Abstract. We study the Likelihood function of data given $f_{NL}$ for the so-called local type of non-Gaussianity. In this case the curvature perturbation is a non-linear function, local in real space, of a Gaussian random field. We compute the Cramer–Rao bound for $f_{NL}$ and show that for small values of $f_{NL}$ the three-point function estimator saturates the bound and is equivalent to calculating the full Likelihood of the data. However, for sufficiently large $f_{NL}$, the naive three-point function estimator has a much larger variance than previously thought. In the limit in which the departure from Gaussianity is detected with high confidence, error bars on $f_{NL}$ only decrease as $1/\ln N_{\text{pix}}$ rather than $N_{\text{pix}}^{-1/2}$ as the size of the data set increases. We identify the physical origin of this behaviour and explain why it only affects the local type of non-Gaussianity, where the contribution of the first multipoles is always relevant. We find a simple improvement to the three-point function estimator that makes the square root of its variance decrease as $N_{\text{pix}}^{-1/2}$ even for large $f_{NL}$, asymptotically approaching the Cramer–Rao bound. We show that using the modified estimator is practically equivalent to computing the full Likelihood of $f_{NL}$ given the data. Thus other statistics of the data, such as the four-point function and Minkowski functionals, contain no additional information on $f_{NL}$. In particular, we explicitly show that the recent claims about the relevance of the four-point function are not correct. By direct inspection of the Likelihood, we show that the data do not contain enough information for...
any statistic to be able to constrain higher order terms in the relation between the Gaussian field and the curvature perturbation, unless these are orders of magnitude larger than the size suggested by the current limits on $f_{NL}$. As our main focus is the scaling with $N_{\text{pix}}$ of the various quantities, calculations are done in flat sky approximation and without the radiation transfer function.

**Keywords:** CMBR experiments, CMBR theory, cosmological perturbation theory

**ArXiv ePrint:** astro-ph/0606001

### Contents

1. **Introduction** 2

2. **Subtleties of the $f_{NL}$ expansion** 7
   2.1. Notation .................................. 7
   2.2. Previous results: the missing enhanced terms .......... 8
   2.3. The relevance of the additional terms .................. 10

3. **Likelihood calculation** 11
   3.1. Full Likelihood and Cramer–Rao bound: leading terms .... 11
   3.2. The expansion of the Likelihood to second order is consistent 13
   3.3. Higher order terms in the relation between $\Phi$ and $g$ are negligible 14

4. **No need to worry: simple estimators can saturate the Cramer–Rao bound** 15
   4.1. Explicit origin of the increased variance ................ 15
   4.2. Improved estimator ................................ 16
   4.3. The improved normalization only depends on the large scales 17
   4.4. Relation to the full Likelihood calculation ............. 18

5. **Comments on estimating $f_{NL}$ using the four-point function** 18
   5.1. The variance of the proposed four-point estimator has enhanced terms ... 19
   5.2. Is there additional information in the four-point function? ...... 20
   5.3. Explicit relation between the three-point and four-point estimators .... 21

6. **Summary** 21

**Acknowledgments** 22

**Appendix A. Proof that the enhancement is at most of order $N_{\text{pix}}$** 22

**Appendix B. Higher order corrections in the definition of local non-Gaussianities** 24

**References** 25

### 1. Introduction

In single-field slow-roll inflation the level of non-Gaussianity is sharply predicted and very small, less than $10^{-6}$ [1, 2]. This is quite far from the present experimental sensitivity and probably not attainable with either CMB observations or galaxy redshift surveys. As a
result, deviations from a purely Gaussian statistics of density perturbations, if observed, could provide important constraints on models of early cosmology, forcing us to abandon the single-field slow-roll paradigm.

Of course there are many ways in which a signal could be ‘non-Gaussian’. Given a data set, such as the WMAP maps, there are two possible ways to proceed. One could calculate all kinds of statistics of the data and compare the results with the expectation for a Gaussian field searching for anomalies. This is a fine strategy as long as one adjusts the significance of the result to account for the number of possible deviations that have been explored. There are several anomalies in the WMAP data reported in the literature that have been found in this way (see for example [3]–[5]). Unfortunately their significance is hard to assess and as a result one is not sure how seriously to take them.

The second approach is to think about the possible physical mechanisms that can lead to non-Gaussianities and search for their particular signatures. In the context of primordial effects one should investigate what types of non-Gaussianity can plausibly be produced in various inflationary models. This approach is further bolstered by the fact that at least at the level of the three-point function, primordial signals seem to fall into two definite classes. Thus there are only two different signatures one has to look for.

The analysis of inflationary models that go beyond the single-field slow-roll class has identified several examples with a relatively high level of non-Gaussianity, within reach of present or forthcoming experiments. For nearly Gaussian fluctuations, the quantity most sensitive to departures from perfect Gaussianity is the three-point correlation function. In general, each model will give a different correlation between the Newtonian potential modes:

\[
\langle \Phi(k_1)\Phi(k_2)\Phi(k_3) \rangle = (2\pi)^3\delta^3(k_1 + k_2 + k_3)F(k_1, k_2, k_3).
\]  

The function \(F\) describes the correlation as a function of the triangle shape in momentum space.

The predictions for the function \(F\) in different models divide quite sharply into two qualitatively different classes as a consequence of qualitatively different ways of producing correlations among modes [8]. The first possibility is that the source of density perturbations is not the inflaton but a second light scalar field \(\sigma\). In this case non-Gaussianities are generated by the non-linear relation between the fluctuation \(\delta\sigma\) of this field and the final perturbation \(\Phi\) we observe. This non-linearity is local as it acts when the modes are much outside the horizon; schematically we have \(\Phi(x) = g(x) + f_{NL}(g^2(x) - \langle g^2 \rangle) + \cdots\), where \(g\) is a Gaussian random field. The quadratic piece introduces a three-point function for \(\Phi\) of the form

\[
F(k_1, k_2, k_3) = f_{NL} \cdot 2\Delta^2 \cdot \left( \frac{1}{k_1^3k_2^3} + \frac{1}{k_1^3k_3^3} + \frac{1}{k_2^3k_3^3} \right),
\]

where \(\Delta\) is the power spectrum normalization, \(\langle \Phi(k_1)\Phi(k_2) \rangle = (2\pi)^3\delta^3(k_1 + k_2)\Delta\cdot k_1^{-3}\), which has been taken as exactly scale invariant. Examples of this mechanism are the curvaton scenario [9] and the variable decay width model [10], which naturally give rise

\[\text{Even with perfectly Gaussian primordial fluctuations, the observables, e.g. the temperature anisotropy, will not be perfectly Gaussian as a consequence of the non-linear relation between primordial perturbations and what we will eventually observe. These effects are usually of order } 10^{-5} \text{ (see for example [6, 7]) and thus beyond (but not much) present sensitivity. In the following we will disregard these contributions.}\]
to $f_{NL}$ greater than 10 and 5, respectively. Various subtleties in estimating the size of this type of non-Gaussianity will be the focus of this paper.

The second class of models are single-field models with a non-minimal Lagrangian, where the correlation among modes is created by higher derivative operators [11]–[15]. In this case, the correlation is strong among modes with comparable wavelength and it decays when we take one of $k$s to zero keeping the other two fixed. Although different models of this class give a different function $F$, all these functions are qualitatively very similar. We will call this kind of functions equilateral because the signal is maximal for equilateral configurations in Fourier space, whereas for the local form (2) the most relevant configurations are the squeezed triangles with one side much smaller than the others. We will not discuss the equilateral type of non-Gaussianity in this paper too much. We will just point out that the effects studied in this paper do not apply in that case so that the situation is much simpler.

The strongest constraint on $f_{NL}$ comes from analysing the three-point function of the WMAP data set. WMAP is the best available data set because it has the largest number of pixels measured with good signal to noise. From the first-year data the constraint is [16]:

$$-27 < f_{NL} < 121 \quad \text{at 95\% C.L.} \quad (3)$$

This constraint is better than that obtained by the WMAP collaboration both using the one-year WMAP data [17] and the three-year ones [18]. This is so because the WMAP team used a non-optimal estimator which did not adequately treat the effect of anisotropic noise, as already noted in [17,19]. In [16], we showed that the effect of the anisotropic noise can be substantially reduced with the addition of a linear piece to the estimator. Always in [16] we also constrained the level of the equilateral three-point function. For both types, the departures from Gaussianity still allowed by the data are at the same level.

Given the interest in constraining the level of non-Gaussianity, one may wonder if a statistic other than the three-point function might extract more information about $f_{NL}$. There are various contradictory, or at least apparently contradictory, answers to this question in the literature. On the one hand in [20] and [16] it is argued that the three-point function saturates the Cramer–Rao bound up to terms of order $f_{NL} A^{1/2}$, where $A$ is the square of the amplitude of curvature perturbations: $A^{1/2} \sim 10^{-5}$. On the other hand calculations of the signal to noise in the four-point function by [21] and [22] point to a different conclusion. These papers claim that, even though in the limit of $f_{NL} \to 0$ the signal to noise ratio of the four-point function is negligible, it grows more rapidly with the number of pixels in the data set than for the three-point function. As a result for values of $f_{NL}$ rather small, say around $\sim 50$ for an experiment like Planck, the signal to noise in the four-point function is larger than for the three-point function and stronger constraints on $f_{NL}$ could be placed by studying the four-point function. Of course this result is puzzling. One is immediately drawn to the question, what about the five-point function? And why not the eleven-point function? Applying the same arguments as in [21] and [22] would show that the signal to noise ratio becomes larger the higher the $n$-point function considered. Clearly there is a contradiction.

