3]. Moreover, one can find here the fraction of sky covered by each survey.
- As before, we divide each survey in equal sized redshift bins. The boundaries of all the redshift bins are given by

$$\begin{aligned}
\text{eBOSS (ELG)}&: \{0.6, 0.8, 0.95, 1.09, 1.21\} && V_{\rm bin} = 2.8(h^{-1}\text{Gpc})^3,\\
\text{DESI (ELG)}&: \{0.1, 0.6, 0.79, 0.94, 1.07,\\
&\qquad 1.19, 1.3, 1.4, 1.5, 1.59, 1.69, 1.78\} && V_{\rm bin} = 5.4(h^{-1}\text{Gpc})^3,\\
\text{Euclid (ELG)}&: \{0.6, 1., 1.28, 1.53, 1.75, 1.97\} && V_{\rm bin} = 14.0(h^{-1}\text{Gpc})^3,\\
\text{eBOSS (LRG)}&: \{0.6, 0.75, 0.87, 0.98\} && V_{\rm bin} = 2.0(h^{-1}\text{Gpc})^3,\\
\text{DESI (LRG)}&: \{0.1, 0.6, 0.79, 0.94, 1.07\} && V_{\rm bin} = 5.4(h^{-1}\text{Gpc})^3,\\
\text{BOSS (LRG)}&: \{0., 0.4, 0.52, 0.61, 0.68, 0.75, 0.8\} && V_{\rm bin} = 1.3(h^{-1}\text{Gpc})^3,\\
\text{eBOSS (QSO)}&: \{0.6, 1., 1.28, 1.53, 1.75, 1.97, 2.17\} && V_{\rm bin} = 6.6(h^{-1}\text{Gpc})^3,\\
\text{DESI (QSO)}&: \{0.1, 0.8, 1.08, 1.31, 1.51, 1.7, 1.89\} && V_{\rm bin} = 11.0(h^{-1}\text{Gpc})^3.
\end{aligned}$$

- We use $k_{\rm min}$ determined by the volume of each bin. Moreover, we choose the same binning of the k-range as in Section 4.1.1.
- The ansatz for shotnoise is now $\bar{n}(z) = b_1^2(z)n(z)$, where we correct for galaxy bias, similar as before.
- When we marginalize, we take the same prior as before. We assume for each EFT coefficient a Gaussian prior with $\sigma = 10$, except for the EFT parameter ξ for which we take $\sigma = 1$.
- As ansatz for the higher order corrections we use $B^{\rm er} = B_{332}$.

[10] To good approximation: the entry of the Fisher matrix, corresponding to f_{NL}, will scale as $F \sim D(a)^6/D(a)^6 \sim 1$, if we neglect the loop corrections to B_{NG}. Then, forgetting about the marginalization over the EFT parameters, we find the same $\sigma(f_{NL})$ in each redshift bin, since we took the bins so that they have the same volume.

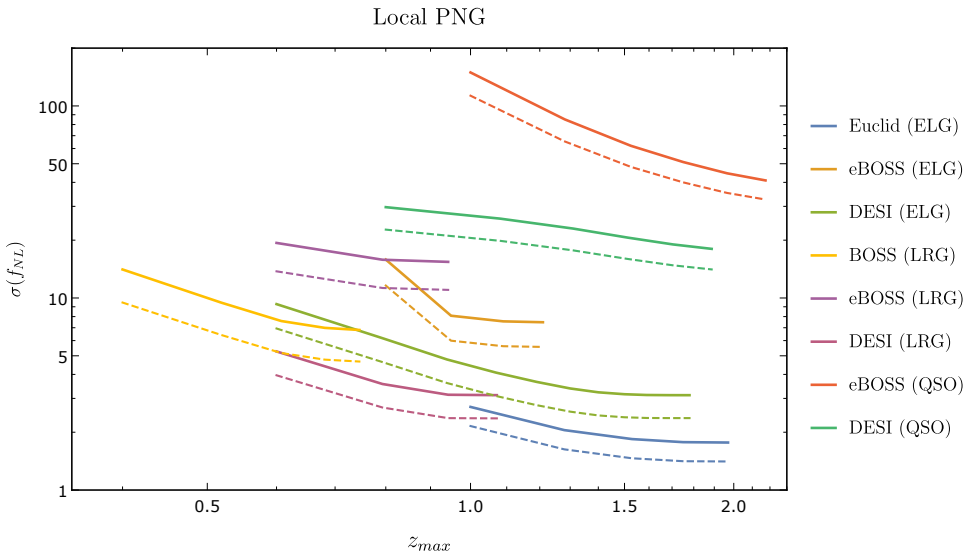


Figure 4: We show $\sigma(f_{NL})$ as function fo $z_{\rm max}$ for local PNG. We use the specification from Euclid (blue), BOSS (yellow), eBOSS (orange, purple and red) and DESI (green/yellow, pink and green). We show both the marginalized (solid lines) and unmarginalized results (dashed lines).

The results for local PNG are shown in Figure 4. We plot $\sigma(f_{NL})$ as function fo $z_{\rm max}$ for the four surveys. We show both the marginalized and unmarginalized results. After combining the different galaxy catalogs of a survey, we get $\sigma(f_{NL})$ for each survey, as summarized in Table 1a for local PNG and in Tables 1b and 1c for equilateral and quasi-single field PNG respectively.

We compare our results with [3] by looking at our unmarginalized results in Table 1a and their results in the last column in Table 1 of their paper. We find much weaker constraints, varying from 4 to 8 times smaller. This can be explained by the fact that we account for the theoretical error, which freezes the errorbars. Therefore, including the theoretical error gives rise to more conservative constraints. Moreover, scale-dependent bias could actually help us constrain local PNG also in the bispectrum. In fact, redoing the analysis for Euclid up to $k_{\rm max}(z) = 0.1D(z)$, as used in [3], with the same specifications, except that we ignore the theoretical error, gives $\sigma(f_{NL})$ equal to 0.57, 0.71 and 1.3 (unmarginalized, including and neglecting priors, respectively). This is roughly a factor three improvement from the results in Table 1a. The fact that this still does not challenge the results from [3] seems to indicate that scale-dependent bias helps to improve the constraints on primordial non-Gaussianity.

From the combined catalogs, we ultimately find $\sigma(f_{NL}^{\rm loc}) = 1.8$, $\sigma(f_{NL}^{\rm eq}) = 11.4$ and $\sigma(f_{NL}^{\rm qsf}) = 8.9$, with priors on the EFT parameters, assuming the surveys are not independent. These results do not change dramatically if we do not put priors on the EFT parameters. If the surveys are independent, the constraints improve approximately with a factor $1/\sqrt{2}$ upon combining Euclid and DESI.

4.2 Correlation coefficients

To gain intuition on how much the EFT parameters affect the constraints on f_{NL} for local, equilateral and quasi-single field PNG, we compute the correlation coefficients between parameters θ_i and θ_j. These are defined as

$$r_{ij} = \frac{F_{ij}^{-1}}{\sqrt{F_{ii}^{-1} F_{jj}^{-1}}}.$$

The correlation coefficient takes a value between 1 (perfectly correlated) and -1 (perfectly anti-correlated). In particular, the parameters are perfectly correlated with themselves. In Figure 5, we plot the absolute value of the correlation coefficients for each pair of parameters. We make the following assumptions

$\sigma(f_{NL}^{\rm loc})$	unmarg.	with prior	no prior
BOSS	4.67	6.81	17.3
eBOSS	4.91	6.6	14.15
Euclid	1.41	1.77	3.66
DESI	1.66	2.18	4.68

(a) Local PNG

$\sigma(f_{NL}^{\rm eq})$	unmarg.	with prior	no prior
BOSS	16.89	29.86	37.99
eBOSS	17.25	26.88	33.4
Euclid	7.46	11.37	13.66
DESI	7.18	11.4	13.48

(b) Equilateral PNG

$\sigma(f_{NL}^{\rm qsf})$	unmarg.	with prior	no prior
BOSS	12.57	23.65	27.26
eBOSS	13.1	21.43	23.49
Euclid	5.52	8.92	9.74
DESI	5.37	8.98	9.66

(c) Quasi-single-field PNG

Table 1: The final $\sigma(f_{NL})$ for local, equilateral and quasi-single field PNG for each survey, combining all expected galaxy catalogues. For the marginalized $\sigma(f_{NL})$, we put a Gaussian prior on each EFT coefficient with $\sigma = 10$, except for the EFT parameter ξ, for which we take $\sigma = 1$. In the last row, we also show the marginalized results, without prior on the EFT coefficients.

- We use the same binning in k-space as in Section 4.1.1.
- We use the redshift binning and shotnoise from Euclid, as given in Section 4.1.2.
- As ansatz for the theoretical error we use $B^{\rm er} = B_{332}$.
- We do not marginalize over the EFT parameters. The marginalized results are quoted in the text below.

We find that the groups of parameters $\{\xi,\ \epsilon_1,\ \epsilon_2,\ \epsilon_3\}$ and $\{\gamma,\ \gamma_1,\ \gamma_2\}$ have strong correlation among themselves. The correlation between f_{NL} and the other parameters is, however, small.

For local PNG, we find f_{NL} is mainly correlated with γ, γ_1 and ϵ_3, with correlation coefficients 0.44, -0.43 and 0.29 respectively. The other correlation coefficients are in absolute value smaller than 0.2. If we include a Gaussian prior on the EFT parameters, with the same variances as before, the correlation coefficients become 0.12, 0.08 and 0.14.

In case of equilateral PNG, we find f_{NL} has appreciable correlation with ξ, ϵ_1, ϵ_3 and γ_2 ,with correlationcoefficients 0.43, -0.39, -0.47 and -0.27 respectively. Including priors on the EFT parameters, we find they become 0.14, -0.05, -0.30 and approximately zero. This could motivate further study on the Gaussian EFT coefficients, see for instance [38]. It is surprising that $f_{NL}^{\rm eq}$ is not extremely degenerate with ξ, since the latter comes with an additional k^2 scaling, similar to equilateral non-Gaussianity. It turns out, however, that the full shapes are sufficiently distinct. This will make it easier to constrain equilateral PNG from the bispectrum than naively thought. Then, for quasi-single-field PNG, we find that f_{NL} is mostly correlated with ϵ_3 and γ, with correlation coefficients -0.26 and 0.31. Including the priors, they reduce to -0.18 and approximately zero.

Summarizing, although the ignorance about EFT parameters does affect the final result, for reasonable priors this is only a small effect, especially for local PNG. This indeed agrees with what is seen in Figures 3 and 4, and Table 1a.

4.3 Higher loop corrections

We can ask ourselves whether it is useful to compute the bispectrum up to two loops in gravitational non-linearities. Note that our analysis does not depend on the actual value of the two-loop bispectrum, as there is no theoretical parameter in front. This means we can simply assume that we have computed all diagrams, neglect the counterterms, and assume the theoretical error is given by the SPT three loop bispectrum B_{3L}^G. Again, we do not know what it is, so we have to

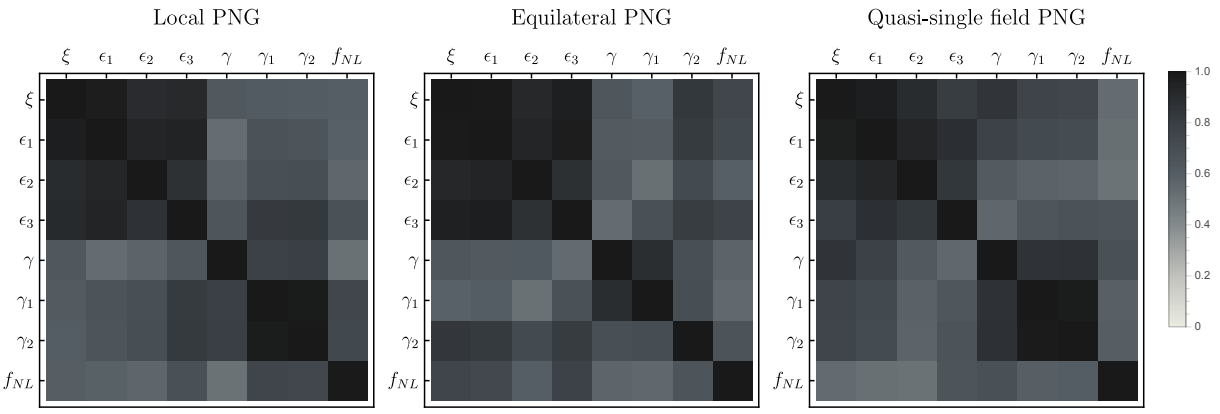


Figure 5: In these plots we show the correlation coefficients r_{ij} for each pair of theoretical parameters. We include shotnoise with the same specifications as Euclid and included all information up to redshift $z_{\max} = 2$. A value of 1 (black) corresponds to perfectly correlated or anti-correlated. A value of 0 (white) corresponds to no correlation.