It is the aim of this paper to clarify this contradiction. We will show that both calculations have missed an interesting subtlety of the local type of non-Gaussianity in
Estimators for local non-Gaussianities

the case of scale invariant, or nearly scale invariant, spectrum of primordial perturbations. As a result, the calculation of the noise of various estimators (including the three-point function) for finite \( f_{NL} \) is missing some relevant term. Some of the terms that are naively down by powers of \( f_{NL} A_{1/2} \) are actually much larger, being enhanced by \( N_{\text{pix}} \). The growth in the signal to noise for high \( f_{NL} \) seen in the above papers is fictitious. We will show that the same subtlety creeps into the calculation of \([20]\) and thus the three-point function estimator considered there also does not saturate the Cramer–Rao bound for large \( f_{NL} \). We want to stress that even though what was missed was a rather subtle point, it has potentially large consequences on the signal to noise of the estimators previously considered. For example when one is in the regime of large signal to noise, the error bars on \( f_{NL} \) from the three-point function decrease as \( 1/\ln N_{\text{pix}} \) rather than \( N_{\text{pix}}^{-1/2} \). The reader at this point should not panic, we will show that the Cramer–Rao bound in this regime still scales as \( N_{\text{pix}}^{-1/2} \) and that it is rather straightforward to extract all of this information from the data either by calculating the full Likelihood or slightly tweaking the three-point function estimator.

What is the missing subtlety? To understand it, it is best to recall what is the main effect of the local non-Gaussianity: it correlates large and small scales. In the three-point function, a long wavelength mode modulates the amplitude of all the short wavelengths by the same amount, regardless of the wavelength of the short mode. Furthermore, in \([8]\) we showed that most of the signal in the three-point function is coming from squeezed triangular configuration in Fourier space. More importantly for this discussion, if one considers the signal to noise as a function of the wavenumber of the long wavelength mode \( k_L \), one gets an equal amount of information from every logarithmic interval in \( k_L \). This in fact is the source of the problem. The smallest \( k_L \) in the survey are by definition the ones with the largest cosmic variance as there are the least of them in the survey. As one increases the resolution of the survey the contribution to the signal to noise from the long wavelengths only decreases logarithmically and thus the large cosmic variance of the long modes translates into large variances in the estimators of \( f_{NL} \).

In fact in \([16]\) indications of the importance of the long wavelength modes were emphasized in the context of the effect of anisotropic noise on the estimator of the three-point function. The noise in the map is anisotropic because WMAP spent different amounts of time observing each pixel on the sky. As a result the level of small scale power, for large multipoles where the noise becomes important, varies across the sky. This map of small scale power can randomly align with the particular large scale mode giving a spurious \( f_{NL} \) signal. Of course on average this effect is zero as there is no intrinsic correlation between the map of observing time per pixel and the large scale temperature. However for a particular realization, some modes will be correlated (spurious positive \( f_{NL} \)) and others anti-correlated (spurious negative \( f_{NL} \)). The contribution to the signal from the long wavelength modes will not add exactly to zero, as we have few of them in the survey. The random left over spurious signal effectively increases the variance of the estimator. This effect was noted in the WMAP team analysis \([17,19]\), where the constraint on \( f_{NL} \) got worse as they increased the size of the data set by including more of the small scales. In \([16]\) the estimator was improved by including a linear piece which substantially reduces the effect allowing us to get better constraints. This effect is the reason why the three-year data analysis by the WMAP team \([18]\) did not appreciably improve the limits.
To clarify the situation we will study the full Likelihood of the data given $f_{NL}$. We will keep careful track of enhanced terms and thus do a consistent expansion in $f_{NL}$. We will then calculate the Cramer–Rao bound, extending the results of [20] to non-zero $f_{NL}$. We will show how the additional terms in the variance of the three-point function estimator make it become sub-optimal. This can be easily fixed using an improved estimator which asymptotically saturates the Cramer–Rao bound. The use of this estimator is equivalent to the full Likelihood of the data.

The fact that the improved three-point function estimator is equivalent to the full Likelihood of the data, implies that there is no additional information in the four-point function. We will also show this explicitly, illustrating how at best the four-point function is equivalent to the three-point function. No other statistic such as Minkowski functionals, various wavelet based statistics and other esoteric constructions are worth trying to constrain $f_{NL}$. None can be better than the three-point function. This is true up to corrections of order $f_{NL}A^{1/2} \lesssim 10^{-3}$.

The apparent large signal to noise in the four-point function led to the suggestion [21,22] that the four-point function could even be sensitive to higher order terms in the relation between $\Phi$ and $\delta\sigma$: $\Phi(x) = g(x) + f_{NL}(g^2(x) - \langle g^2 \rangle) + f_{NL}^2 \alpha g^3(x) \ldots$ The claim was that the four-point function could constrain the real parameter $\alpha$. Of course the third term is a minuscule correction to the first two, even for the largest allowed values of $f_{NL}$. Thus it is difficult to understand how one could be sensitive to it. Again the missing terms in the variance of the estimator were responsible for the apparent sensitivity to $\alpha$. Using the full Likelihood we will show that for any realistic experiment there is in fact not enough information about $\alpha$ in the data to constrain it, unless $\alpha > 1/(f_{NL}A^{1/2}) \gtrsim 10^3$.

Finally we will also show that for a realistic experiment where $\ln N_{\text{pix}}$ is large, the value of $f_{NL}$ for which improving the naive three-point function estimator is important is rather large. One should start worrying about it once there is a many $\sigma$ detection of $f_{NL}$. As a result our improved estimator will probably be only of academic interest. Our paper mainly provides clarification of various misconceptions in the literature. Given this and to reduce the length of our equations, we will work in the flat sky approximation and neglect the CMB transfer functions, directly working with a two-dimensional random field with a local non-Gaussianity. Expressions for the three- and four-point functions including the radiation transfer function and with spherical geometry can be found for instance in [22,23].

We want to stress\(^6\) that our approximation captures the qualitative features of the real problem, like the dependence of the various expressions on the number of data $N_{\text{pix}}$. On the other hand one should not trust the numerical factors that we will find, because they would be changed in a complete treatment, where projection effects from three to two dimensions are taken into account together with the full radiation transfer function. The reason why the qualitative features are captured by our approximation is that, although a given mode on the sky receives contribution from a range of different 3D wavelengths, this effect is limited to an interval $\Delta k \sim k$. This implies that a squeezed configuration of the three-point function of primordial perturbations $k_1 \ll k_2, k_3$ (which, as we will see, are the only configurations material to our conclusions) maps to a correlation among

---

\(^6\) We thank the unknown referees for correspondence about this point.
different multipoles with $l_1 \ll l_2, l_3$. In other words although the three-point function of the two-dimensional temperature map is not exactly of the local form, it still roughly behaves as $\langle a_{l_1} a_{l_2} a_{l_3} \rangle \propto l_1^{-2} l_2^{-2} + \text{perm.}$, which is the 2D analogue of equation (2), with additional modulations induced by the transfer function. As we will discuss, our results just depend on the $l_1^{-2}$ dependence for $l_1 \to 0$ (compared for instance with $l_1^0$ which is typical of other shapes of non-Gaussianity), so that it is enough to stick to an exactly local non-Gaussianity in two dimensions. Also the flat sky approximation will contribute to change the numerical factors, but leave unaltered the qualitative features.

We also point out that even though we do all our calculations in two dimensions, as relevant to the CMB, our conclusions are equally valid for three-dimensional surveys, such as galaxy surveys or future 21 cm observations.

2. Subtleties of the $f_{NL}$ expansion

The aim of this section is to explicitly show the presence of terms in the variance of the three-point function estimator of $f_{NL}$ that are enhanced by factors of $N_{\text{pix}}$ and thus contribute significantly even though they are naively suppressed by $f_{NL} A^{1/2}$. We start by introducing our notation and reproducing previous calculations of the variance of the estimator. We then identify the terms that had been previously missed and give a rule of thumb to easily determine when they are important. We will show that in practice the enhanced terms do not correct current upper limits and that they will only become important after a very high signal to noise detection of $f_{NL}$.

2.1. Notation

We work in the flat sky approximation, neglect the transfer function, and assume that the error is dominated by cosmic variance. In this paper we are mostly interested in the scaling properties of the estimators for the non-Gaussianities used for example in [16, 17], [20]–[22], [24, 25]. These properties are not modified by these approximations, while on the other hand their use makes the presence of some physical effects much clearer, as we will later see. Let us briefly set up our conventions. For the Fourier transform we have:

$$\Phi_{l} = \frac{\Omega}{N_{\text{pix}}} \sum_{\theta} e^{-i \theta \cdot \hat{l}} \Phi_{\theta},$$

where $\Omega$ and $N_{\text{pix}}$ are respectively the angular size and the number of pixels of the sky survey. It is immediate to obtain from this the continuum limit $\Phi_{l} = (\Omega/N_{\text{pix}}) \sum_{\theta} e^{-i \theta \cdot \hat{l}} \Phi_{\theta} \simeq \int d^2 \theta e^{-i \theta \cdot \hat{l}} \Phi_{\theta}$. We also have:

$$\Phi_{\theta} = \frac{1}{\Omega} \sum_{l} e^{-i \theta \cdot \hat{l}} \Phi_{l} \simeq \int \frac{d^2 l}{(2\pi)^2} e^{-i \theta \cdot \hat{l}} \Phi_{l},$$

and the useful relations:

$$\sum_{l} e^{-i \theta \cdot (\hat{l}_1 - \hat{l}_2)} = N_{\text{pix}} \delta_{\theta, \hat{l}_1 - \hat{l}_2}, \quad \sum_{\theta} e^{-i \theta \cdot (\hat{l}_1 - \hat{l}_2)} = N_{\text{pix}} \delta_{\theta, \hat{l}_1 - \hat{l}_2}. $$

---

7 This is confirmed by the fact we are able to recover the same scaling properties found in [20]–[22], [24, 25] where these approximations were not used.
Estimators for local non-Gaussianities

We are only interested in local non-Gaussianities as the effect we will discuss does not apply to other types. In that case, the observed field $\Phi_{\vec{\theta}}$ is given by a non-linear function of a Gaussian field $g_{\vec{\theta}}$ which is local in real space:

$$
\Phi_{\vec{\theta}} = f(g_{\vec{\theta}}) = g_{\vec{\theta}} + f_{NL}(g_{\vec{\theta}}^2 - \sigma^2).
$$

(7)

We will call this field temperature although our results apply to other measures, not just the CMB temperature. In Fourier space the local relation reads:

$$
\Phi_{\vec{l}} = (f(g))_{\vec{l}} = g_{\vec{l}} + f_{NL}((g \circ g)_{\vec{l}} - \sigma^2 \Omega \delta_{l,0}),
$$

(8)

where $g = \sum_{\vec{k}} g_{\vec{k}} g_{\vec{\theta}}$. We will explicitly address later the case of possible higher order corrections in $f_{NL}$ to these definitions. The covariance matrix is defined as

$$
\langle g_{\vec{l}1} g_{\vec{l}2} \rangle = C_{l1, l2} = C_l \Omega \delta_{l1, l2},
$$

(9)

where $l^2 = \vec{l} \cdot \vec{l}$, and $C_l = 2 \pi A / l^2$, which then implies that

$$
\sigma^2 = \langle g_{\vec{\theta}1} g_{\vec{\theta}2} \rangle = {2 \pi A \over \Omega} \sum_l {1 \over l^2} \simeq A \over 2 \ln N_{\text{pix}},
$$

(10)

where in the last passage we have used the continuum limit, and the fact that $N_{\text{pix}} \simeq \Omega l_{\text{max}}^2 / (4 \pi)$, with $l_{\text{max}}$ the maximum of the observed $l$s. From here on, in order to simplify the notation, we will remove the vector symbol from $\vec{l}$ and $\vec{\theta}$ in all the mathematical expressions when the meaning and the distinction from the modulus $l = |\vec{l}|$ and $\theta = |\vec{\theta}|$ is clear from the context.

2.2. Previous results: the missing enhanced terms

In [16, 17] the analysis for the non-Gaussianities of the local kind was performed using a trilinear estimator with signal to noise weighting. In the limit of flat sky, unit transfer function, and isotropic noise it reduces to:

$$
\mathcal{E} = \frac{1}{N} \sum_l \frac{1}{\Omega C_l} \Phi_l \chi_{l-\vec{l}},
$$

(11)

where we have defined the field $\chi_l = (\Phi \circ \Phi)_l - \Omega \sigma^2 \delta_{l,0}$, and where we consider only non-degenerate configurations with all the $\Phi$s taken with $l \neq 0$. The normalization

$$
N = \sum_l \frac{1}{\Omega C_l} \langle \Phi_l \chi_{l-1} \rangle_1 \simeq 8 N_{\text{pix}} \sigma^2,
$$

(12)

with the subscript 1 meaning that the expectation value is taken with $f_{NL} = 1$, has been chosen so that the estimator is unbiased, $\langle \mathcal{E} \rangle = f_{NL}$. The definition in equation (7) tells us that the temperature field in the sky $\Phi_{\vec{\theta}}$ is to a good approximation a Gaussian field, with a small non-Gaussian correction of order $f_{NL} \Phi \sim f_{NL} A^{1/2} < 10^{-3}$. So one is tempted

8 Some of the expressions, like equation (12) above, when expressed in terms of $N_{\text{pix}}$ and $\sigma$, will slightly depend on the geometry of the survey which changes the boundary of the domain of integration in Fourier space. Also we will have similar corrections going from flat to full sky. However these effects do not change significantly our results.
to expect that higher order corrections in $f_{\text{NL}}$ in the various expressions are suppressed with respect to the leading terms by powers of $f_{\text{NL}}A^{1/2}$ and thus irrelevant. For example, the variance of the estimator in equation (11) starts with a piece which is of zeroth order in $f_{\text{NL}}$, and which gives:

$$\langle \Delta E^2 \rangle_{f_{\text{NL}}=0} = \frac{1}{\Omega^2 N^2} \sum_{l \ell'} \frac{1}{C_l C_{l'}} \langle \Phi_l \chi_{-l} \Phi_{l'} \chi_{-l'} \rangle_{f_{\text{NL}}=0} = \frac{1}{N} = \frac{1}{8 N_{\text{pix}} \sigma^2}, \quad (13)$$

where $\Delta E = E - \langle E \rangle$. One would naively assumes that this is the dominant term in the variance, with small corrections of order $f_{\text{NL}}A^{1/2}$, which should contribute at most at order $10^{-3}$ given the current bound on $f_{\text{NL}}$. This is what was assumed for example in [16, 17, 20], where the estimator (11) was in fact found minimizing the variance at zeroth order in $f_{\text{NL}}$ among all trilinear estimators. However, as we will soon see, for the case of local non-Gaussianities, and only for them, there is another parameter which enters into the expansion: $N_{\text{pix}}$. We will see in fact that in certain expressions such as the variance of the estimator above for example, there are terms that although suppressed by powers of $f_{\text{NL}}A^{1/2}$, are enhanced by powers of $N_{\text{pix}}$, and so, depending on the real value of $f_{\text{NL}}$, they might need to be taken into account.

In order to verify that this is actually the case, and to understand the implications of this fact, let us sketch the computation of the variance of the estimator $E$ keeping higher order terms in $f_{\text{NL}}$. The variance of $E$ will involve the computation of a six-point function, which will split in the sum of the product of several different combinations of connected $n$-point functions, i.e. the product of three two-point functions, of two three-point functions, and of a four-point function and a two-point function. Concentrating on the last kind of contribution, we will have terms like:

$$\langle \Delta E^2 \rangle \supset \frac{1}{N^2 \Omega^4} \sum_{l_1 l_2 l_1 l_2} \frac{1}{C_{l_1} C_{l_2}} \langle \Phi_{l_1} \Phi_{l_2} \Phi_{-l_1 -l_2} \Phi_{l_1} \Phi_{l_2} \Phi_{-l_1 -l_2} \rangle_c, \quad (14)$$

where $\langle \Phi_{l_1} \Phi_{l_2} \cdots \Phi_{n} \rangle_c$ stays for the connected $n$-point function. Now, apart from numerical factors, one of the terms in the expansion of the connected four-point function reads:

$$\langle \Phi_{l_2} \Phi_{-l_1 -l_2} \Phi_{-l_1 -l_2} \rangle_c \supset f_{\text{NL}}^2 \Omega^4 \delta_{l_1, -l_1} C_{l_2} C_{l_1} C_{l_2}. \quad (15)$$

Considering the effect of this term in the variance in equation (14), where we also take the two-point function at zeroth order in $f_{\text{NL}}$, we obtain:

$$\langle \Delta E^2 \rangle \supset \frac{f_{\text{NL}}^2}{N^2} \sum_{l_1 l_2 l_2} C_{l_1}^2 \propto \frac{f_{\text{NL}}^2 A_{\text{pix}}^2}{N^2} \propto \frac{f_{\text{NL}}^2}{\ln^2 N_{\text{pix}}}. \quad (16)$$

Thus this contribution to the relative variance does not decrease as $1/(N_{\text{pix}} \ln N_{\text{pix}})$, as one would have naively expected, but there is an enhancement of $N_{\text{pix}}$ which make it decrease only as $1/\ln^2 N_{\text{pix}}$.