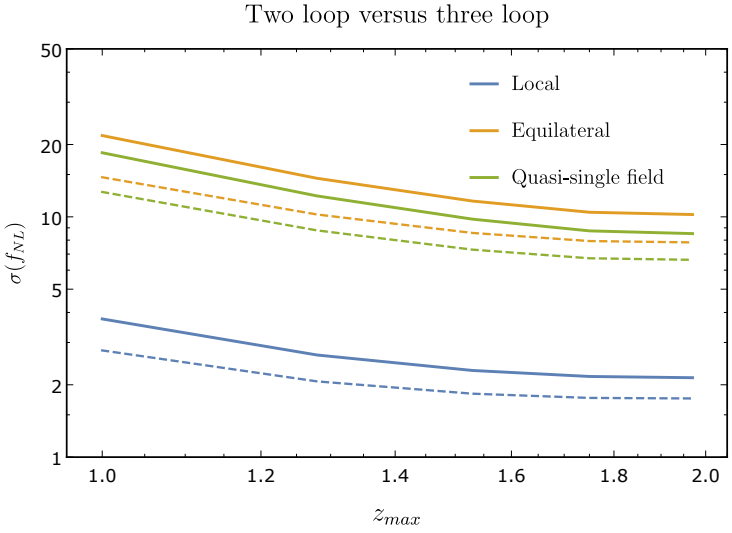


Figure 6: In this plot, we show $\sigma(f_{NL})$ as function of $z_{\max}$ for local (blue), equilateral (orange) and quasi-single field (green) PNG. We use both B_{2L} (solid lines) and B_{3L} (dashed lines) as order of magnitude estimation for the theoretical error. We use the specifications of Euclid in the analysis.

make an ansatz for it. Here we use the ansatz for the higher loop corrections from [15], given in (27), since it is easy to compute[11]. For a rough estimate this should suffice. We estimate B^G_{3L} using scaling universes [25]. We choose[12] E_b with $n = -1.5$, $k_{NL} = 0.5\ h\text{Mpc}^{-1}$ and $l = 3$ (see (27)). Using the specifications from Euclid, we perform the Fisher analysis with both B_{332} and B_{3L} as order of magnitude estimates for the noise. We collect the result in Figure 6. This shows that the constraints would improve if one computed the two-loop corrections to the bispectrum. The precise values are given in Table 2, where we consider all surveys again. The tighest constraints on local, equilateral and quasi-single field PNG improve with a factor 1.2, 1.3 and 1.3 respectively with this particular choice for B^G_{3L}. If it turns out we can get constraints on PNG close to the theoretical benchmarks, it would then be worth computing the two-loop corrections. It might be time consuming, but otherwise much cheaper than doubling the survey volume.

[11] An alternative - more in line with our two loop ansatz - would be to compute the reducible diagram of B_{433} as order of magnitude estimate of B^G_{3L}. However, we point out that using only one diagram is dangerous. For instance, in [14] we considered only one of the two reducible two loop diagrams in our qualitative analysis, namely B^I_{332}. In the squeezed limit, the two loop contribution turns out to be much larger if we include B^{II}_{332}.

[12] We choose a larger k_{NL} than for the two loop estimate, since this scale determines when the three loop correction equals the lower order corrections. Since we are doing a perturbative expansion, we assume this happens at a smaller scale than when the two loop correction becomes equal to its lower order corrections. Moreover, each loop will scale as k/k_{NL} to the power $3 + n$, where n will be of order of the scaling of the power spectrum at the non-linear scale k_{NL}. The power spectrum is steeper on smaller scales, therefore we take a more negative value for n. Each loop has this scaling, so we have to take $l = 3$, in case of three loops.

$\sigma(f_{NL}^{\rm loc})$	2 loop	3 loop
BOSS	8.73	6.05
eBOSS	7.12	6.07
Euclid	2.14	1.75
DESI	2.62	2.09

(a) Local PNG

$\sigma(f_{NL}^{\rm eq})$	2 loop	3 loop
BOSS	27.8	19.14
eBOSS	22.99	18.44
Euclid	10.22	7.83
DESI	10.2	7.81

(b) Equilateral PNG

$\sigma(f_{NL}^{\rm qsf})$	2 loop	3 loop
BOSS	23.66	16.65
eBOSS	19.15	15.59
Euclid	8.52	6.62
DESI	8.46	6.6

(c) Quasi-single-field PNG

Table 2: The final $\sigma(f_{NL})$ for equilateral and quasi-single-field PNG (left and right), for each survey, combining all expected galaxy catalogues. For the marginalized $\sigma(f_{NL})$, we put a Gaussian prior on each EFT coefficient with $\sigma = 10$, except for the EFT parameter ξ, for which we take $\sigma = 1$. In the last row, we also show the marginalized results without prior on the EFT coefficients.

Approach	$B^{\rm th} = B_0^G + \ldots$	$B^{\rm er} = B_{332} + \ldots$	$\sigma(f_{NL}^{\rm loc})$	$\sigma(f_{NL}^{\rm eq})$	$\sigma(f_{NL}^{\rm qsf})$
EFT (G+NG)	$+B_0^{NG} + B_{EFT}^G + B_{EFT}^{NG}$		1.77	11.37	8.92
EFT G+SPT NG	$+B_0^{NG} + B_{EFT}^G$	$B_{\rm 2L}^G + B_{EFT}^{NG}$	1.78	11.37	8.92
SPT (G+NG)	$+B_0^{NG}$	$+B_{EFT}^{NG} + B_{EFT}^G$	6.11	27.61	21.76
SPT (G+NG tree)	$+B_{\rm tree}^{NG}$	$+B_{EFT}^{NG} + B_{EFT}^G + B_{\rm 1L}^{NG}$	7.17	30.58	24.23

Table 3: We show the contraints on primordial non-Gaussianity of the local, equilateral and quasi-single-field type (last three columns). In the first row, we use the EFToLSS for both the Gaussian and non-Gaussian part of the bispectrum ('EFT (G+NG)'). In the second row, we only use the EFT for the Gaussian part of the bispectrum, and include the non-Gaussian counterterms in the higher order corrections ('EFT G + SPT NG'). Then, in the third row, we use only the SPT for describing the bispectrum ('SPT (G+NG)'). In the last row, we only include the tree level non-Gaussian contribution to the bispectrum ('SPT (G+NG tree)'). The second and third column denote all the contributions to the theoretical description of the bispectrum $B^{\rm th}$, and the higher order corrections $B^{\rm er}$ respectively.

4.4 EFT of LSS versus SPT

In the EFT of LSS we are forced to include free parameters over which we have to marginalize. Above, we saw that this marginalization weakens the constraints on $f_{\rm NL}$, be it only mildly. One might therefore wonder how much the improvement actually is over a more conservative approach, in which one uses only SPT results for $B^{\rm th}$ and moves all other gravitational contributions to the theoretical error. In this section we confirm that the EFT approach always performs sizably better. We consider a couple of options, for different choices of $B^{\rm th}$ and $B^{\rm er}$, and compute the constraints. We use the specifications from Euclid as given in the previous section.

Table 3 shows our results, which include the usual Gaussian priors for the EFT parameters whenever they are included in $B^{\rm th}$. The second and third columns give the theoretical description of the bispectrum $B^{\rm th}$ and what we consider to be the unknown $B^{\rm er}$. For the latter, we sum the absolute values of all contributions indicated in the table. The EFT contributions, except for ξ, are multiplied by a factor 10, consistent with the priors we chose when we included them in $B^{\rm th}$. The first row of Table 3 corresponds to the results we find in the 'with prior' columns of the 'Euclid' rows of Tables 1a, 1b and 1c.

We find that including the non-Gaussian counterterms does not improve the bounds on f_{NL}. This could have been anticipated from the qualitative results in Figure 5 - 7 of [14]. We see that the counterterms are negligible in many configurations. Apparently, they are negligible in most configurations. However, using the EFT for the Gaussian part of the bispectrum performs significantly improves the results compared with just the SPT predictions. We find that the *EFT of LSS improves the constraints on PNG approximately by a factor 3.* Finally, neglecting the one-loop non-Gaussian contribution to the bispectrum makes only about a 10% difference. This is consistent with the observation that the non-Gaussian counterterms are not very important, as the non-Gaussian one loop correction itself is not very relevant.

5 Discussion and Outlook

In this work, we have presented how the EFT of LSS helps us improve the constraints on primordial non-Gaussianities (PNG), using the matter bispectrum as observable. We have accounted for intrinsic theoretical uncertainties in the perturbative description, and studied in details their modeling in a Fisher forecast.

Our main results are given in Table 3. The forecasted values for $\sigma(f_{NL})$ for the local, equilateral and quasi-single field types of PNG are presented. Moreover, we show that the EFT approach improves the constraints on PNG by almost a factor 3 with respect to the results from SPT.

Limitations Let us first discuss the limitations of these results. We would like to compare the constraints we find with theoretically interesting benchmarks and constraints coming from the CMB. However, we should be careful in making a direct comparison, as there are other sources of non-linearities and noise that we have not accounted for in our analysis. First, we have modeled the matter bispectrum. To relate it to the observed galaxy bispectrum, we have to include galaxy bias and redshift space distortions. These introduce new uncertainties, leading to worse constraints. However, considering the results found in [2, 3], scale-dependent bias might actually *improve* the constraints on (only) local PNG, since it enhances the non-Gaussian signal in the bispectrum. Second, except for shotnoise, we neglected all observational sources of noise. Survey geometry and survey mask may increase the errorbars as well. For instance the authors of [6] found that the errorbars increased by a factor of 4-5. Errors in determining the redshift of galaxies are another source of observational noise. Third, we made some simplifications in the Fisher analysis itself, such as neglecting the covariance between different points of the bispectrum. Combining this with the covariance induced by the survey geometry could further increase the errorbars by a factor of 8 [39].

Improvements On the other hand, there are also ways the constraints could be improved. First, we have used the specifications of Euclid to get a reasonable estimate for the limitations due to shotnoise. This determined our final forecasted result for $\sigma(f_{NL})$. It might well be possible, in a more futuristic survey, to optimize the number densities of galaxies and redshift range to be more suitable for constraining PNG (see for instance [2]). Moreover, we should perform a joint analysis of all large scale structure surveys. We have assumed for simplicity that we can do as well as the single best survey, which turned out to be Euclid for the four surveys we considered. In principle, we can do better if the surveys are not all precisely overlapping. Similarly, we should combine the results from different observables. For instance, we should perform a joint analysis of the power spectrum and bispectrum. This could improve the results by a factor of 2 for local PNG. Combining the results found in [3], and using the multitracer technique proposed in [40] instead, which could improve upon the constraints from the power spectrum by a factor of about 7. In addition, the trispectrum might turn out to be an important probe for non-Gaussianity, since linear theory works for a larger range of scales compared to the bispectrum [41]. The one loop corrections to the trispectrum in the EFT of LSS have recently been computed in [42]. Last, we divided the full redshift range in smaller redshift bins, and only considered correlations within each redshift bin. If we also include correlations among galaxies separated by a larger distance along the line of sight, we might extract more information from a given survey. We will discuss the issue elsewhere [31]. Finally, our focus here was on near future galaxy surveys, but of course our results will be relevant in the future also for 21 cm survey (see e.g. [43]).

One of our main results is that the EFT approach helps constraining PNG. The improvement comes completely from the EFT corrections to the late time gravitational non-linear evolution of matter. Both the SPT loops and EFT corrections to the primordial non-Gaussian signal, discussed in [14], do not help much improving the constraints.

Comparing our results with the theoretically interesting benchmark $\sigma(f_{NL}^{\rm eq}) \sim 1$, we see that it does not look promising for equilateral PNG. Even with zero shotnoise, as we can see in Figure

3, we barely touch the theoretical targets. Our lack of understanding of matter non-linearities is already an important obstacle to reach $\sigma(f_{NL}^{\rm eq}) \sim 1$. The same applies to quasi-single field PNG. Additional sources of non-linearities such as bias and redshift space distortions will make things worse. On the other hand, for local primordial non-Gaussianity, things look more promising. Matter non-linearities can be modeled well enough to get close to $\sigma(f_{NL}^{\rm loc}) \sim 1$ from large scale structure experiments.

We can ask whether N-body simulations can help reaching better constraints on primordial non-Gaussianity. As pointed out in [15], using end to end simulations, without any perturbative input, will most likely be insufficient to reach $\sigma(f_{NL}^{\rm eq}) \sim 1$. The reason is that simulations do not solve the exact problem but make a series of approximations, such as for example the particle mesh and tree approaches to solve Poisson equation, finite size effects and approximate initial conditions. Currently, simulations reach approximately 1% precision [44,45]. Heuristically, looking at our Figure 1, we see that the PNG signal we are trying to extract is much smaller than that, so large improvements in the precision of simulations are needed. Alternatively, one can use N-body simulations to determine the EFT parameters[13]. We can then look directly at the unmarginalized columns in Table 1c. We see that, even in the very optimistic case that all relevant EFT parameters at one loop are fixed, $\sigma(f_{NL}^{\rm eq})$ still remains around 7.

Theoretical error Another goal of our paper is to clarify some aspects of the modeling of theoretical uncertainties in forecasting observational constraints, and, eventually, in analyzing data (see Section 3.5 and Appendix B). We introduced the concept of correlation length in Section 3.4, along the lines of [15]. In Appendix B, we argued that the choice of correlation length in integrating out the theoretical error is subtle and no "right" choice can be established a priori. However, in our particular analysis, we hardly find any dependence on the correlation length (see Figure 7). In future studies, with different observables and different perturbative approaches, we believe that an analysis on the choice of correlation length should be always performed.