There is a very physical reason for the presence of such enhanced terms for the case of local non-Gaussianities. As we have already said, most of the signal for this kind of

---

9 Notice that the variance of the estimator scales faster than the naive $1/N_{\text{pix}}$ by a factor of $\ln N_{\text{pix}}$. This behaviour is typical of non-Gaussianities of the local kind, where the signal comes from the correlation of the modes of all different scales.
non-Gaussianities comes from squeezed configurations, where one of the $l$s is small and the others are large. More precisely, for the three-point function, the signal to noise from all the squeezed configurations with the smallest $l$ in a given decade is roughly the same for every decade, although there are much fewer modes in a decade of low $l$. This means that the low $l$ modes are always very important for the estimator $E$. Now, the point is that there is an intrinsic variance associated with a configuration with a certain small $l$, simply because there are very few of those small $l$s, just $2l + 1$; and this is unaffected by the fact that $N_{\text{pix}}$ of the survey increases, because this just increases the $l_{\text{max}}$ of the experiment. Therefore the relative variance of the estimator due to these terms decrease only logarithmically with $N_{\text{pix}}$, because this is how the relative importance of the configurations with small $l$s decreases with $N_{\text{pix}}$. This physical explanation guarantees us that these enhanced terms are not present in the case of equilateral non-Gaussianities, where the importance of the small $l$s and of the squeezed configurations is marginal.

It is useful to develop a quick thumb rule to understand if a term which is a sum of product of different $C_l$s is enhanced or not: a term will be enhanced only if at least one $C_l$ is raised to a power larger than one. In fact in this case the summation over the multipoles for this term will be dominated by the lowest $l$s, so that some $l_{\text{min}}$ will appear in the denominator. This makes these terms enhanced by powers of $l_{\text{max}}/l_{\text{min}}$ (see also appendix A).

### 2.3. The relevance of the additional terms

Our discussion shows that the treatment of expressions containing $f_{\text{NL}}$ is delicate in the case of local non-Gaussianities. The expansion parameter is not just $f_{\text{NL}}A^{1/2}$, but there are terms which can parametrically go as $f_{\text{NL}}A^{1/2}N_{\text{pix}}$, and therefore cannot be neglected. We need to understand the relevance of these terms both for existing limits on $f_{\text{NL}}$ as well as their impact on future measurements.

After a careful calculation, we find the following expression for the variance of $E$:

$$\langle \Delta E^2 \rangle = \frac{1}{4AN_{\text{pix}} \ln N_{\text{pix}}} \left( 1 + \frac{8f_{\text{NL}}^2 A N_{\text{pix}}}{\pi \ln N_{\text{pix}}} + \cdots \right),$$

where $\cdots$ represents terms suppressed by powers of $f_{\text{NL}}^2 A$ without any further $N_{\text{pix}}$ enhancement. This result shows an important feature of this estimator. Imagine that we have a series of experiments with increasing $N_{\text{pix}}$, and that at some point we detect a non-null $f_{\text{NL}}$. Then, at first the variance will decrease as $1/(N_{\text{pix}} \ln N_{\text{pix}})$, but, after a critical $N_{\text{pix}}$ which depends on the actual value of $f_{\text{NL}}$, and which is basically, apart for logarithms, when the signal to noise is of order 1, the variance will begin to decrease very slowly as $1/\ln^2 N_{\text{pix}}$, because of the enhanced variance of the term proportional to $f_{\text{NL}}^2$. In the analysis performed in [16, 17] the variance for a non-zero $f_{\text{NL}}$ was assumed to be the same as for $f_{\text{NL}} = 0$, expecting that the $f_{\text{NL}}$ corrections would have been small. In the light of the results of this section, we see that this procedure is not always justified. However, for those analysis, we can verify that the error introduced is very small, as already numerically checked with non-Gaussian Monte Carlo simulations in [17, 23]. We can quantify the error this way: the relative correction to the variance for an $f_{\text{NL}}$ at $n \sigma_0$ from the origin, where $\sigma_0$ is the variance computed at $f_{\text{NL}} = 0$, is of order $2n^2 / (\pi \ln^2 N_{\text{pix}})$. For the WMAP experiment $\ln^2 N_{\text{pix}} \sim 35$; therefore this correction is large for $n$ larger...
than \( \sim 6,7 \). Therefore, if we wish to give a 2\( \sigma \) confidence interval around a certain central value, we see that the enhanced terms will become important for a central value around 4, 5\( \sigma \) far from the origin, i.e. in the case of a clear detection of a non-zero \( f_{\text{NL}} \). In the analysis of [16,17], the central value of \( f_{\text{NL}} \) is of the order of only one \( \sigma \) far from the origin. Because of this, the approximation done in [16,17] of considering for a non-zero \( f_{\text{NL}} \) the variance at \( f_{\text{NL}} = 0 \) is numerically justified, with a small error at the per cent level, well beneath the error coming from other sources, for example from the uncertainty in the cosmological parameters, which gives an error on the variance of order ten per cent [16].

Summarizing, we conclude that the enhanced terms will not be important until there is a clear detection of a non-zero \( f_{\text{NL}} \). If that happens, they will have to be taken into account. At that point, the variance of the estimator \( E \) will begin to decrease as \( 1/\ln N_{\text{pix}} \). Given the very slow convergence of \( E \) in this regime, one is lead to wonder whether a better estimator exists.

3. Likelihood calculation

For non-Gaussianities of the local type, it is easy to calculate the full Likelihood for \( f_{\text{NL}} \) given the data and determine to what extent the data are able to constrain \( f_{\text{NL}} \). This is true even in the high signal to noise limit where the previous estimator has an increased variance.

With the full Likelihood it is possible to determine what is the minimum variance that an estimator of \( f_{\text{NL}} \) can have, the so called Cramer–Rao bound. The bound on the variance is \( \langle \partial^2 L/\partial f_{\text{NL}}^2 \rangle^{-1} \), where \( L \) is minus the logarithm of the Likelihood. In [20] it was proved that the estimator \( E \) of the former section, whose variance scales as \( 1/(N_{\text{pix}} \ln N_{\text{pix}}) \), satisfies this bound at order zero in \( f_{\text{NL}} A^{1/2} \). However, we have just learned that this expansion in powers of \( f_{\text{NL}} A^{1/2} \) breaks down when \( f_{\text{NL}} \) is detected because of the presence of enhanced terms. It is therefore worth asking what happens to the Cramer–Rao bound in the same regime, and check if there are enhanced term also in this case.

By the end of this section, we will see that the Likelihood allows for an expansion in powers of \( f_{\text{NL}}^2 A \), without \( N_{\text{pix}} \) enhancements, and therefore that the Cramer–Rao bound in the presence of a non-null \( f_{\text{NL}} \) is affected only marginally by terms suppressed by powers of \( f_{\text{NL}}^2 A \). This will tell us that the estimator \( E \) of the previous section is just a bad estimator in the large signal to noise regime, and that in principle there can be estimators whose variance in this regime scales as \( 1/(N_{\text{pix}} \ln N_{\text{pix}}) \).

3.1. Full Likelihood and Cramer–Rao bound: leading terms

The Likelihood function can be simply obtained inverting equation (7), and expressing the probability for the Gaussian variables \( g \) as a function of the temperature field \( \Phi \):

\[
g_\theta = f^{-1}(\Phi_\theta) = \Phi_\theta - \tilde{f}_{\text{NL}} (\Phi_\theta^2 - \sigma^2) + 2 \tilde{f}_{\text{NL}}^2 \Phi_\theta (\Phi_\theta^2 - \sigma^2) + \cdots.
\]

The Likelihood will be a function of the parameter \( \tilde{f}_{\text{NL}} \), while we keep \( f_{\text{NL}} \) to denote the true value of the non-Gaussianity parameter. The dots represent higher order terms in
\( \tilde{f}_{NL} \) coming from the inversion of the function \( f(g) \), which for the moment we neglect. We will come back to them shortly. In Fourier space, expression (18) translates into:

\[
\tilde{g}_l = (f^{-1}(\Phi))_l = \Phi_l - \tilde{f}_{NL} \left( (\Phi \circ \Phi)_l - \sigma^2 \delta_{l,0} \right) + 2f_{NL}^2 \left( (\Phi \circ \Phi \circ \Phi)_l - \sigma^2 \Phi_l \right) + \cdots .
\]  

(19)

Starting from minus the logarithm of the probability

\[
\mathcal{L}_g = \frac{1}{2} \sum_{l_1 l_2} C_{l_1 l_2}^{-1} g_{l_1} g_{l_2},
\]

(20)

we change variable from \( g_l \) to \( \Phi_l \) taking into account the change in the measure:

\[
\mathcal{L} = \frac{1}{2} \sum_{l_1 l_2} C_{l_1 l_2}^{-1} (f^{-1}(\Phi))_{l_1} (f^{-1}(\Phi))_{l_2} - \text{Tr} \ln(J),
\]

(21)

where \( \text{Tr} \) stays for trace in Fourier space, and \( J \) is the Jacobian

\[
J = \frac{\partial (f^{-1}(\Phi))_{l_1}}{\partial \Phi_{l_2}}.
\]

(22)

We can now expand to second order in \( \tilde{f}_{NL} \) to obtain:

\[
\mathcal{L} = \frac{1}{2} \sum_l \left( \frac{1}{\Omega C_l} \left( \Phi_l \Phi_{-l} - 2\tilde{f}_{NL} \chi_l \Phi_{-l} + \tilde{f}_{NL}^2 \left( \chi_l \chi_{-l} + 4\Phi_l \eta_{-l} \right) \right) \right)
\]

\[
+ 2\tilde{f}_{NL} \frac{N_{\text{pix}}}{\Omega} \Phi_{l=0} - 4f_{NL}^2 \frac{N_{\text{pix}}}{\Omega} \chi_{l=0} - 2\tilde{f}_{NL}^2 N_{\text{pix}} \sigma^2,
\]

(23)

where we have introduced the field \( \eta_l = (\chi \circ \Phi)_l \).

\(^{10}\) Although the Likelihood contains all the information on the parameter \( f_{NL} \) one can derive from an experiment, its computation as a function of \( \tilde{f}_{NL} \) can be very challenging in practice. All analysis to date have used an estimator of \( f_{NL} \) rather than to calculate the full Likelihood (e.g. \([16, 17]\)).

The statistical properties of \( \mathcal{L} \) depend on the underlying true value \( f_{NL} \). To make this explicit we can write

\[
\Phi_l = g_l + f_{NL} \left( (g \circ g)_l - \Omega \sigma^2 \delta_{l,0} \right).
\]

(24)

Plugging back in the expression for the Likelihood, we obtain:

\[
\mathcal{L} = \frac{1}{2} \sum_l \frac{1}{\Omega C_l} \left( g_l g_{-l} + 2(f_{NL} - \tilde{f}_{NL})(\tilde{\chi}_l g_{-l} - 2f_{NL}(g_l \eta_{-l} - \langle g_l \eta_{-l} \rangle)) \right)
\]

\[
+ (f_{NL} - \tilde{f}_{NL})^2 (\tilde{\chi}_l \tilde{\chi}_{-l} + 2g_l \eta_{-l} + 2(g_l \eta_{-l} - \langle g_l \eta_{-l} \rangle))
\]

\[
+ 2\tilde{f}_{NL} \frac{N_{\text{pix}}}{\Omega} \Phi_{l=0} - 4f_{NL}^2 \frac{N_{\text{pix}}}{\Omega} \chi_{l=0} - 2f_{NL}^2 N_{\text{pix}} \sigma^2,
\]

(25)

where we have analogously defined \( \tilde{\chi}_l = (g \circ g)_l - \Omega \sigma^2 \delta_{l,0} \), and \( \tilde{\eta}_l = (\tilde{\chi} \circ \Phi)_l \).

\(^{10}\) Notice that in this section, to keep the formulae as simple as possible, we have assumed that the average of \( \Phi \) in the patch of the sky survey is observed. As it will become clear later, even if this was not the case, it would not change relevantly the results.
We can use the expression of the Likelihood we have just derived to find the Cramer–Rao bound on the variance of an unbiased estimator for \( f_{\text{NL}} \):

\[
\left\langle \frac{\partial^2 \mathcal{L}}{\partial f_{\text{NL}}^2} \right\rangle^{-1} = \left( \sum_l \frac{1}{\Omega C_l} \langle \tilde{\chi}_l \tilde{\chi}_{-l} + 2g_l \tilde{\eta}_{-l} \rangle \right)^{-1} = \frac{1}{8N_{\text{pix}} \sigma^2},
\]

(26)

where we have used that:

\[
\sum_l \frac{1}{\Omega C_l} \langle g_l \tilde{\eta}_{-l} \rangle = \sum_l \frac{1}{\Omega C_l} \left\langle \frac{g_l}{\Omega} \sum_{l'} g_{-l'-l} \left( \sum_{l''} g_{l''-l''} - \Omega \sigma^2 \delta_{l',0} \right) \right\rangle = \sum_{l' l''} \frac{2}{\Omega^2 C_l} \Omega C_l \delta_{l',-l'-l''} \Omega C_{l''} \delta_{l',-l'-l''} = 2 \frac{N_{\text{pix}}}{\Omega} \sum_l C_l = 2N_{\text{pix}} \sigma^2,
\]

(27)

and analogously:

\[
\sum_l \frac{1}{\Omega C_l} \langle \tilde{\chi}_l \tilde{\chi}_{-l} \rangle = 4N_{\text{pix}} \sigma^2.
\]

(28)

We see from equation (13) that the estimator of the last section saturates the Cramer–Rao bound for sufficiently small \( f_{\text{NL}} \).

### 3.2. The expansion of the Likelihood to second order is consistent

At this point, one may wonder if the higher order terms in \( f_{\text{NL}} \) might relevantly alter this result with terms that, though suppressed by powers of \( f_{\text{NL}} \), are enhanced by factors of \( N_{\text{pix}} \), as in the former section for the variance of the estimator \( \mathcal{E} \). It is quite straightforward to check that this is not the case. For example, at quartic level in \( f_{\text{NL}} \) there are terms like:

\[
\mathcal{L} \supset f_{\text{NL}}^4 \sum_l \frac{1}{\Omega C_l} (g \circ g \circ g)_l (g \circ g \circ g)_{-l} = f_{\text{NL}}^4 \frac{1}{\Omega^4} \sum_{l_1 l_2 l_3 l_4} g_{l_1} g_{l_2} g_{l_3} g_{l_4} \frac{1}{C_l} \Omega^2 C_{l_1} C_{l_2} C_{l_3} C_{l_4} \sim f_{\text{NL}}^4 \sigma^4 N_{\text{pix}}.
\]

(29)

whose expectation value contributes to the Cramer–Rao bound in equation (26) with terms like:

\[
f_{\text{NL}}^2 \sum_{l_1 l_2 l_3} \frac{1}{\Omega^2} \frac{C_{l_1} C_{l_2} C_{l_3} C_{-l_1 -l_2 -l_3}}{C_l} = f_{\text{NL}}^2 \sum_{l_1 l_2 l_3} \frac{1}{\Omega^2} \frac{C_{l_1} C_{l_2} C_{l_3}}{C_{l_1 + l_2 + l_3}} \sim f_{\text{NL}}^2 \sigma^4 N_{\text{pix}}.
\]

(30)

We recognize this term as not being enhanced also thanks to our thumb rule according to which a term is not enhanced if there are no \( C_l \)'s raised to a power larger than one. We conclude that for these terms there is no \( N_{\text{pix}} \) enhancement, and therefore are suppressed by genuine powers of \( f_{\text{NL}}^2 \) with respect to the leading terms in the Cramer–Rao bound in equation (26). Higher order terms will appear in even powers of \( f_{\text{NL}} \), and will give similar contributions of the form:

\[
f_{\text{NL}}^{2n-2} \sum_{l_1 l_2 \ldots l_{n+1}} \frac{1}{\Omega^n} \frac{C_{l_1} C_{l_2} \ldots C_{l_{n+1}}}{C_{l_1 + l_2 + \ldots + l_{n+1}}} \sim f_{\text{NL}}^{2n-2} \sigma^{2n} N_{\text{pix}},
\]

(31)
so that we see there is no $N_{\text{pix}}$ enhancement for all these terms. At quartic level in $f_{\text{NL}}$ there are also terms appearing from the expansion of the Jacobian in the Likelihood, as for example:

$$f_{\text{NL}}^4 \frac{N_{\text{pix}}}{\Omega^4} \sum_{l_1,l_2,l_3} g_{l_1} g_{l_2} g_{l_3} g_{-l_1-l_2-l_3}, \quad (32)$$

whose expectation value gives subleading not-enhanced terms to the Cramer–Rao bound of the form:

$$f_{\text{NL}}^2 \frac{N_{\text{pix}}}{\Omega^2} \sum_{l_1,l_2} C_{l_1} C_{l_2} = f_{\text{NL}}^2 \frac{N_{\text{pix}}}{\Omega^2} \left( \sum_l C_l \right)^2 \propto f_{\text{NL}}^2 \sigma^4 N_{\text{pix}}, \quad (33)$$

and the same conclusion applies unaltered to the higher order terms coming from the expansion of the Jacobian:

$$f_{\text{NL}}^{2n-2} \frac{N_{\text{pix}}}{\Omega^n} \sum_{l_1,l_2} \cdots C_{l_1} C_{l_2} \cdots C_{l_{2n-1}} = f_{\text{NL}}^{2n-2} \frac{N_{\text{pix}}}{\Omega^n} \left( \sum_l C_l \right)^n \propto f_{\text{NL}}^{2n-2} \sigma^{2n} N_{\text{pix}}. \quad (34)$$

We conclude that these higher order terms, being not enhanced by $N_{\text{pix}}$, do not alter significantly the Cramer–Rao bound in equation (26).

In order to verify the consistency of the expansion at quadratic order in $f_{\text{NL}}$ for the Likelihood as well, we need to check that the higher order terms are irrelevant not only on average, but also on each realization. In appendix A, we show that their variance scales at most as $N_{\text{pix}}^2$, which makes their contribution to the Likelihood suppressed by powers of $f_{\text{NL}} A^{1/2}$ with respect to the contribution of the quadratic terms. We conclude that the expansion of the Likelihood up to quadratic order is consistent.