In Figure 2 we have seen that assuming the wrong shape for the theoretical error can lead to biased results in a χ^2-analysis. Therefore, if we want to fit to data, we need good estimates for the higher order corrections. For instance by using estimates from N-body simulations, or alternatively, by computing additional two-loop diagrams.

Outlook Our work can be extended and improved in different ways.

- Instead of dividing the survey volume in redshift bins and only consider correlations within these bins, it would be interesting to see how much we gain including all possible cross-correlations across redshift bins. This is work in progress [31].
- It would be interesting to perform a similar Fisher analysis with an updated study of covariance effects due to geometry, masking and non-Gaussian gravitational evolution.
- We should join all forces. It would be interesting to do a joint analysis of multiple observations, such as the CMB, LSS surveys and the 21 cm observations. Moreover, all the different LSS surveys should be combined to have maximum constraining power. Furthermore, the results from the power spectrum, bispectrum and trispectrum should be combined too. Finally, on the theory side, one should also try to use results from N-body simulations as soon as our perturbative description starts to break down, i.e. when the theoretical error becomes dominant.

Acknowlegdements

We thank Valentin Assassi and Daniel Baumann for initial collaboration on this project. We would like to thank Emiliano Sefusatti for useful discussions. We are especially thankful to Tobias

[13] In fact, in our analysis, we assumed that a one EFT parameter, ξ, was fixed by fitting the power spectrum to simulations.

Baldauf for the many useful discussions. We would like to thank Fabian Schmidt and Giovanni Cabass for feedback on the draft. Y.W. would like to thank Koen Kuijken and Ana Achúcarro for providing a sounding board now and then. Y.W. is supported by a de Sitter Fellowship of the Netherlands Organization for Scienti c Research (NWO). E.P. is supported by the Delta-ITP consortium, a program of the Netherlands organization for scientic research (NWO) that is funded by the Dutch Ministry of Education, Culture and Science (OCW).

A Explicit Results for the Bispectrum

The purpose of this appendix is to provide, and to some extent clarify, explicit expressions for the bispectrum in the presence of primordial non-Gaussianity. This is essentially a summary of [14], so we refer the reader to that work for a thorough discussion and explanation of these results. We adopt the same notation, which we summarize in Appendix D.

A.1 Perturbation Theory in the EFT of LSS

In the EFToLSS, the equations of motion for the density contrast δ and the velocity divergence $\theta = \partial_i v^i$ on large scales are

$$\delta_\tau \delta + \theta = \mathcal{S}_\alpha \ , \tag{34a}$$

$$(\delta_\tau + \mathcal{H})\theta + \frac{3}{2}\Omega_m \mathcal{H}^2 \delta = \mathcal{S}_\beta + \tau_\theta \ . \tag{34b}$$

Here the source terms $\mathcal{S}_{\alpha,\beta}$ are the standard nonlinear terms in the Euler equations, which, in Fourier space, are given by the following convolutions,

$$\mathcal{S}_\alpha(\mathbf{k},\tau) \equiv -\int_{\mathbf{p}} \alpha(\mathbf{p},\mathbf{k}-\mathbf{p})\theta(\mathbf{p},\tau)\delta(\mathbf{k}-\mathbf{p},\tau) \ , \qquad \alpha(\mathbf{k}_1,\mathbf{k}_2) \equiv \frac{\mathbf{k}_1\cdot(\mathbf{k}_1+\mathbf{k}_2)}{\mathbf{k}_1^2} \ , \tag{35a}$$

$$\mathcal{S}_\beta(\mathbf{k},\tau) \equiv -\int_{\mathbf{p}} \beta(\mathbf{p},\mathbf{k}-\mathbf{p})\theta(\mathbf{p},\tau)\theta(\mathbf{k}-\mathbf{p},\tau) \ , \qquad \beta(\mathbf{k}_1,\mathbf{k}_2) \equiv \frac{(\mathbf{k}_1+\mathbf{k}_2)^2}{2}\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{\mathbf{k}_1^2\mathbf{k}_2^2} \ . \tag{35b}$$

Clearly, we neglected large scale vorticity and large scale velocity dispersion in (34). However, the backreaction from unknown short scale physics is taken into account through the effective stress tensor τ_θ. A complete description and motivation of this term in the presence of primordial non-Gaussianity was the main purpose of [14]. Here we just quote the leading contributions for the types of non-Gaussianities we consider, to first order in $f_{\rm NL}$,

$$\begin{aligned}\tau_\theta = &-d^2\,\triangle\delta - e_1\,\triangle(\delta^2) - e_2\,\triangle(s^2) - e_3\,\partial_i(s^{ij}\partial_j\delta) \\ &- f_{\rm NL}\big[g\big(\triangle\Psi - \partial_i(\delta\partial^i\Psi)\big) + g_1\,\triangle(\Psi\delta) + g_2\partial_i\partial_j(\Psi s^{ij})\big] \ ,\end{aligned} \tag{36}$$

where $\triangle$ denotes the Laplacian, and the coefficients in this expression are functions of time only. The equations are formally solved using a Green's function method

$$\delta(\mathbf{k},a) = D_1(a)\delta_1(\mathbf{k}) + \int_{a_{in}}^{a} G_\delta(a,a')\left[\mathcal{S}_\beta + \tau_\theta - \mathcal{H}\partial_a(a\mathcal{S}_\alpha)\right]. \tag{37}$$

Here $\delta_1(\mathbf{k})$ is the growing mode initial condition, and $D_1(a)$ is the linear growth factor

$$D_1(a) = \frac{5}{2}\mathcal{H}_0^2\Omega_m^0\frac{\mathcal{H}}{a}\int_{a_{in}}^{a}\frac{\mathrm{d}a'}{\mathcal{H}^3(a')} \ , \tag{38}$$

which in Einstein-de Sitter reduces to $D_1(a) = a/a_{in}$, where a is the scale factor. This equation can be solved perturbatively in terms of the linear solution, yielding

$$\delta(\mathbf{k},a) = \sum_{n=1}^{\infty}\delta_{(n)}^{\rm SPT}(\mathbf{k},a) + \delta_{(n)}^{\rm c}(\mathbf{k},a), \tag{39}$$

where $\delta^{\rm SPT}_{(n)}$ are the standard perturbation theory (SPT) terms (see [5]), sourced solely by $\mathcal{S}_{\alpha,\beta}$:

$$\delta^{\rm SPT}_{(n)}(\mathbf{k},a) \;\approx\; D_1^n(a)\int_{\mathbf{k}_1}\dots\int_{\mathbf{k}_n}(2\pi)^3\delta_D\big(\mathbf{k}-\mathbf{k}_{1\dots n}\big)\,F_n(\mathbf{k}_1,\dots,\mathbf{k}_n)\,\delta_1(\mathbf{k}_1)\dots\delta_1(\mathbf{k}_n); \tag{40}$$

and $\delta^{\rm c}_{(n)}$ is the 'counterterm' contribution, i.e. the terms proportional to one of the free parameters in (36), which we write as

$$\begin{aligned}\delta^{c}_{(n)}(\mathbf{k},a) = &\int_{\mathbf{k}_1}\dots\int_{\mathbf{k}_n}(2\pi)^3\delta_D\big(\mathbf{k}-\mathbf{k}_{1\dots n}\big)\,F^c_n(\mathbf{k}_1,\dots,\mathbf{k}_n|a)\,\delta_{(1)}(\mathbf{k}_1,a)\dots\delta_{(1)}(\mathbf{k}_n,a)\\ &+f_{\rm NL}\int_{\mathbf{k}_1}\dots\int_{\mathbf{k}_n}(2\pi)^3\delta_D\big(\mathbf{k}-\mathbf{k}_{1\dots n}\big)\,H^c_n(\mathbf{k}_1,\dots,\mathbf{k}_n|a)\,\psi(\mathbf{k}_1)\dots\delta_{(1)}(\mathbf{k}_n,a)\;.\end{aligned} \tag{41}$$

Note that we have not specified the time dependence of the free parameters yet, which is why the counterterm kernels are still time dependent. It turns out that at first order in the perturbations, this is not really an issue, as the time dependence is just given by an integral of the Green's function over some unknown function of time, which yields some other unknown function. However, in order to get the momentum dependence right at second order, we have to make an assumption about the first order terms. A convenient ansatz is

$$d^2(a) = [\mathcal{H}(a)f(a)]^2[D_1(a)]^{m_d+1}\,\bar{d}^2\;, \tag{42}$$

$$g(a) = [\mathcal{H}(a)f(a)]^2[D_1(a)]^{m_g+1}\,\bar{g}\;, \tag{43}$$

which in Einstein-de Sitter reduces to $(d^2,g)\propto a^{m_d}$. Then the expressions for the counterterm kernels at second order are as follows. We split up the kernels in the following way

$$F^c_2(\mathbf{k}_1,\mathbf{k}_2|a) = F^\tau_2(\mathbf{k}_1,\mathbf{k}_2|a) + F^{\alpha\beta}_2(\mathbf{k}_1,\mathbf{k}_2|a) + F^\delta_2(\mathbf{k}_1,\mathbf{k}_2|a), \tag{44}$$

where the terms coming from the new counterterms at second order are

$$F^\tau_2(\mathbf{k}_1,\mathbf{k}_2|a) = -\sum_{i=1}^{3}\epsilon_i(a)E_i(\mathbf{k}_1,\mathbf{k}_2),\;\text{with} \tag{45}$$

$$E_1(\mathbf{k}_1,\mathbf{k}_2) \equiv \mathbf{k}^2_{12}, \tag{46}$$

$$E_2(\mathbf{k}_1,\mathbf{k}_2) \equiv \mathbf{k}^2_{12}\left[\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{k_1^2k_2^2}-\frac{1}{3}\right], \tag{47}$$

$$E_3(\mathbf{k}_1,\mathbf{k}_2) \equiv \left[-\frac{1}{6}\mathbf{k}^2_{12}+\frac{1}{2}\mathbf{k}_1\cdot\mathbf{k}_2\left[\frac{\mathbf{k}_{12}\cdot\mathbf{k}_2}{k_2^2}+\frac{\mathbf{k}_{12}\cdot\mathbf{k}_1}{k_1^2}\right]\right],\;\text{and} \tag{48}$$

$$\epsilon_i(a) \equiv -\frac{1}{[D_1(a)]^2}\int_{a_{in}}^{a}\mathrm{d}a'\;G_\delta(a,a')\,[D_1(a')]^2\,e_i(a')\;. \tag{49}$$

Furthermore, the terms coming from plugging the lowest order counterterm back into the equations of motion are

$$F^{\alpha\beta}_2(\mathbf{k}_1,\mathbf{k}_2|a) = -\xi(a)E_{\alpha\beta}(\mathbf{k}_1,\mathbf{k}_2),\;\text{with} \tag{50}$$

$$E_{\alpha\beta}(\mathbf{k}_1,\mathbf{k}_2) \equiv \frac{1}{2m_d+9}\Bigg[2\beta(\mathbf{k}_1,\mathbf{k}_2)(k_1^2+k_2^2) \tag{51}$$

$$+\frac{2m_d+7}{2(m_d+2)}\Big(\alpha(\mathbf{k}_1,\mathbf{k}_2)\big(k_2^2+(m_d+2)k_1^2\big)+\{1\leftrightarrow 2\}\Big)\Bigg],\;\text{and} \tag{52}$$

$$\xi(a) = \frac{2}{(m_d+1)(2m_d+7)}[D_1(a)]^{m_d+1}\,\bar{d}^2\;. \tag{53}$$

Finally, the term from the second order contribution to the density is

$$F^\delta_2(\mathbf{k}_1,\mathbf{k}_2|a) = -\xi(a)E_\delta(\mathbf{k}_1,\mathbf{k}_2),\;\text{with} \tag{54}$$

$$E_\delta(\mathbf{k}_1,\mathbf{k}_2) = \frac{(m_d+1)(2m_d+7)}{(m_d+2)(2m_d+9)}\,\mathbf{k}^2_{12}\,F_2(\mathbf{k}_1,\mathbf{k}_2)\;. \tag{55}$$

Similarly, for the non-Gaussian kernels we have

$$H_2^c(\mathbf{k}_1,\mathbf{k}_2|a) = H_2^\tau(\mathbf{k}_1,\mathbf{k}_2|a) + H_2^{\alpha\beta}(\mathbf{k}_1,\mathbf{k}_2|a) + H_2^\Psi(\mathbf{k}_1,\mathbf{k}_2|a)\ , \tag{56}$$

where second order counterterm contribution is

$$H_2^\tau(\mathbf{k}_1,\mathbf{k}_2|a) = -\sum_{i=1}^{2}\gamma_i(a)G_i(\mathbf{k}_1,\mathbf{k}_2),\ \text{with} \tag{57}$$

$$G_1(\mathbf{k}_1,\mathbf{k}_2) = \mathbf{k}_{12}^2, \tag{58}$$

$$G_2(\mathbf{k}_1,\mathbf{k}_2) = \frac{(\mathbf{k}_{12}\cdot\mathbf{k}_2)^2}{k_2^2} - \frac{1}{3}\mathbf{k}_{12}^2,\ \text{and} \tag{59}$$

$$\gamma_i(a) \equiv -\frac{1}{D_1(a)}\int_{a_{in}}^{a} \mathrm{d}a'\ G_\delta(a,a')\,D_1(a')g_i(a'). \tag{60}$$

Subsequently, the convoluted first order counterterms give

$$H_2^{\alpha\beta}(\mathbf{k}_1,\mathbf{k}_2|a) = -\gamma(a)G_{\alpha\beta}(\mathbf{k}_1,\mathbf{k}_2),\ \text{with} \tag{61}$$

$$G_{\alpha\beta}(\mathbf{k}_1,\mathbf{k}_2) \equiv \frac{4}{2m_g+7}\,\beta(\mathbf{k}_1,\mathbf{k}_2)k_1^2 \tag{62}$$

$$+\frac{2m_g+5}{(m_g+1)(2m_g+7)}\big[(m_g+1)\alpha(\mathbf{k}_1,\mathbf{k}_2)+\alpha(\mathbf{k}_2,\mathbf{k}_1)\big]k_1^2,\ \text{and} \tag{63}$$

$$\gamma(a) = \frac{2}{m_g(2m_g+5)}[D_1(a)]^{m_g+1}\,\bar{g}\ . \tag{64}$$

Lastly, the second order contribution to the lowest order counterterm is given by

$$H_2^\Psi(\mathbf{k}_1,\mathbf{k}_2|a) = -\gamma(a)G_\Psi(\mathbf{k}_1,\mathbf{k}_2),\ \text{with} \tag{65}$$

$$G_\Psi(\mathbf{k}_1,\mathbf{k}_2) = \frac{m_g(2m_g+5)}{(m_g+1)(2m_g+7)}\left[\mathbf{k}_{12}^2\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_2^2} - \mathbf{k}_{12}\cdot\mathbf{k}_1\right]\ . \tag{66}$$

A.2 One Loop Bispectrum

Having obtained perturbative solutions for the evolved density contrast in terms of the initial field, we can compute correlation functions using the statistical properties of the initial distribution. Along the lines of the discussion above, we decompose the bispectrum as

$$B_{\rm tot} = B^{\rm G}_{\rm SPT} + B^{\rm G}_{\rm EFT} + f_{\rm NL}\left(B^{\rm NG}_{\rm SPT} + B^{\rm G}_{\rm EFT}\right). \tag{67}$$

The expressions for the Gaussian part of the bispectrum at one loop were given in [12] and [13], and read

$$B^{\rm G}_{\rm SPT} = B_{112} + \Big[B_{114} + B^{\rm (I)}_{123} + B^{\rm (II)}_{123} + B_{222}\Big], \tag{68}$$

$$B^{\rm G}_{\rm EFT} = \xi B^{\rm G}_{\xi}\ + \sum_{i=1}^{3}\epsilon_i B_{\epsilon_i}, \tag{69}$$

where

$$B_{112} = 2F_2(\mathbf{k}_1, \mathbf{k}_2)P_{11}(k_1)P_{11}(k_2) + 2 \text{ cycl. perms} \tag{70a}$$

$$B_{222} = 8\int_{\mathbf{p}} F_2(-\mathbf{p}, \mathbf{p}+\mathbf{k}_1)F_2(\mathbf{p}+\mathbf{k}_1, -\mathbf{p}+\mathbf{k}_2)F_2(\mathbf{k}_2+-\mathbf{p}, \mathbf{p}) \tag{70b}$$
$$P_{11}(p)P_{11}(|\mathbf{p}+\mathbf{k}_1|)P_{11}(|\mathbf{p}-\mathbf{k}_2|),$$

$$B^{(\mathrm{I})}_{321} = 6P_{11}(k_3)\int_{\mathbf{p}} F_3(-\mathbf{p}, \mathbf{p}-\mathbf{k}_2, -\mathbf{k}_3)F_2(\mathbf{p}, \mathbf{k}_2-\mathbf{p})P_{11}(p)P_{11}(|\mathbf{p}-\mathbf{k}_2|) + 5 \text{ perms}, \tag{70c}$$

$$B^{(\mathrm{II})}_{321} = 6F_2(\mathbf{k}_2, \mathbf{k}_3)P_{11}(k_2)P_{11}(k_3)\int_{\mathbf{p}} F_3(-\mathbf{k}_3, \mathbf{p}, -\mathbf{p})P_{11}(p) + 5 \text{ perms}, \tag{70d}$$

$$B_{411} = 12P_{11}(k_2)P_{11}(k_3)\int_{\mathbf{p}} F_4(\mathbf{p}, -\mathbf{p}, -\mathbf{k}_2, -\mathbf{k}_3)P_{11}(p) + 2 \text{ cycl. perms}. \tag{70e}$$

Here $P_{11}(k)$ is the linear power spectrum, whose time dependence is implied. Furthermore

$$B^{\mathrm{G}}_{\xi} \equiv -2\big[E_{\alpha\beta}(\mathbf{k}_1, \mathbf{k}_2) + E_{\delta}(\mathbf{k}_1, \mathbf{k}_2)\big]P_{11}(k_1)P_{11}(k_2) + 2 \text{ perms}, \tag{71a}$$

$$B_{\epsilon_i} \equiv -2\,E_i(\mathbf{k}_1, \mathbf{k}_2)\,P_{11}(k_1)P_{11}(k_2) + 2 \text{ perms}. \tag{71b}$$

The non-Gaussian contribution at one loop is

$$B^{\mathrm{NG}}_{\mathrm{SPT}} = B_{111} + \Big[B^{(\mathrm{I})}_{113} + B^{(\mathrm{II})}_{113} + B^{(\mathrm{I})}_{122} + B^{(\mathrm{II})}_{122}\Big], \tag{72}$$

$$B^{\mathrm{NG}}_{\mathrm{EFT}} = \xi B^{\mathrm{NG}}_{\xi} + \gamma B_{\gamma} + \sum_{i=1}^{2}\gamma_i B_{\gamma_i}, \tag{73}$$

where

$$B^{(\mathrm{I})}_{113} = 3P_{11}(k_2)\int_{\mathbf{p}} F_3(\mathbf{k}_1+\mathbf{p}, -\mathbf{p}, \mathbf{k}_2)B_{111}(k_1, p, |\mathbf{k}_1+\mathbf{p}|) + 5 \text{ perms}, \tag{74a}$$

$$B^{(\mathrm{II})}_{113} = 3B_{111}(k_1, k_2, k_3)\int_{\mathbf{p}} F_3(\mathbf{k}_1, \mathbf{p}, -\mathbf{p})P_{11}(p) + 2 \text{ perms}, \tag{74b}$$

$$B^{(\mathrm{I})}_{122} = 4\int_{\mathbf{p}} F_2(\mathbf{k}_3+\mathbf{p}, -\mathbf{p})F_2(\mathbf{p}, \mathbf{k}_2-\mathbf{p})B_{111}(k_1, |\mathbf{k}_3+\mathbf{p}|, |\mathbf{k}_2-\mathbf{p}|)P_{11}(p) + 2 \text{ perms}, \tag{74c}$$

$$B^{(\mathrm{II})}_{122} = 2F_2(\mathbf{k}_1, \mathbf{k}_2)P_{11}(k_2)\int_{\mathbf{p}} F_2(\mathbf{p}, \mathbf{k}_1-\mathbf{p})B_{111}(k_1, p, |\mathbf{k}_1-\mathbf{p}|) + 5 \text{ perms}, \tag{74d}$$

and

$$B^{\mathrm{NG}}_{\xi} \equiv -(k_1^2+k_2^2+k_3^2)B_{111}(k_1, k_2, k_3), \tag{75a}$$

$$B_{\gamma} \equiv -\big[G_{\alpha\beta}(\mathbf{k}_1, \mathbf{k}_2) + G_{\Psi}(\mathbf{k}_1, \mathbf{k}_2)\big]P_{11}(k_1)P_{1\psi}(k_2) + 5 \text{ perms}, \tag{75b}$$

$$B_{\gamma_i} \equiv -G_i(\mathbf{k}_1, \mathbf{k}_2)P_{11}(k_1)P_{1\psi}(k_2) + 5 \text{ perms}. \tag{75c}$$

Again, the time dependence of the correlation functions is implied. Finally, we have to specify the type of non-Gaussianity. In this work we consider the primordial bispectra given in 2.2. The corresponding correlation between the linear density field and the first order non-Gaussian counterterm ψ (see (41)) is given by

$$P_{1\psi}(k) \equiv \frac{(k/\mu)^{\Delta}}{M(k)}P_{11}(k), \tag{76}$$

where μ is some arbitrary momentum scale, introduced for dimensional reasons, which cancels when multiplied with the EFT parameter in the full contribution to the bispectrum. We therefore

set it to unity in the numerical evaluation. In our case the k-dependence is respectively given by $\Delta = \{0, 1, 2\}$ for local, quasi-single field and equilateral type non-Gaussianities. The transfer function was defined in (7). For this work we choose the time dependence of the lowest order counterterms (42) to match the divergence it is supposed to cancel, which corresponds to the choice $m_d = m_g = 1$. This is was argued for in [27]. Moreover, our main results do not depend much on this assumption.

A.3 Ansatz Two Loop Bispectrum

As an ansatz for the two loop bispectrum we compute the two reducible two loop diagrams, given by [12],

$$B^{\mathrm{I}}_{332} = 2F_2(\mathbf{k}_1, \mathbf{k}_2)\frac{P_{13}(k_1)}{2}\frac{P_{13}(k_2)}{2} + 2 \text{ cycl. perms} \tag{77a}$$

$$B^{\mathrm{II}}_{332} = 6\frac{P_{13}(k_3)}{2}\int_{\mathbf{p}} F_3(-\mathbf{p}, \mathbf{p}-\mathbf{k}_2, -\mathbf{k}_3)F_2(\mathbf{p}, \mathbf{k}_2-\mathbf{p})P_{11}(p)P_{11}(|\mathbf{p}-\mathbf{k}_2|) + 5 \text{ perms}, \tag{77b}$$

with

$$P_{13}(k) = 6P_{11}(k)\int_{\mathbf{p}} F_3(\mathbf{k}, \mathbf{p}, -\mathbf{p})P_{11}(p). \tag{78}$$

As an estimate for the theoretical error we use

$$B_{332} = |B^{\mathrm{I}}_{332}| + |B^{\mathrm{II}}_{332}|. \tag{79}$$

B Theoretical noise

This appendix contains the details of the implementation of the theoretical error and further investigates some issues related to it. First, we give the intermediate steps to derive (25) and provide an alternative derivation of the effective Fisher matrix in the presence of theoretical error. Then, we study the effect of the correlation length, both by means of a toy model and by running the analysis for several correlations lengths. Finally, we discuss in more detail the effect of the two possible ansätze for B^{er}.

B.1 Derivation of (25)

Let us first show how to go from equations (22), (23) and (24) to the effective Fisher matrix given in (25). We will need to use the Woodbury matrix identity several times, which relates the inverse of sums of matrices to their individual inverses

$$(A+B)^{-1} = A^{-1} - A^{-1}\left(A^{-1}+B^{-1}\right)^{-1}A^{-1}. \tag{80}$$

By using this identity, we can rewrite (23) as

$$F^{-1}_{\alpha\beta} = (N^{-1}_{\alpha\beta} + D_{\alpha\beta})^{-1} = D^{-1}_{\alpha\beta} - D^{-1}_{\alpha\gamma}\left(N + D^{-1}\right)^{-1}_{\gamma\delta}D^{-1}_{\delta\beta}. \tag{81}$$

This allows us to compute

$$\begin{aligned}
F_{i\gamma}F^{-1}_{\gamma\delta}F_{\delta j} &= \sum_{\hbar,p} B_i(\hbar)C^{-1}(\hbar,\hbar_\gamma)B^{\mathrm{er}}(k_\gamma)\left(D^{-1} - D^{-1}(N+D^{-1})^{-1}D^{-1}\right)_{\gamma\delta}B^{\mathrm{er}}(k_\delta)C^{-1}(\hbar_\delta, p)B_j(p) \\
&= \sum_{\hbar,p} B_i(\hbar)\delta_{\hbar,\hbar_\alpha}\left(C^{-1}(\hbar_\alpha,\hbar_\beta) - \frac{1}{B^{\mathrm{er}}(k_\alpha)}(N+D^{-1})^{-1}_{\alpha\beta}\frac{1}{B^{\mathrm{er}}(k_\beta)}\right)\delta_{\hbar_\beta,p}B_j(p) \\
&= \sum_{\hbar_\alpha,\hbar_\beta}\frac{B_i(\hbar_\alpha)}{B^{\mathrm{er}}(k_\alpha)}\left(D - (N+D^{-1})^{-1}\right)_{\alpha\beta}\frac{B_i(\hbar_\beta)}{B^{\mathrm{er}}(k_\beta)} \\
&= \sum_{\hbar_\alpha,\hbar_\beta}\frac{B_i(\hbar_\alpha)}{B_\alpha}\left(D(N^{-1}+D)^{-1}D\right)_{\alpha\beta}\frac{B_i(\hbar_\beta)}{B_\beta}.
\end{aligned}$$

Here summation over the Greek indices is understood. Upon applying the Woodbury identity again, the effective Fisher matrix then becomes

$$\begin{aligned} F_{ij}^{\text{eff}} &= \sum_{\hbar_\alpha,\hbar_\beta} \frac{B_i(\hbar_\alpha)}{B_\alpha} \left(D - D(N^{-1}+D)^{-1}D\right)_{\alpha\beta} \frac{B_j(\hbar_\beta)}{B_\beta} + \left(C_\Theta^{-1}\right)_{ij} \\ &= \sum_{\hbar_\alpha,\hbar_\beta} \frac{B_i(\hbar_\alpha)}{B_\alpha} \left(N + D^{-1}\right)^{-1}_{\alpha\beta} \frac{B_j(\hbar_\beta)}{B_\beta} + \left(C_\Theta^{-1}\right)_{ij} \\ &= \sum_{\hbar_\alpha,\hbar_\beta} B_i(\hbar_\alpha) \left(N^{\text{eff}} + C_B\right)^{-1}_{\alpha\beta} B_j(\hbar_\beta) + \left(C_\Theta^{-1}\right)_{ij}, \end{aligned} \tag{82}$$

with $N^{\text{eff}}_{\alpha\beta} = B^{\text{er}}(k_\alpha) N_{\alpha\beta} B^{\text{er}}(k_\beta)$.