### 3.3. Higher order terms in the relation between $\Phi$ and $g$ are negligible

Since in this section we have been very careful in keeping track of higher order terms in $f_{\text{NL}}$, it is useful to comment on the possibility that additional contributions come from the presence of higher order terms in the relation between $\Phi$ and the underlying Gaussian field of equation (7):

$$\Phi_\theta = g_\theta + f_{\text{NL}} (g_\theta^2 - \sigma^2) + \alpha f_{\text{NL}}^2 g_\theta (g_\theta^2 - \sigma^2) + \cdots \quad (35)$$

with $\alpha$ an unknown real parameter. Physically we expect these corrections to be there with $\alpha$ of order 1. We need the Likelihood at second order in $f_{\text{NL}}$, so that one might worry that our results now depend on $\alpha$. This would be very strange as physically these third-order terms in the expansion above are a very small correction to the, already small, second-order terms. The data should not be sensitive at all to $\alpha$. One can check that this is indeed what happens. We leave the details of the algebra to appendix B. We will find that, although $\alpha$ enters into the Likelihood at order $f_{\text{NL}}^2$, terms containing $\alpha$ cancel on each realization up to terms suppressed by $1/N_{\text{pix}}$. Therefore there is no sensitivity to $\alpha$ and it can safely be set to zero.

In summary, we have written the Likelihood for $f_{\text{NL}}$ up to second order in $f_{\text{NL}}$, proving that this expansion is consistent, with the higher order terms only giving negligible contributions suppressed by powers of $f_{\text{NL}} A^{1/2}$. This has also allowed us to verify that
Estimators for local non-Gaussianities

higher order terms in $f_{NL}$ in the Likelihood do not give rise to enhanced contributions to the Cramer–Rao bound. Thus we conclude that the result in equation (26) is only corrected by terms of order $f_{NL}A^{1/2}$, without $N_{\text{pix}}$ enhancement. After what observed in the previous section, this was not a priori guaranteed.

4. No need to worry: simple estimators can saturate the Cramer–Rao bound

Now that we have determined the Cramer–Rao bound, in this section we look for a new estimator which continues to have a variance close to the bound even in the high signal to noise regime. We do this to further understand the origin of the enhanced terms and point out how a simple change in the estimator based on our intuitive understanding can make the estimator saturate the bound.

We will start from the original estimator $E$ in section 2, and we will explicitly show the way in which the enhanced terms cause the slow convergence of the estimator in the large signal to noise regime. After this it will become easy to guess a new estimator that, apart from small corrections, saturates the Cramer–Rao bound when the enhanced terms are large.

4.1. Explicit origin of the increased variance

Let us therefore start from the estimator $E$ in equation (11) and express it in terms of Gaussian variables as:

$$E = \frac{1}{N\Omega} \sum_l \frac{1}{C_l} \left( g_l \tilde{\chi}_{-l} + f_{NL} (\tilde{\chi}_l \tilde{\chi}_{-l} + 2g_l \tilde{\eta}_{-l}) \right) \equiv E_0 + f_{NL} E_1 + \cdots \quad (36)$$

where $+ \cdots$ represent higher order terms in $f_{NL}$, and where we have defined:

$$E_0 = \frac{1}{N\Omega} \sum_l \frac{1}{C_l} g_l \tilde{\chi}_{-l}, \quad (37)$$

$$E_1 = \frac{1}{N\Omega} \sum_l \frac{1}{C_l} (\tilde{\chi}_l \tilde{\chi}_{-l} + 2g_l \tilde{\eta}_{-l}). \quad (38)$$

Notice that $\langle E_0 E_1 \rangle = 0$, $\langle E_0 \rangle = 0$, and $\langle E_1 \rangle = 1$. Therefore the variance of the estimator can be written as:

$$\langle \Delta E^2 \rangle = \langle E_0^2 \rangle + f_{NL}^2 \langle \Delta E_1^2 \rangle. \quad (39)$$

As we discussed, the variance at zeroth order in $f_{NL}$ is

$$\langle \Delta E_0^2 \rangle_{f_{NL}=0} = \langle E_0^2 \rangle = \frac{1}{8N_{\text{pix}} \sigma^2} \quad (40)$$

which saturates the Cramer–Rao bound in (26). However, as we noted in section 2, the variance of $E_1$ behaves like:

$$\langle \Delta E_1^2 \rangle \sim \frac{2}{\pi \ln^2 N_{\text{pix}}}, \quad (41)$$
Estimators for local non-Gaussianities

decreasing only logarithmically. Therefore it is going to dominate the variance of the estimator for $N_{\text{pix}}/\ln N_{\text{pix}} \gtrsim 1/(f_{NL}^2 A)$. Apart for the logarithm, this is when the signal to noise becomes of order 1.

The enhanced variance of $\mathcal{E}_1$ could have been anticipated. This term contains, at leading order in $f_{NL}$, all the signal of non-Gaussianity. We know that for the local kind of non-Gaussianity the signal comes from squeezed configurations with two of the three $l$s very large, and one of the $l$s very small. The contribution to $\mathcal{E}_1$ of squeezed configurations with the smallest of the $l$s within a certain decade is roughly independent of the decade, although there are very few low $l$ multipoles. The value of $\mathcal{E}_1$ on a given realization depends quite strongly on the particular value of the few lowest multipoles; this explains why it converges to its average depends quite strongly on the particular value of the lowest multipoles goes away. However this happens only logarithmically in $N_{\text{pix}}$ as this is the way in which the contribution from the lowest multipoles decays.

4.2. Improved estimator

Now that we understand better the problem of the estimator $\mathcal{E}$, it is easy to find an improved estimator for the large signal to noise regime. We can think about the large variance of $\mathcal{E}$ as coming from a ‘wrong normalization’. Although the estimator is clearly unbiased its value strongly depends on the amplitude of the low $l$ modes, so that if on a particular realization we have a small amplitude in the first multipoles, the value of the estimator will be small and vice versa. This effect cancels on average (that is why the estimator is unbiased) but it is the source of the large variance. Anyway this effect can clearly be corrected as we surely know the amplitude of the low $l$ modes in each particular realization: we just have to divide by a ‘realization dependent’ normalization. We define a new estimator $\tilde{\mathcal{E}}$

$$\tilde{\mathcal{E}} = \frac{N\Omega}{\sum_l (1/C_l) (\chi l^{\chi-l} + 2 \Phi_l \eta^{\eta-l})} \mathcal{E} = \frac{\sum_l (1/C_l) \Phi_l \chi^{\chi-l}}{\sum_l (1/C_l) (\chi l^{\chi-l} + 2 \Phi_l \eta^{\eta-l})} = f_{NL} + \frac{\mathcal{E}_0}{\mathcal{E}_1} + \cdots \quad (42)$$

where in the second line we have expressed everything in terms of Gaussian variables. Neglected terms are suppressed with respect to the ones we kept by genuine powers of $f_{NL} A^{1/2}$. Neglecting these terms, the new estimator $\tilde{\mathcal{E}}$ is unbiased: $\langle \tilde{\mathcal{E}} \rangle = f_{NL}$, as $\mathcal{E}_0/\mathcal{E}_1$ is an odd function of the Gaussian variables $g$ and it has thus zero average.

We can now verify that the new estimator converges to the Cramer–Rao bound. We can write $\mathcal{E}_1 = 1 + \delta \mathcal{E}_1$, where $\delta \mathcal{E}_1$ is of the order of $\langle \delta \mathcal{E}_1^2 \rangle^{1/2} \sim 2^{1/2}/(\pi^{1/2} \ln N_{\text{pix}})$. For large $N_{\text{pix}}$ we can thus expand the denominator

$$\tilde{\mathcal{E}} \simeq f_{NL} + \mathcal{E}_0 - \mathcal{E}_0 \delta \mathcal{E}_1. \quad (43)$$

The variance introduced by the third piece scales like $1/(N_{\text{pix}} \ln^3 N_{\text{pix}})$, more rapidly than the Cramer–Rao bound $\propto 1/(N_{\text{pix}} \ln N_{\text{pix}})$. After a while we are therefore left with the

---

11 As shown in appendix $A$, terms in $\mathcal{E}$ of higher order in $f_{NL}$, even including a possible contribution from terms proportional to $\alpha$, contribute to the variance of the estimator with terms which are not enhanced by powers of $N_{\text{pix}}$ more than what $\mathcal{E}_1$ already is. Therefore they are suppressed with respect to the contribution of $\mathcal{E}_1$ by a genuine power of $f_{NL} A^{1/2}$.
variance of $E_0$ that, as we know, satisfies the Cramer–Rao bound. It is worth noticing that already at the level of the WMAP experiment $\ln^2 N_{\text{pix}} \simeq 35$, so the deviation of this estimator from the Cramer–Rao bound is already rather small. The important point is that this good behaviour of $\tilde{E}$ is not spoiled when we enter in the large signal to noise regime.