B.2 Alternative derivation of the effective Fisher matrix

Next, we present a slightly different derivation of the effective Fisher matrix, by marginalizing at the level of the likelihood function. Let us first expand $\chi^2 = -2\log(\mathcal{L})$ in the nuisance parameters Θ_α. We would like to expand about some value $\bar{\Theta}_\alpha$ to get

$$\chi^2(\Theta_i,\Theta_\alpha) = \chi^2(\Theta_i,\bar{\Theta}_\alpha) + (\Theta_\alpha - \bar{\Theta}_\alpha)\chi^2_\alpha(\Theta_i,\bar{\Theta}_\alpha) + \frac{1}{2}(\Theta_\alpha - \bar{\Theta}_\alpha)(\Theta_\beta - \bar{\Theta}_\beta)\chi^2_{\alpha\beta}(\Theta_i,\bar{\Theta}_\alpha), \tag{83}$$

where summation over repeating indices is understood, and the index α on χ^2 denotes a derivative with respect to the corresponding nuisance parameter. It is an equality, since the variables are Gaussian distributed. We can rewrite this expression in more compact notation as

$$\chi^2 = \chi^2_0 + \delta\Theta^\alpha X_\alpha + \frac{1}{2}\delta\Theta^\alpha\delta\Theta^\beta Y_{\alpha\beta}, \tag{84}$$

where χ^2_0 is the chi-squared we would get if we ignored the presence of the nuisance parameters Θ_α. By completing the square and adding some prior information on the nuisance parameters (i.e. a covariance matrix), we can integrate them out to get an effective chi-squared. In other words, we would like to evaluate the following integral

$$\int d^N\Theta_\alpha \exp\left(-\frac{1}{2}\left(\chi^2_0 + \delta\Theta^\alpha X_\alpha + \frac{1}{2}\delta\Theta^\alpha\delta\Theta^\beta Y_{\alpha\beta}\right)\right) \exp\left(-\frac{1}{2}\delta\Theta^\alpha (N^{-1})_{\alpha\beta}\delta\Theta^\beta\right). \tag{85}$$

where $N_{\alpha\beta}$ is the covariance matrix of the theoretical error parameters. The integration results into

$$\sqrt{(2\pi)^N \cdot \det((\tfrac{1}{2}Y + N^{-1})^{-1})}\ \exp\left(-\frac{1}{2}\chi^2_0 + \frac{1}{4}X_\gamma \left(Y + 2N^{-1}\right)^{-1}_{\gamma\delta} X_\delta\right), \tag{86}$$

and therefore,

$$\chi^2_{\text{eff}} = \chi^2_0 - \frac{1}{2}X_\gamma \left(Y + 2N^{-1}\right)^{-1}_{\gamma\delta} X_\delta + \ln\left(\det((\tfrac{1}{2}Y + N^{-1}))\right). \tag{87}$$

Please note that all these terms do in general depend on Θ_i. Since the joint probability distribution of all parameters is a multivariate Gaussian, we know that Y is independent of Θ_i and X only depends linearly on Θ_i. In that case, we get

$$\left(\chi^2_{\text{eff}}\right)_{ij} = \left(\chi^2_0\right)_{ij} - \frac{1}{2}X_{i\gamma}\left(Y + 2N^{-1}\right)^{-1}_{\gamma\delta} X_{\delta j}. \tag{88}$$

The full Fisher matrix is given by

$$F_{\mu\nu} = \begin{pmatrix} \frac{1}{2}Y_{\alpha\beta} + N^{-1}_{\alpha\beta} & \frac{1}{2}X_{\alpha j} \\ \frac{1}{2}X_{i\beta} & F_{ij} \end{pmatrix} \tag{89}$$

where we have to evaluate the matrices at the maximum likelihood value of the parameters. This means that the effective chi-squared is given by

$$\left(\chi^2_{\rm eff}\right)_{ij} = \left(\chi^2_0\right)_{ij} - 2F_{i\gamma}F^{-1}_{\gamma\delta}F_{\delta j}, \tag{90}$$

or, in other words, the effective Fisher matrix for the theoretical parameters is given by

$$F^{\rm eff}_{ij} = F_{ij} - F_{i\gamma}F^{-1}_{\gamma\delta}F_{\delta j}. \tag{91}$$

This is what we found before in the main text.

Alternatively, starting from (88) we can write down immediately the expression for the effective likelihood

$$\begin{aligned}\mathcal{L}^{\rm eff} &= \frac{1}{\sqrt{\det\left(\frac{1}{2}Y+N^{-1}\right)}}\exp\left[-\frac{1}{2}\left(\chi^2_0 - X_\gamma(\tfrac{1}{2}Y+N^{-1})^{-1}_{\gamma\delta}X_\delta\right)\right]\\ &= \frac{\exp\left[-\frac{1}{2}\sum_{\hbar,p}\Delta B(\hbar)\left(C_B^{-1}(\hbar,p) - C_B^{-1}(\hbar,\hbar_\gamma)B_{2L}(k_\gamma)(D+N^{-1})^{-1}_{\gamma\delta}B^{\rm er}(k_\delta)C_B^{-1}(\hbar_\delta,p)\right)\Delta B(p)\right]}{\sqrt{\det\left(D+N^{-1}\right)}}\\ &= \frac{1}{\sqrt{\det\left(D+N^{-1}\right)}}\exp\left[-\frac{1}{2}\sum_{\hbar_\alpha,\hbar_\beta}\Delta B(\hbar_\alpha)\left(C_B+N^{\rm eff}\right)^{-1}_{\alpha\beta}\Delta B(\hbar_\beta)\right].\end{aligned} \tag{92}$$

Here the difference between the data and theory vector ΔB is evaluated at the fiducial values for the nuisance parameters $\bar{\Theta}_\alpha$ (i.e. at zero in our case). Taking now two derivatives with respect to the remaining theoretical parameters Θ_i, we find the effective Fisher matrix

$$F^{\rm eff}_{ij} = \sum_{\hbar_\alpha,\hbar_\beta} B_i(\hbar_\alpha)\left(C_B+N^{\rm eff}\right)^{-1}_{\alpha\beta}B_j(\hbar_\beta) \tag{93}$$

with $N^{\rm eff}_{\alpha\beta} = B^{\rm er}(k_\alpha)N_{\alpha\beta}B^{\rm er}(k_\beta)$.

B.3 Theoretical error - a toy model

In our approach, the theoretical error on the value of the bispectrum is modeled in the following way. For every bin, we introduce a nuisance parameter that is drawn from a Gaussian distribution with average zero and variance set by the estimated size of the theoretical error for that bin. Importantly, we allow for non-vanishing correlations among these nuisance parameters, i.e. we allow for a non-diagonal covariance matrix for them. The purpose of this appendix is to show that both the limit of zero and maximal correlation among the parameters have a clear interpretation, neither of which resembles the way we think the theoretical error should act. To be more precise, we prove, by means of a simple toy model that still captures the essence of the real analysis, the intuitive statements that:

- for zero correlation length, the theoretical error just acts as shot noise per bin;
- for maximal correlation length, the theoretical error acts as some free coefficient multiplying a fixed shape function, which by definition we think is the wrong function.

B.3.1 Toy model

We consider measuring some observable d a total of N times and collecting the data d_i. Our model is $d_i = x + e_i$, with x a Gaussian random variable with variance σ_x^2, whose average, $\bar{x}$, we would like to determine as well as possible. The e_i are additional Gaussian variables that represent the systematic error or theoretical uncertainty in every measurement. Their averages and variances

are $\bar{e_i}$ and $\sigma^2_{e_i}$, respectively. One can think of this scenario as determining the average weight of a group of people, knowing that their weights are Gaussian distributed with variance σ^2_x, where we use a different weighing scale with a systematic error $\bar{e_i}$ and some uncertainty in the measurement characterized by $\sigma^2_{m_i}$ every time we weigh someone. Since the e_i are uncorrelated with x, this leads to the likelihood

$$\log L = -\sum_{i=1}^{N} \frac{(d_i - \bar{x} - \bar{e_i})^2}{2\sigma_d^2}, \tag{94}$$

where $\sigma_d^2 = \sigma_x^2 + \sigma_m^2$, assuming for convenience that $\sigma_{m_i} = \sigma_m$ (the arguments below do not depend on this assumption). Without any prior information on the systematic errors, they are completely degenerate with $\bar{x}$, so we do not expect to be able to learn anything about $\bar{x}$ in this case. This can be verified using a Fisher analysis. We have

$$F_{ab} = \begin{pmatrix} \frac{N}{\sigma_d^2} & \frac{1}{\sigma_d^2}\overrightarrow{1}^T \\ \frac{1}{\sigma_d^2}\overrightarrow{1} & \frac{1}{\sigma_d^2}1_{N\times N} \end{pmatrix}, \tag{95}$$

where $a, b = \bar{x}, \bar{e_i}$. Since we are ignorant about the systematic errors, we compute the marginalized error on $\bar{x}$,

$$\sigma^2_{\bar{x},marg} = \left(F^{-1}\right)_{\bar{x}\bar{x}}, \tag{96}$$

which can be computed using the block matrix inversion formula (see also (22)):

$$\text{let} \quad F = \begin{pmatrix} A & B^T \\ B & D \end{pmatrix}, \tag{97}$$

then

$$\sigma^2_{\bar{x},marg} = (A - B^T D^{-1} B)^{-1} = \left(\frac{N}{\sigma_d^2} - \frac{1}{\sigma_d^2}\overrightarrow{1}^T 1_{N\times N}\overrightarrow{1}\right)^{-1} = \frac{1}{0}, \tag{98}$$

as expected. In a realistic situation we do have some prior information about the systematic errors. Here, and in the paper, we assume they are also Gaussian random variables with some variance $\sigma^2_{\bar{e}_i}$. Moreover, we allow for nontrivial correlations among the $\bar{e_i}$, which for the scales could mean they were produced by the same machine for instance. This means we obtain the updated likelihood

$$\log L = -\sum_{i=1}^{N} \frac{(d_i - \bar{x} - \bar{e_i})^2}{2\sigma_d^2} - \bar{e_i}\left(C^{-1}\right)_{ij}\bar{e_j}, \tag{99}$$

where

$$C_{ij} = < \bar{e_i}\bar{e_j} > . \tag{100}$$

In the following, we investigate the effect of zero and maximal correlation length on $\sigma^2_{\bar{x},marg}$.

B.3.2 Zero correlation

Let us first assume zero correlation among the systematic errors, leading to a diagonal covariance matrix,

$$C_{ij} = \sigma_i^2 \delta_{ij}. \tag{101}$$

In terms of the weighing scales this could mean all scales really come from different companies with uncorrelated systematic errors. We now show that in this case the ignorance about the

systematic errors acts as shot noise per bin; it simply updates the variance of the measurements $\sigma_d^2 \to \sigma_{d_i}^2$. For any covariance matrix, the Fisher matrix is

$$F = \begin{pmatrix} \frac{N}{\sigma_d^2} & \frac{1}{\sigma_d^2}\overrightarrow{1}^T \\ \frac{1}{\sigma_d^2}\overrightarrow{1} & \frac{1}{\sigma_d^2}\delta_{ij} + (C^{-1})_{ij} \end{pmatrix}. \tag{102}$$

Then, using the block matrix inversion formula, zero correlation leads to an error

$$\sigma^2_{\bar{x},marg} = \left(\frac{N}{\sigma_d^2} - \frac{1}{\sigma_d^4}\overrightarrow{1}^T \frac{1}{\frac{1}{\sigma_d^2}+\frac{1}{\sigma_i^2}}\delta_{ij}\overrightarrow{1} \right)^{-1} \tag{103}$$

$$= \left[\sum_{i=1}^{N} \left(\frac{1}{\sigma_d^2} - \frac{1}{\sigma_d^4}\frac{1}{\frac{1}{\sigma_d^2}+\frac{1}{\sigma_i^2}} \right) \right]^{-1} \tag{104}$$

$$= \left[\sum_{i=1}^{N} \left(\frac{1}{\sigma_d^2+\sigma_i^2} \right) \right]^{-1} \equiv \left[\sum_{i=1}^{N} \frac{1}{\sigma_{d_i}^2} \right]^{-1}, \tag{105}$$

which is the same error one gets from assuming the likelihood function

$$\log L = -\sum_{i=1}^{N} \frac{(d_i - \bar{x})^2}{2\sigma_{d_i}^2}. \tag{106}$$

This shows that indeed for zero correlation the systematic errors acts as shot noise per bin. In particular, this means the error on $\bar{x}$ can be made arbitrarily small by increasing the number of measurements (if the $\sigma_{\bar{e}_i}$ do not grow too fast for additional measurements). The intuitive reason is of course that in this model we expect the systematic errors to average out to zero in the long run. This is clearly not what is expected of the theoretical error in the measurement of the bispectrum.

B.3.3 Maximal correlation

Next we assume maximal correlation, which by definition means

$$C_{ij} = <\bar{e_i}\bar{e_j}> = \sqrt{<\bar{e_i}^2>}\sqrt{<\bar{e_j}^2>} = \sigma_i\sigma_j. \tag{107}$$

Since this matrix has rank one (all columns are multiples of the same vector), it is not invertible in more than one dimension. One way to deal with this is to introduce a regulator, such as a small matrix $\epsilon\delta_{ij}$, to break the degeneracy. Using our block matrix inversion formula, this is however not necessary. In the notation of ((97)), we wish to compute D^{-1}. Let us write $D = S + C^{-1}$, where $S = \frac{1}{\sigma_d^2}\delta_{ij}$. The Woodbury identity then gives

$$(S + C^{-1})^{-1} = S^{-1} - S^{-1}(S^{-1} + C)^{-1}S^{-1}. \tag{108}$$

Hence, we have to compute the inverse of $S^{-1} + C$, where $S^{-1} = \sigma_d^2\delta_{ij}$, and $C = \sigma_i\sigma_j$. Conveniently, since C is of the form $\overrightarrow{\sigma}(\overrightarrow{\sigma})^T$, we can use the Sherman-Morrison formula to compute the inverse

$$(S^{-1} + C)^{-1} = S - \frac{S(\sigma_i\sigma_j)S}{1 + \sigma_i S_{ij}\sigma_j}. \tag{109}$$

Plugging this into the previous formula, we find

$$D^{-1} = (S + C^{-1})^{-1} = S^{-1} - S^{-1} + \frac{\sigma_i\sigma_j}{1 + \frac{(\sum\sigma_i^2)}{\sigma_d^2}} = \frac{\sigma_i\sigma_j}{1 + \frac{(\sum\sigma_i^2)}{\sigma_d^2}}. \tag{110}$$

This finally leads to the error on $\bar{x}$:

$$\sigma^2_{\bar{x},marg} = \left[\frac{N}{\sigma_d^2} - \frac{\frac{(\sum \sigma_i)^2}{\sigma_d^2}}{\sigma_d^2 + \left(\sum \sigma_i^2\right)}\right]^{-1}. \tag{111}$$

In order to interpret this result, let us rewrite this expression as follows

$$\sigma^2_{\bar{x},marg} = \left[\frac{N}{\sigma_d^2} - \frac{\left(\sum \frac{\sigma_i}{\sigma}\right)^2}{\sigma_d^4}\frac{1}{\frac{1}{\sigma^2} + \frac{\sum\left(\frac{\sigma_i}{\sigma}\right)^2}{\sigma_d^2}}\right]^{-1}, \tag{112}$$

where we have introduced the dimensionful parameter σ to keep the dimensions clean. Now observe that we get exactly the same error on $\bar{x}$ from the following likelihood function

$$\log L = -\sum_{i=1}^{N} \frac{(d_i - \bar{x} - \frac{\sigma_i}{\sigma}\bar{e})^2}{2\sigma_d^2} - \frac{\bar{e}^2}{2\sigma^2}. \tag{113}$$

whose Fisher matrix is

$$F = \begin{pmatrix} \frac{N}{\sigma_d^2} & \frac{1}{\sigma_d^2}\sum \frac{\sigma_i}{\sigma} \\ \frac{1}{\sigma_d^2}\sum \frac{\sigma_i}{\sigma} & \frac{1}{\sigma^2} + \frac{\sum\left(\frac{\sigma_i}{\sigma}\right)^2}{\sigma_d^2} \end{pmatrix}, \tag{114}$$

This means that the maximal correlation case is equivalent to having a single, unknown parameter multiplying a known 'shape' function σ_i/σ. In terms of the weighing scales this would mean that we know in advance exactly the ratios between the systematic errors of the scales. In terms of the bispectrum this would mean that we claim to know the theoretical error is exactly some number times the two loop estimate we put in, which it is clearly not. Finally note that if we choose all σ_i to be equal, which for convenience we take to be σ, we find

$$\sigma^2_{\bar{x}} = \left[\frac{N}{\sigma_d^2} - \frac{N^2}{\sigma_d^4}\frac{1}{\frac{1}{\sigma^2} + \frac{N}{\sigma_d^2}}\right]^{-1} = \frac{\sigma_d^2}{N} + \sigma^2, \tag{115}$$

meaning the error on $\bar{x}$ can never get below the uncertainty in the degenerate parameter $\bar{e}$. In terms of the weighing problem this makes perfect sense, as this case is equivalent to simply using one and the same scale for every measurement. In this case we never expect to beat the unknown systematic error in the scale. In terms of the bispectrum this shows the importance of the relation between the shapes of the non-Gaussian signal and the theoretical error. In fact, in the maximal correlation limit we treat the theoretical error exactly the same as the EFT terms.

B.3.4 Conclusions

From the above example it is clear that in neither limit the implementation of the theoretical error is completely satisfactory. Moreover, if the shapes are not too similar, the estimates from both limits are probably too optimistic. For this reason we recommend a conservative use of the method. In particular, we choose to use the correlation length that gives the weakest constraints on $f_{\rm NL}$, as we show in the next subsection.

B.4 Choice of correlation length

In order to find the most conservative correlation length to work with, we ran a test computation of $\sigma(f_{\rm NL})$. We did this at redshift zero with $k_{\rm min} = 0.001\ h{\rm Mpc}^{-1}$ and $k_{\rm max} = 1\ h{\rm Mpc}^{-1}$, where we divide the k-range in 9 (blue), 15 (orange), 27 (green), 45 (red) and 81 (purple) bins. We find

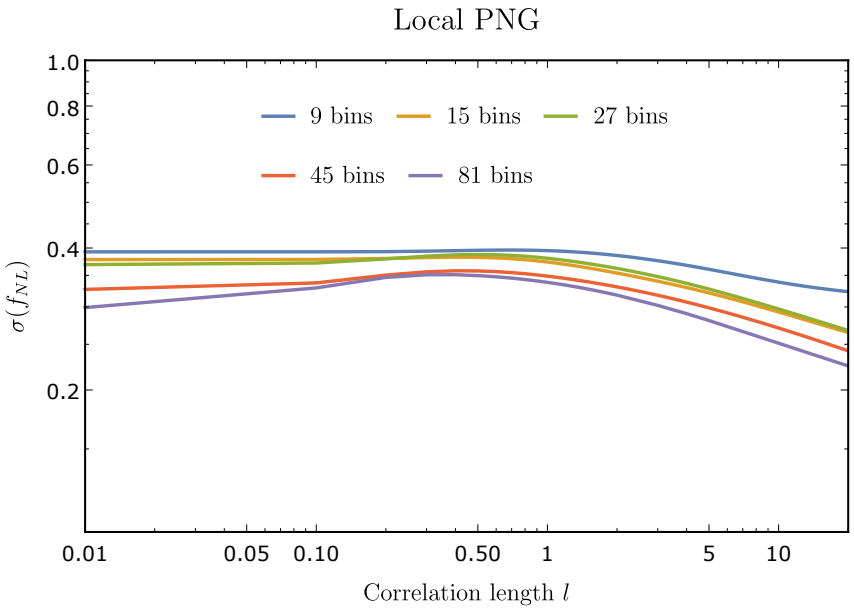


Figure 7: A test computation of $\sigma(f_{NL})$ at redshift $z = 0$ including theoretical error as function of correlation length. We choose $k_{\rm min} = 0.001\ h{\rm Mpc}^{-1}$ and $k_{\rm max} = 1\ h{\rm Mpc}^{-1}$ where we divide the k-range in 9 (blue), 15 (orange), 27 (green), 45 (red) and 81 (purple) bins.

that the weakest constraints are obtained for $l \approx 0.5$, see Figure 7. This is therefore the value we take for the analysis in the paper.

Remarkably, the error is actually very insensitive to the correlation length, despite the very different nature of the effect of small and large correlation length. We believe the reason for this to be the fact that our ansatz for the theoretical error is a much steeper function of k than the non-Gaussian signal. The transition from the k's for which the error is negligible to the region where it is completely dominant is therefore very small, and the shape of the error is therefore not very important in this case.

Another observation is that the error keeps increasing as we increase the correlation length beyond 10 decades, whereas the k's we consider only run over a couple of decades. This makes the nuisance parameters almost maximally correlated for all these large correlation lengths. At the moment, we have no good explanation for the fact that the error seems to keep improving, other than it being a numerical fluke, perhaps related to the inversion of the correlation matrix.

B.5 Ansätze for higher loop corrections

We compare the ansätze for the higher loop corrections in Figure 8. It is a zoom-in of Figure 1, where we now show in addition the ansatz used in [15], for the two and three loop contribution to the bispectrum. Please note that in [14], we only showed one of the two reducible two loop diagrams contributing to B_{332}. Therefore, the plots look different now, in particular in the squeezed configuration of the bispectrum. We see that in the squeezed configuration, the ansatz B_{332} is an order of magnitude smaller than E_b. This explains why we have to multiply B_{332} by a factor 10 in section 3.5 to get reliable results. Furthermore, we note that at redshift zero, B_{332} allows one to go to higher $k_{\rm max}$ in the squeezed configuration, whereas E_b allows one to go further in the equilateral configuration (for f_{NL} bigger than 10). This explains why using B_{332} as an ansatz gives more optimistic results for local PNG, whereas E_b gives more optimistic results for equilateral PNG (see section 4.1). Keep in mind also that the time dependence of the theoretical error terms is different from the signal, making the signal stronger at higher redshifts.

Next, we consider $\sigma(f_{NL})$ as function of $k_{\rm max}$ at various redshifts in Figure 9. We do not include shotnoise, but we integrate out the theoretical error. The result for local PNG is shown in the left panel. We see that at redshift zero, the signal freezes out at some $k_{\rm max} < 1h{\rm Mpc}^{-1}$. Furthermore, in agreement with what we expect from Figure 8, we see that using B_{332} as ansatz for the two loop corrections gives more optimistic results. More specifically, $\sigma(f_{NL})$ is a factor 5 smaller. At higher redshifts, we find that $\sigma(f_{NL})$ does not freeze when we reach $k_{\rm max} = 1h{\rm Mpc}^{-1}$. We think this is due to the fact that we keep gaining information as we go to more and more

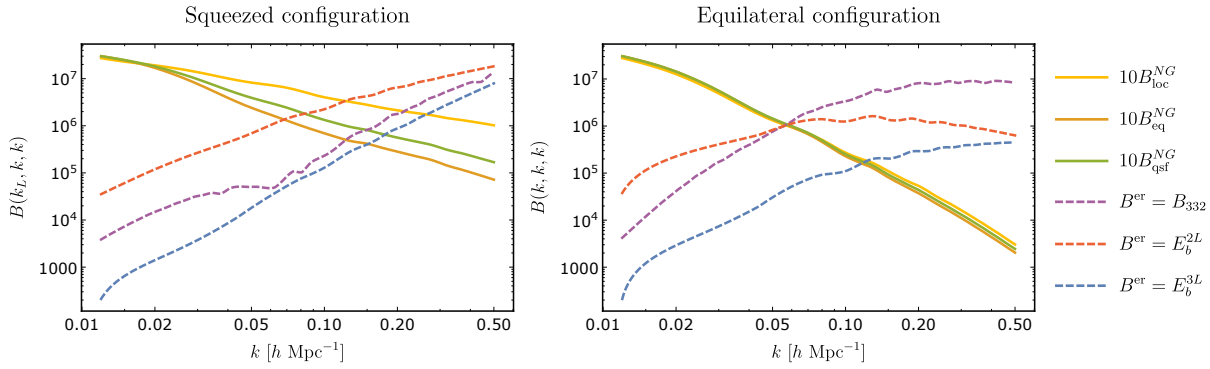


Figure 8: Comparison of the ansätze for the higher loop corrections. We plot E_b from equation (27) for two and three loops (yellow and green dashed lines) versus B_{332} (green dashed line) defined in equation (77). For the two loop ansatz using E_b we take $n = -1.4$, $k_{\mathrm{NL}} = 0.45h\mathrm{Mpc}^{-1}$ and $l = 2$ and for the three loop ansatz we use $n = -1.5$, $k_{\mathrm{NL}} = 0.50h\mathrm{Mpc}^{-1}$ and $l = 3$. We compare these ansätze with the non-Gaussian contribution to the bispectrum up to one loop with $f_{NL} = 10$ for local, equilateral and quasi-single-field PNG (red, blue and purple solid lines). In the left panel we compare the different contributions in the configuration $B(k_L, k, k)$ where we varied k and fixed $k_L = 0.012h\mathrm{Mpc}^{-1}$. The smaller k the more squeezed the configuration is. In the right panel we show the equilateral configuration $B(k, k, k)$.

squeezed configurations. This is also important for equilateral PNG, shown in the right panel, even though, $\sigma(f_{NL})$ does freeze out in this case. Interestingly, compared to scaling estimates for k_{max} for equilateral PNG (see for instance [15]) we find that we can go to much smaller scales than naively thought. The squeezed limit allows us to extract more information, also for equilateral PNG. The fact that $k_{\mathrm{max}} = 1h\mathrm{Mpc}^{-1}$ is not large enough to ensure that $\sigma(f_{NL})$ is frozen when we ignore shotnoise explains the results we find in section 4.1.