4.3. The improved normalization only depends on the large scales

Our understanding of the enhanced variance of the estimator $E$ relies on the fact that always a significant fraction of the signal is coming from low $l$ modes with a great intrinsic variance. If this is true, it better be that the solution to this problem depends strongly on the low $l$s. Here we therefore verify that $E_1$ can be written to good approximation in terms of just the first few modes. In appendix B we have shown that

$$S_2 = \sum_l \frac{1}{\Omega C_l} g_l \bar{\eta}_{-l}$$  \hspace{1cm} (44)

is fully correlated with the quantity

$$S_1 = \frac{N_{\text{pix}}}{\Omega^2} \sum_l g_l g_{-l},$$  \hspace{1cm} (45)

up to corrections $\mathcal{O}(1/N_{\text{pix}})$, so that on each realization $\Delta S_2 = 3\Delta S_1$ with very good accuracy. In the same fashion one can prove that also the quantity

$$S_3 = \sum_l \frac{1}{\Omega C_l} \bar{\chi}_l \bar{\chi}_{-l}$$  \hspace{1cm} (46)

is fully correlated to $S_1$ (up to corrections $\mathcal{O}(1/N_{\text{pix}})$) and that $\Delta S_3 = 4\Delta S_1$. This implies that also $E_1$ is fully correlated with $S_1$ and therefore, on each realization, we can write:

$$E_1 = \langle E_1 \rangle + \frac{10\Delta S_1}{N}.$$  \hspace{1cm} (47)

Now, the important point is that in order to compute the quantity $\Delta S_1$ on a given realization one needs, to good approximation, only the first few modes. This can be seen from the computation of the variance of $S_1$ in appendix B:

$$\langle \Delta S_1^2 \rangle = 2 \frac{N_{\text{pix}}^2}{\Omega^2} \sum_l C_l^2 \simeq 2\pi A^2 N_{\text{pix}}^2 \frac{\Omega}{\int_{l_{\text{min}}}^{l_{\text{max}}} \frac{1}{l^2} \, dl} = 2\pi A^2 N_{\text{pix}}^2 \frac{\Omega}{\int_{l_{\text{min}}}^{l_{\text{max}}} \frac{1}{l^2} \, dl}$$  \hspace{1cm} (48)

which shows that the contribution to the variance of the high $l$s is completely irrelevant. We therefore conclude that the value of $E_1$ on each realization can be determined just by looking at the first few modes, in agreement with our intuition.

This last remark has also relevant consequences from the computational point of view. In fact it seems at first sight very hard to use the new estimator in the analysis of CMB data, as it contains four-point functions and one has to deal with the complications of the spherical geometry and of the transfer function. On the other hand the dependence on only the first few modes makes the modification computationally quite light.
4.4. Relation to the full Likelihood calculation

The Likelihood (23) contains all the information on the parameter \( f_{\text{NL}} \). To reconstruct it from the data we just need the coefficients of the terms linear and quadratic in \( \tilde{f}_{\text{NL}} \). These two combinations of the data are sufficient statistics for \( f_{\text{NL}} \). Notice that given our discussion above it is not so complicated to analyse the full Likelihood function: we just need the same kind of terms entering in the estimator \( \tilde{E} \) above. A natural question is whether one can get better constraints on \( f_{\text{NL}} \) using the full Likelihood function instead of an estimator.

First of all it is straightforward to check that the maximum Likelihood estimator, which can be easily derived from equation (23), has the same good properties of our improved estimator discussed above. It is unbiased up to corrections \( \mathcal{O}(1/N_{\text{pix}}^{1/2}) \) and it asymptotically saturates the Cramer–Rao bound up to corrections decaying as \( 1/\ln^2 N_{\text{pix}} \).

The Cramer–Rao bound, being the average value of the second derivative of the log-Likelihood, gives the average value of error bars that one gets. The difference in using the full Likelihood is that the curvature of it changes realization by realization, as it is given by the \( \tilde{f}_{\text{NL}}^2 \) term in equation (23). Usually this distinction between the curvature of the Likelihood in a particular realization and its average value is irrelevant, as the difference scales like \( 1/N_{\text{pix}} \). This is not true in our case. The variance of the curvature of the Likelihood function only scales as \( 1/\ln^2 N_{\text{pix}} \). This is again intuitive: given the strong dependence on the lowest multipoles, a realization with an excess of power in the low \( l \)s compared with the average will be more constraining than one with suppressed power on large scales. In the first case in fact it is easier to see the non-Gaussian correlation between the low \( l \)s and the short scale power. This difference is anyway not that large: for real experiments that have a chance of detecting \( f_{\text{NL}} \), \( 1/\ln N_{\text{pix}} \) is rather small.

We reach an important conclusion. The use of our improved estimator, or equivalently the maximum Likelihood one, is equivalent to the full Likelihood of the data up to small corrections suppressed by \( 1/\ln N_{\text{pix}} \). This closes the door to any additional attempt to improve the limits on \( f_{\text{NL}} \).

5. Comments on estimating \( f_{\text{NL}} \) using the four-point function

It has been proposed in [21] and more recently in [22] that an estimator for \( f_{\text{NL}} \) based on the four-point function has a variance which decreases as \( 1/N_{\text{pix}}^2 \), so that it might be better than a three-point function estimator for large enough \( f_{\text{NL}} \). Of course our analysis in the previous sections shows that the Cramer–Rao bound scales as \( 1/(N_{\text{pix}} \ln N_{\text{pix}}) \), so that no estimator can do better than this. Moreover we proved that a slight modification of the three-point function estimator makes the new estimator \( \tilde{E} \) approach asymptotically the Cramer–Rao bound. Rather than stop here and rely on the above ‘theorems’ we want to show explicitly in this section what goes wrong in the naive calculation of the variance of the four-point function estimator. We will see that the \( 1/N_{\text{pix}}^2 \) scaling of the variance does not hold once the signal to noise is larger than one and the ‘enhanced’ terms are taken into account. We will also show explicitly that there is no additional information about \( f_{\text{NL}} \) in the four-point function that is not already captured by the three-point function.
5.1. The variance of the proposed four-point estimator has enhanced terms

Let us begin proving that the variance of the estimator introduced in [21, 22] does not scale as $1/N_{\text{pix}}^2$. The proposed estimator, analogously to the estimator $E$ in the three-point function case, is the linear combination of four-point correlators which maximizes the signal to noise in the limit $f_{\text{NL}} \to 0$:

$$
E_4 = \frac{1}{N_4} \sum_{l_1l_2l_3} \frac{\langle \Phi_{l_1} \Phi_{l_2} \Phi_{l_3} \Phi_{l_4} \rangle_{c,1}}{C_{l_1} C_{l_2} C_{l_3} C_{l_4}} \Phi_{l_1} \Phi_{l_2} \Phi_{l_3} \Phi_{l_4}
$$

(49)

where $N_4$ is a normalization constant which makes the estimator unbiased $\langle E_4 \rangle = f_{\text{NL}}^4$. The sum is restricted to momentum conserving, $l_4 = -l_1 - l_2 - l_3$, non-degenerate quadrilaterals. The subscript 1 in $\langle \Phi_{l_1} \ldots \Phi_{l_4} \rangle_{c,1}$ means that the connected four-point function is evaluated with $f_{\text{NL}} = 1$. The variance of this estimator is:

$$
\langle \Delta E_4^2 \rangle = \frac{1}{N_4} \sum_{l_1l_2l_3} \frac{\langle \Phi_{l_1} \Phi_{l_2} \Phi_{l_3} \Phi_{l_4} \rangle_{c,1}^2}{C_{l_1} C_{l_2} C_{l_3} C_{l_4}} - f_{\text{NL}}^4.
$$

(50)

Here we are interested in the scaling with $N_{\text{pix}}$ of the different terms; therefore we do not keep track of the various combinatorial and numerical factors, and also of possible logarithmic corrections. Using the fact that the connected four-point function behaves, at leading order in $f_{\text{NL}}$, like:

$$
\langle \Phi_{l_1} \Phi_{l_2} \Phi_{l_3} \Phi_{l_4} \rangle_c \sim f_{\text{NL}}^2 \cdot C_{l_1} C_{l_1+l_2} C_{l_4} + \text{symm},
$$

(51)

we find that the denominator of the first term contains terms that behave like:

$$
\sum_{l_1l_2l_3} \frac{\langle \Phi_{l_1} \Phi_{l_2} \Phi_{l_3} \Phi_{l_4} \rangle_{c,1}^2}{C_{l_1} C_{l_2} C_{l_3} C_{l_4}} \sim \sum_{l_1l_2l_3} \frac{C_{l_1}^2 C_{l_2}^2 C_{l_4}^2}{C_{l_1} C_{l_2} C_{l_3} C_{l_4}} \sim \sum_{l_1l_2l_3} \frac{C_{l_1}^2 C_{l_1+l_2} C_{l_4}}{C_{l_2} C_{l_3}}.
$$

(52)

We already met these kind of summations and we know they are dominated by $l_1 + l_2 \sim l_{\text{min}}$ and enhanced by a factor of $N_{\text{pix}}$ with respect to the naive scaling:

$$
\sum_{l_1l_2l_3} \frac{\langle \Phi_{l_1} \Phi_{l_2} \Phi_{l_3} \Phi_{l_4} \rangle_{c,1}^2}{C_{l_1} C_{l_2} C_{l_3} C_{l_4}} \sim N_{\text{pix}}^2 \sum_l C_l^2 \propto N_{\text{pix}}^2. A^2.
$$

(53)

For the numerator we have to compute the eight-point function, which in general will be the sum of the product of four two-point function, of two connected three-point functions and one two-point function, of two connected four-point functions, and so on. At zeroth order in $f_{\text{NL}}$, we have only two-point functions and we obtain something which scales as the square root of the denominator. Therefore, if ones stops at this level, one finds that the variance decreases as $1/N_{\text{pix}}^2$. This is the result obtained in [21] and [22].