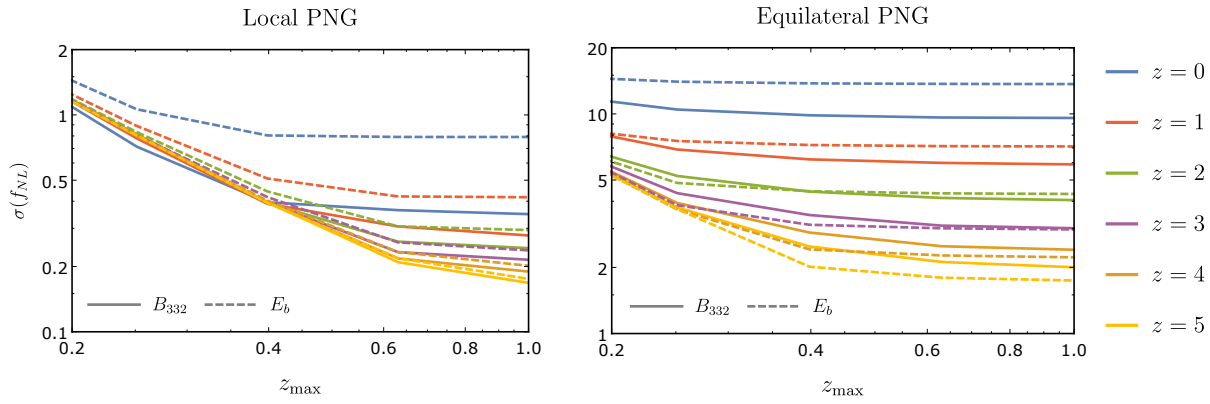


Figure 9: We show $\sigma(f_{NL})$ as function of k_{max} using B_{332} and E_b as ansatz for the two loop corrections (solid and dashed lines). In the left panel we show the results for local PNG and in the right panel for equilateral PNG. The redshift takes values between $z = 0$ and $z = 5$. We use $k_{\mathrm{min}} = 0.001h\mathrm{Mpc}^{-1}$ and $V = (2\pi/k_{\mathrm{min}})^3$ at each redshift.

C Choice of binning and volume of the bins

In this appendix, we motivate the decision of section 3.3 to use logarithmic binning and exactly computed values of V_{123}.

C.1 Exact computation of V_{123}

We will now explain how to compute V_{123} exactly and systematically by dividing all the bins in 'interior' and 'edge' bins. Moreover, the selection of bins is now determined by whether it

contains at least some valid triangles instead of the usual selection rule that the central point should be a triangle.
Recall that V_{123} is defined as

$$V_{123} = \int_{\mathbf{q}_1} \int_{\mathbf{q}_2} \int_{\mathbf{q}_3} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \tag{116}$$

We choose logarithmic binning, i.e. we have

$$q_i \equiv |\mathbf{q}|_i \in \left[k_i e^{-\frac{1}{2}\Delta \ln k}, k_i e^{\frac{1}{2}\Delta \ln k} \right]. \tag{117}$$

The integrand above only depends on the relative orientations of the vectors and their lengths. Fixing $\mathbf{q}_1$ along the $\hat{z}$-direction and $\mathbf{q}_2$ to be in the (x, z)-plane, their relative orientation is given by $\theta_{12} = \theta_2$. Now the lengths of these vectors, together with c_{12}, the cosine of θ_{12}, completely determine $\mathbf{q}_3$. The length of $\mathbf{q}_3$ is then restricted to be in $\left[k_3 e^{-\frac{1}{2}\Delta \ln k}, k_3 e^{\frac{1}{2}\Delta \ln k} \right]$, which means

$$c_{12} \in [-1, 1] \cap \left[\frac{\left(k_3 e^{-\frac{1}{2}\Delta \ln k} \right)^2 - q_1^2 - q_2^2}{2 q_1 q_2}, \frac{\left(k_3 e^{\frac{1}{2}\Delta \ln k} \right)^2 - q_1^2 - q_2^2}{2 q_1 q_2} \right], \tag{118}$$

where q_1 and q_2 also take values within their bin. Then, if $[-1, 1]$ contains the range on the right for all values of q_1 and q_2, we are dealing with an 'interior bin', and we get

$$\int dc_{12} dq_1 dq_2 \; q_1^2 q_2^2 = k_1^2 k_2^2 k_3^2 \sinh^3(\Delta \ln k). \tag{119}$$

Finally, accounting for the fact that we fixed θ_1, $\phi_{1,2}$ and the factors of $(2\pi)^3$ we find

$$V_{123} = \frac{1}{(2\pi)^9} 8\pi^2 k_1^2 k_2^2 k_3^2 \sinh^3(\Delta \ln k). \tag{120}$$

This approximation breaks down when the two ranges are partly overlapping, in which case we have an 'edge bin'.
let us evaluate V_{123} more precisely. We have seen that the integral simplifies to

$$V_{123} = \frac{8\pi^2}{(2\pi)^9} \int_{q_1} \int_{q_2} \int_{c_{12}} q_1^2 q_2^2 \tag{121}$$

where the c_{12} is restricted to be in the range given above. This integral can therefore be rewritten as

$$V_{123} = \frac{8\pi^2}{(2\pi)^9} \int_{q_1} \int_{q_2} \tfrac{1}{2} q_1 q_2 \max\left[0, \left(\min\left[(q_1 + q_2)^2, k_3^2 e^{\Delta \ln k} \right] - \max\left[(q_1 - q_2)^2, k_3^2 e^{-\Delta \ln k} \right] \right) \right].$$

In other words we integrate over the overlap

$$\left[(q_1 - q_2)^2, (q_1 + q_2)^2 \right] \cap \left[k_3^2 e^{-\Delta \ln k}, k_3^2 e^{\Delta \ln k} \right]. \tag{122}$$

There are multiple possibilities:

- The overlap is always zero.

  This happens whenever $|q_1 + q_2|_{\max} \leq k_3 e^{-\frac{1}{2}\Delta \ln k}$ or $|q_1 - q_2|_{\min} \geq k_3 e^{\frac{1}{2}\Delta \ln k}$. This means we should exclude the cases $k_3 \geq (k_1 + k_2) e^{\Delta \ln k}$ and $k_3 < k_2$. The latter is already excluded since we have $k_1 \leq k_2 \leq k_3$. The first leads to a constraint to select the bins, namely

  $$k_3 < (k_1 + k_2) e^{\Delta \ln k}. \tag{123}$$

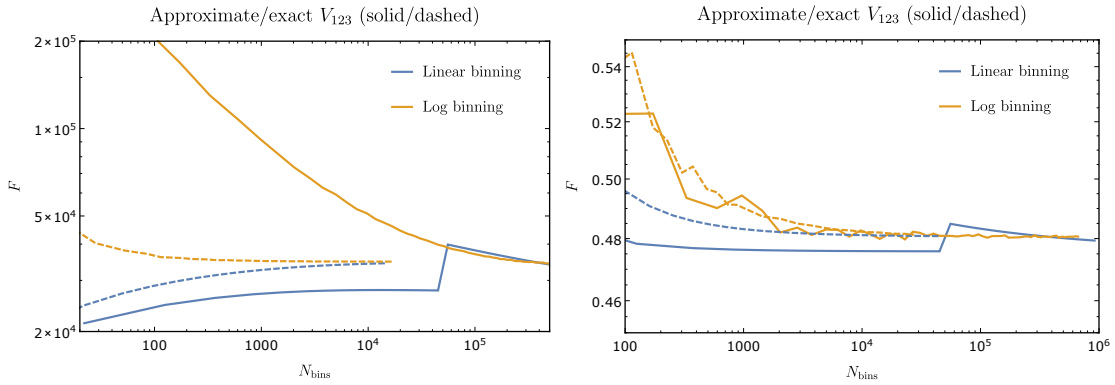
- The first range always contains the second range.
  This happens when $|q_1 - q_2|_{\max} \leq k_3 e^{-\frac{1}{2}\Delta \ln k}$ ánd $|q_1 + q_2|_{\min} \geq k_3 e^{\frac{1}{2}\Delta \ln k}$. So we need both
  $$\begin{cases} k_3 \geq k_2 e^{\Delta \ln k} - k_1 \\ k_3 \leq (k_1 + k_2) e^{-\Delta \ln k} \end{cases}. \tag{124}$$
  In this case the volume takes the simple form
  $$V_{123} = \frac{8\pi^2}{(2\pi)^9} k_1^2 k_2^2 k_3^2 \sinh^3(\Delta \ln k). \tag{125}$$

- Any other type of overlap.
  For the other cases we have to compute the actual volume of the bin. We will numerically perform the integral given above. This is when one of the two options below is satisfied
  $$\begin{cases} k_3 < k_2 e^{\Delta \ln k} - k_1 \\ k_3 > (k_1 + k_2) e^{-\Delta \ln k} \end{cases} \tag{126}$$
  Not for all these edge bins the central point has to be a triangle, since there are some cases for which $k_3 > k_1 + k_2$, considering the second inequality. Thus, we can either decide to define another point in the bin to represent the central triangle or we can merge these bins with one of their neighbors. In the first case a valid central triangle in the bin (k_1, k_2, k_3) is given by
  $$\left(k_1 e^{\frac{k_3}{2(k_2+k_1)}}, k_2 e^{\frac{k_3}{2(k_2+k_1)}}, k_3 e^{-\frac{k_3}{2(k_2+k_1)}} \right) \tag{127}$$
  The other option is to merge (i.e. we add the volumes) the bin with one of its neighbors, which has the advantage that we never have to change the representing triangle of a given bin. For practical reasons, we choose this option. We implement this by merging each bin $\mathbf{k}$ for which $k_3 > k_1 + k_2$ with the bin $\mathbf{p}$ which has $p_1 = k_1$, $p_2 = k_2$ and p_3 the biggest value below or equal to $k_1 + k_2$.

C.2 Logarithmic versus linear binning

Let us now compare logarithmic binning with linear binning. We will show two examples of a computation of a Fisher matrix and show that the linear binning might cause problems.
We assume the following form for the Fisher matrix
$$F = \sum_{k_1,k_2,k_3} f(k_1, k_2, k_3) \frac{V_{123}}{s_{123}}, \tag{128}$$
for some function $f(k_1, k_2, k_3)$. We will consider a 'local'-type function f^{loc} and an 'equilateral'-type function f^{eq}. The local function corresponds to assuming the late time power spectrum scales as $P(k) \sim k^{-3}$, where F represents the (f_{NL}, f_{NL})-component of the Fisher matrix for local PNG. Forgetting about the right normalization, this gives
$$f^{\text{loc}}(k_1, k_2, k_3) = \left(\frac{k_1^{3/2}}{k_2^{3/2} k_3^{3/2}} + \frac{k_2^{3/2}}{k_1^{3/2} k_3^{3/2}} + \frac{k_3^{3/2}}{k_2^{3/2} k_1^{3/2}} \right)^2. \tag{129}$$
Similarly, we can define a function that corresponds to equilateral PNG
$$f^{\text{eq}}(k_1, k_2, k_3) = \frac{k_1 k_2 k_3}{(k_1 + k_2 + k_3)^6}. \tag{130}$$
We compute F over a range $k \in [0.003, 0.5]$ $h\text{Mpc}^{-1}$ for both logarithmic and linear bins and for both the approximate and exact computation of V_{123}. In Figure 10 we plot F as function of

number of bins (number of triangles) considered.
First of all, in Figure 10a, restricting ourselves to about 10.000 bins, we see that if we use the approximate value for V_{123} (the usual assumption), the linear bins seem to converge quickly to the asymptotic value. The logarithmic bins seem to converge much slower. However, as we keep increasing the number of triangles, suddenly the graph of the linear bins jumps to the graph of the logarithmic bins. This shows that if we would have trusted the linear binning for a smaller number of bins we would have gotten the wrong result. This is quite unexpected and alarming as it seems we cannot always trust linear binning! If we now change to the exact V_{123} we see that both linear and logarithmic binning converge much faster and both to the same value. In fact, it turns out that they reach one percent agreement for about 15.000 triangles. Let us try to understand why this happens. As we are summing over a function which peaks in the squeezed limit we do in fact get most signal from k-triplets which satisfy $k_3 \sim k_2 \gg k_1$. In particular the edge bins will contribute an important part to the final result. We know that precisely for these bins the approximate value for V_{123} does not work, which is probably why the results improve dramatically when using the exact V_{123}. Now one can still wonder why the linear binning performs so badly in this case. A reason might be that we are sampling the values for k_1 much better in case of logarithmic binning. However one could argue exactly the opposite, namely that linear binning samples the values of k_2 and k_3 much better. We have not found a convincing argument why linear binning fails, this remains an open question. As we are also studying the Fisher matrix for local PNG in our paper, we decided to stick to logarithmic binning. Even with the exact value of V_{123} the result converges quite slowly for the local function. In order to be within a couple of percent of the actual outcome of the Fisher analysis we need quite some triangles. For the analysis therefore we divide the k-axis over three logarithmic decades in 45 bins. By this we mean that if for instance each k_i from the triplet can take values in the range $[0.001, 1]$ $h\text{Mpc}^{-1}$, it can take one of the 45 logarithmically separated values.
We did the same analysis for the equilateral function. In Figure 10b we see again a jump of the graph corresponding to linear binning. This time we do not expect to gain most signal from the edge bins. However, when we use the exact value for V_{123} everything seems to be fine again. The jump takes place at a comparable value of N_{bins}. The graph of the logarithmic binning remains a bit wiggly, but we find one percent agreement between linear and logarithmic binning already for 1000 triangles. For equilateral PNG we therefore divide the k-axis over three logarithmic decades in either 27 or 45 bins.



(a) Local function f^{loc} **(b)** Equilateral function f^{eq}

Figure 10: Computation of F as function of the number of triangles N_{bins} for (a) f^{loc} and (b) f^{eq} as given in equations (129) and (130). We show the results for both the approximate (solid) and exact (dashed) expression for V_{123}. Moreover we denote the results from linear binning with a blue line and logarithmic binning with a orange line.

D Table of parameters

Symbol	Relation	Meaning
a		scale factor
τ	$a\,\mathrm{d}\tau = \mathrm{d}t$	conformal time
$\mathcal{H}$	$\equiv d\ln(a)/d\tau$	conformal Hubble parameter
$\mathcal{H}_0$		present value of $\mathcal{H}$
$\boldsymbol{x}$		comoving coordinate
$\boldsymbol{k}$		momentum
Ω_m		matter density in units of the critical density
Ω_Λ		dark energy density
h		dimensionless Hubble constant
ρ		dark matter density
δ	$\equiv \delta\rho/\rho$	dark matter density contrast
θ	$\equiv \partial_i v^i$	velocity divergence
$\delta_{(n)}$		density contrast in SPT at order n
F_n		kernel function in $\delta_{(n)}$
P_{mn}	$\equiv \langle\delta_{(m)}\delta_{(n)}\rangle'$	power spectrum in SPT
B_{lmn}	$\equiv \langle\delta_{(l)}\delta_{(m)}\delta_{(n)}\rangle'$	bispectrum in SPT
ϕ		Newtonian potential
Φ	$\triangle\Phi = \delta$	rescaled Newtonian potential
φ	$\phi = T(k)\varphi$	primordial potential
φ_g		Gaussian primordial potential
$T(k)$		transfer function
$M(k)$		transfer function in the Poisson equation
D_1		linear growth factor
f	$\equiv d\ln D_1/\ln a$	growth rate
P_φ		primordial power spectrum
B_φ		primordial bispectrum
n_s		scalar spectral index
ψ		correlation in the initial conditions
Ψ	$\Psi(\boldsymbol{x}) \equiv \psi(\boldsymbol{q}(\boldsymbol{x}))$	Eulerian definition of ψ
Δ		scaling dimension in K_{NL}
f_{NL}		amplitude of the primordial bispectrum
d^2	$\equiv c_s^2 + f(c_{vis}^2 + \hat{c}_{vis}^2)$	parameter in τ_v
e_i, g, g_i		parameters in τ_v
ξ		parameter in $\delta^c_{(1)}$
γ		parameter in $\delta^c_{(1)}$
ϵ_i		parameter in $\delta^c_{(2)}$
γ_i		parameter in $\delta^c_{(2)}$

Symbol	Relation	Meaning
$\mathcal{S}_{\alpha,\beta}$		SPT quadratic source terms
τ_θ		EFT source in Euler equation
$\delta^c_{(n)}$		viscosity counterterm at order n
F^c_n		kernel function in $\delta^c_{(n)}$
H^c_n		kernel function in $\delta^c_{(n)}$
G_δ		Green's function for δ
$\mathcal{D}_\delta$		evolution operator in the fluid equation
$P_{1\psi}$	$\equiv \langle\delta_{(1)}\psi\rangle'$	correlation of $\delta_{(1)}$ and ψ
$B^{\rm G}_{\rm SPT}$		Gaussian SPT contributions to B_δ
$B^{\rm NG}_{\rm SPT}$		non-Gaussian SPT contributions to B_δ
$B^{\rm G}_{\rm EFT}$		sum of Gaussian EFT counterterms
$B^{\rm NG}_{\rm EFT}$		sum of non-Gaussian EFT counterterms

References

[1] 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, 0710.4560.

[2] O. Dor *et al.*, "Cosmology with the SPHEREX All-Sky Spectral Survey," 2014, 1412.4872.

[3] M. Tellarini, A. J. Ross, G. Tasinato, and D. Wands, "Galaxy bispectrum, primordial non-Gaussianity and redshift space distortions," 2016, 1603.06814.

[4] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, "Cosmological Non-Linearities as an Effective Fluid," *JCAP*, vol. 1207, p. 051, 2012, 1004.2488.

[5] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, "Large scale structure of the universe and cosmological perturbation theory," *Phys.Rept.*, vol. 367, pp. 1–248, 2002, astro-ph/0112551.

[6] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, "Probing primordial non-Gaussianity with large - scale structure," *Phys. Rev.*, vol. D69, p. 103513, 2004, astro-ph/0312286.

[7] 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, 0705.0343.

[8] T. Baldauf, U. Seljak, and L. Senatore, "Primordial non-Gaussianity in the Bispectrum of the Halo Density Field," *JCAP*, vol. 1104, p. 006, 2011, 1011.1513.

[9] D. Jeong and E. Komatsu, "Primordial non-Gaussianity, scale-dependent bias, and the bispectrum of galaxies," *Astrophys. J.*, vol. 703, pp. 1230–1248, 2009, 0904.0497.

[10] G. Tasinato, M. Tellarini, A. J. Ross, and D. Wands, "Primordial non-Gaussianity in the bispectra of large-scale structure," *JCAP*, vol. 1403, p. 032, 2014, 1310.7482.

[11] N. Roth and C. Porciani, "Can we really measure f_{NL} from the galaxy power spectrum?," *Mon. Not. Roy. Astron. Soc.*, vol. 425, pp. L81–L85, 2012, 1205.3165.

[12] T. Baldauf, L. Mercolli, M. Mirbabayi, and E. Pajer, "The Bispectrum in the Effective Field Theory of Large Scale Structure," *JCAP*, vol. 1505, no. 05, p. 007, 2015, 1406.4135.

[13] R. E. Angulo, S. Foreman, M. Schmittfull, and L. Senatore, "The One-Loop Matter Bispectrum in the Effective Field Theory of Large Scale Structures," *JCAP*, vol. 1510, no. 10, p. 039, 2015, 1406.4143.

[14] V. Assassi, D. Baumann, E. Pajer, Y. Welling, and D. van der Woude, "Effective Theory of Large-Scale Structure with Primordial Non-Gaussianity," 2015, 1505.06668.

[15] T. Baldauf, M. Mirbabayi, M. Simonovi, and M. Zaldarriaga, "LSS constraints with controlled theoretical uncertainties," 2016, 1602.00674.

[16] P. A. R. Ade *et al.*, "Planck 2015 results. XVII. Constraints on primordial non-Gaussianity," 2015, 1502.01592.

[17] D. Baumann, D. Green, and R. A. Porto, "B-modes and the Nature of Inflation," *JCAP*, vol. 1501, no. 01, p. 016, 2015, 1407.2621.

[18] M. Alvarez *et al.*, "Testing Inflation with Large Scale Structure: Connecting Hopes with Reality," 2014, 1412.4671.

[19] A. Lewis and S. Bridle, "Cosmological parameters from CMB and other data: a Monte-Carlo approach," *Phys. Rev.*, vol. D66, p. 103511, 2002, astro-ph/0205436.

[20] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, "The Effective Field Theory of Cosmological Large Scale Structures," *JHEP*, vol. 09, p. 082, 2012, 1206.2926.

[21] L. Mercolli and E. Pajer, "On the velocity in the Effective Field Theory of Large Scale Structures," *JCAP*, vol. 1403, p. 006, 2014, 1307.3220.

[22] N. Bartolo, S. Matarrese, and A. Riotto, "Nongaussianity from inflation," *Phys. Rev.*, vol. D65, p. 103505, 2002, hep-ph/0112261.

[23] P. Creminelli, "On non-Gaussianities in single-field inflation," *JCAP*, vol. 0310, p. 003, 2003, astro-ph/0306122.

[24] X. Chen and Y. Wang, "Quasi-Single Field Inflation and Non-Gaussianities," *JCAP*, vol. 1004, p. 027, 2010, 0911.3380.

[25] E. Pajer and M. Zaldarriaga, "On the Renormalization of the Effective Field Theory of Large Scale Structures," *JCAP*, vol. 1308, p. 037, 2013, 1301.7182.

[26] E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro, "Cosmology and the Bispectrum," *Phys. Rev.*, vol. D74, p. 023522, 2006, astro-ph/0604505.

[27] T. Baldauf, L. Mercolli, and M. Zaldarriaga, "Effective field theory of large scale structure at two loops: The apparent scale dependence of the speed of sound," *Phys. Rev.*, vol. D92, no. 12, p. 123007, 2015, 1507.02256.

[28] S. Foreman, H. Perrier, and L. Senatore, "Precision Comparison of the Power Spectrum in the EFTofLSS with Simulations," 2015, 1507.05326.

[29] J. J. M. Carrasco, S. Foreman, D. Green, and L. Senatore, "The Effective Field Theory of Large Scale Structures at Two Loops," *JCAP*, vol. 1407, p. 057, 2014, 1310.0464.

[30] A. Heavens, "Statistical techniques in cosmology," 2009, 0906.0664.

[31] E. Pajer and Y. Welling, "Time dependence in correlation functions,"

[32] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, "Nonlinear evolution of the bispectrum of cosmological perturbations," *Astrophys. J.*, vol. 496, p. 586, 1998, astro-ph/9704075.

[33] C. E. Powell, "Generating Realisations of Stationary Gaussian Random Fields by Circulant Embedding." https://www.nag.co.uk/doc/techrep/pdf/tr1_14.pdf.

[34] R. Laureijs *et al.*, "Euclid Definition Study Report," 2011, 1110.3193.

[35] K. S. Dawson *et al.*, "The Baryon Oscillation Spectroscopic Survey of SDSS-III," *Astron. J.*, vol. 145, p. 10, 2013, 1208.0022.

[36] K. S. Dawson *et al.*, "The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data," *Astron. J.*, vol. 151, p. 44, 2016, 1508.04473.

[37] M. Levi *et al.*, "The DESI Experiment, a whitepaper for Snowmass 2013," 2013, 1308.0847.

[38] M. Garny, T. Konstandin, R. A. Porto, and L. Sagunski, "On the Soft Limit of the Large Scale Structure Power Spectrum: UV Dependence," *JCAP*, vol. 1511, no. 11, p. 032, 2015, 1508.06306.

[39] E. Sefusatti and R. Scoccimarro, "Galaxy bias and halo-occupation numbers from large-scale clustering," *Phys. Rev.*, vol. D71, p. 063001, 2005, astro-ph/0412626.

[40] U. Seljak, "Extracting primordial non-gaussianity without cosmic variance," *Phys. Rev. Lett.*, vol. 102, p. 021302, 2009, 0807.1770.

[41] L. Verde and A. F. Heavens, "On the trispectrum as a Gaussian test for cosmology," *Astrophys. J.*, vol. 553, p. 14, 2001, astro-ph/0101143.

[42] D. Bertolini, K. Schutz, M. P. Solon, and K. M. Zurek, "The Trispectrum in the Effective Field Theory of Large Scale Structure," 2016, 1604.01770.

[43] A. Cooray, "21-cm Background Anisotropies Can Discern Primordial Non-Gaussianity," *Phys. Rev. Lett.*, vol. 97, p. 261301, 2006, astro-ph/0610257.

[44] K. Heitmann, M. White, C. Wagner, S. Habib, and D. Higdon, "The Coyote Universe I: Precision Determination of the Nonlinear Matter Power Spectrum," *Astrophys. J.*, vol. 715, pp. 104–121, 2010, 0812.1052.

[45] A. Schneider, R. Teyssier, D. Potter, J. Stadel, J. Onions, D. S. Reed, R. E. Smith, V. Springel, F. R. Pearce, and R. Scoccimarro, "Matter power spectrum and the challenge of percent accuracy," *JCAP*, vol. 1604, no. 04, p. 047, 2016, 1503.05920.