However, as we learned in section 2, there are other terms in the numerator which, though suppressed by powers of $f_{\text{NL}}^2 A$, are enhanced by powers of $N_{\text{pix}}$. It turns out that the product of two four-point functions, of the five-point with the three-point one, and of the six-point with the two-point one, all give rise to enhanced terms with the same
scaling. For example a term with two connected four-point functions is

\[ \sum_{l_1 l_2 l_3} \sum_{\tilde{l}_1 \tilde{l}_2 \tilde{l}_3} C_{l_1} C_{l_2} C_{l_3} C_{l_4} \langle \Phi_{l_1} \Phi_{l_2} \Phi_{l_3} \Phi_{l_4} \rangle_c \langle \tilde{\Phi}_{\tilde{l}_1} \tilde{\Phi}_{\tilde{l}_2} \tilde{\Phi}_{\tilde{l}_3} \tilde{\Phi}_{\tilde{l}_4} \rangle_c \]

\[ \simeq f_{NL}^4 \left( \sum_{l_1 l_2 l_3} \sum_{\tilde{l}_1 \tilde{l}_2 \tilde{l}_3} \frac{C_{l_1} C_{l_2} C_{l_3}^2}{C_{l_2} C_{l_3}} + \frac{C_{l_1} C_{l_2}^2}{C_{l_2} C_{l_3}} + \cdots \right) \] (54)

where \( + \cdots \) represents terms of higher order in \( f_{NL} \). The sums will be dominated by the region with \( l_1 + l_2 \sim l_{\text{min}} \) and \( \tilde{l}_1 + \tilde{l}_2 \sim l_{\text{min}} \), so that we obtain

\[ f_{NL}^4 N_{\text{pix}}^4 \left( \sum_l C_l^2 \right)^2 \propto f_{NL}^4 A^4 N_{\text{pix}}^4 \] (55)

as we wanted to show. Putting together the behaviour of the different terms we get the scaling of the variance of the proposed estimator up to logarithmic corrections

\[ \langle \Delta E_4^2 \rangle \sim \frac{1 + f_{NL}^4 A^2 N_{\text{pix}}^2}{A^2 N_{\text{pix}}^2} \] (56)

where we see the importance of enhanced terms at numerator. For comparison, it is useful to write the same schematic relation for the analogous three-point function estimator \( E \) we discussed in section 2:

\[ \langle \Delta E^2 \rangle \sim \frac{1 + f_{NL}^2 A N_{\text{pix}}}{A N_{\text{pix}}} \] (57)

We notice that the enhanced terms become parametrically important for both the estimators when \( f_{NL}^2 A N_{\text{pix}} \sim 1 \) which is, apart for logarithms, the regime when the signal to noise is of order 1.

### 5.2. Is there additional information in the four-point function?

In the limit in which the enhanced terms are negligible, we see that the variance of \( E_4 \) is of the order of the square of the variance of \( E \), which means that \( E_4 \) will give parametrically the same limits on \( f_{NL} \) in this regime. However, in the previous sections, we have shown that in the same limit the three-point function estimator \( E \) saturates the Cramer–Rao bound; therefore in this regime there is nothing which could be added by the use of the four-point function. This agrees with the numerical result of [22], where it is shown that in this regime the limit on \( f_{NL} \) obtained from the four-point function estimator is always slightly worse than the one obtained using the three-point function. Given the Cramer–Rao bound, we can even say something more: in this regime no improvement can be achieved from combining the two estimators.

On the other hand when the enhanced terms become important we see that the variance of both estimators does not decrease anymore (apart for logarithmic terms). In particular there no \( 1/N_{\text{pix}}^2 \) scaling for the four-point function. As it was shown in section 4, in this regime one must consider ‘fractional’ estimators, which are not just polynomial in the data.
5.3. Explicit relation between the three-point and four-point estimators

It is worth pointing out an explicit relationship between the four-point and three-point function estimators, to show that there is really nothing new in the four-point estimator $E_4$, which is not already taken into account using $E$.

Let us remind once again that, for local non-Gaussianities, the signal to noise of the three-point function is concentrated on squeezed configurations, where one of the three $l$s is small, and the other two are large and almost opposite. The three-point function estimator $E$ is basically doing a weighted sum of the signal contained in all the configurations, where the weight is the signal to noise ratio. Therefore, the estimator $E$ can be well approximated by a sum over just the squeezed configurations:

$$E \propto \sum_L \Phi_L \sum_l \frac{\Phi_{-L-l}\Phi_l}{C_l} \equiv \sum_L \Phi_L K_{-L}$$

where $L$ is a large scale multipole and $l$ is a small scale one, and $K_L$ is defined as:

$$K_L = \sum_l \frac{\Phi_{-L-l}\Phi_l}{C_l}.$$ 

(59)

In the presence of a non-zero $f_{NL}$, $K_L$ will contain, when expressed in terms of Gaussian variables a contribution

$$K_L \sim 2f_{NL}g_L.$$ 

(60)

The estimator $E$ correlates this contribution with the long wavelength mode $\Phi_{-L}$.

Analogously in the case of the estimator $E_4$, as we have seen, the signal comes from squeezed configurations, where the four vectors $l$s are approximately opposite in pairs. Therefore, the estimator $E_4$ can be written to a good approximation as a sum over just these configurations:

$$E_4 \propto \sum_L C_L K_L K_{-L}.$$ 

(61)

Again in the presence of a non-zero $f_{NL}$ we are correlating the non-Gaussian contribution inside each of the $K$s, giving an average signal $\propto f_{NL}^2$.

From this it should be clear that the two estimators are clearly not independent and that the four-point one is less efficient because the non-Gaussian contribution must come out of both the $K$s, while it is more efficient to directly correlate $K_L$ with the mode $\Phi_{-L}$ as in equation (58).

6. Summary

It is perhaps unfortunate that our paper is filled with so many equations, the message however is simple. The analysis of the local type of non-Gaussianity for scale invariant perturbations is somewhat more subtle than one might have guessed: a naive $f_{NL}A^{1/2}$ expansion is not always appropriate. The physical origin of the effect is clear: long wavelength modes modulate the amplitude of the short wavelengths and the amplitude of this modulation produced by long wavelengths of every decade in scale is the same. Cosmic variance severely affects this long wavelengths and because their relative information
Estimators for local non-Gaussianities

contribution only decreases logarithmically with the number of pixels, one ends up with large variances for the naive $f_{NL}$ estimators.

The basic point is that when one calculates the normalization of the $f_{NL}$ estimator one uses the average level of large scale fluctuations as opposed to the power in the individual realization one happens to have. As a result, in the limit of large signal to noise, this relatively large uncertainty in the normalization of the estimator severely enhances its variance. Fortunately one knows the amplitude of the modes in a given realization by direct measurement so it is almost trivial to fix the problem by choosing a normalization that depends on the particular realization.

Writing down the full Likelihood one can explicitly calculate how well one should in principle be able to constrain $f_{NL}$ and explicitly check how the effect mentioned above comes in. One can show that a simple modification of the naive estimator recovers all the information that the data contain and that in fact using that estimator is basically equivalent to calculating the full Likelihood, up to corrections $O(1/\ln N_{\text{pix}})$. As a result, one is also convinced that other statistics such as the four-point function, Minkowski functionals, wavelets, etc can at best extract as much information on $f_{NL}$ as the three-point function. In any event they would not contribute additional information on $f_{NL}$, so once the three-point function is measured there is nothing else to be done.

Acknowledgments

We thank Daniel Babich and Eiichiro Komatsu for useful comments. LS is supported in part by funds provided by the US Department of Energy (DOE) under cooperative research agreement DF-FC02-94ER40818. MZ is supported by the Packard and Sloan foundations, NSF AST-0506556 and NASA NNG05GG84G.

Appendix A. Proof that the enhancement is at most of order $N_{\text{pix}}$

In this appendix we want to prove that the variance of the sums that appear in the Likelihood and in the estimators scales at most as $N_{\text{pix}}^2$, i.e. that the possible enhancement with respect to the naive scaling is at most of order $N_{\text{pix}}$. In order to do this, following the discussion in section 2, where we explained the thumb rule for discovering enhanced terms, it is enough to show that in the expression in Fourier space of the variance there is at most one $c_l$ raised at most to the power of two. This corresponds to an enhancement of one factor of $N_{\text{pix}}$, while further enhancements would require either one $c_l$ raised to a power larger than two, or more than one $c_l$ squared.

A good rearrangement of the various terms is obtained if we start in real space, where the terms we are interested in can be written in the general form:

$$
\sum_{\theta_1, \theta_2} C_{\theta_1, \theta_2}^{-1} \bar{g}_{\theta_1}^M \bar{g}_{\theta_2}^N,
$$

(A.1)

where $M$ and $N$ are two positive integers, and $C_{\theta_1, \theta_2}$ is the covariance matrix in real space. This is related to the one in Fourier space by the following relation:

$$
C_{\theta_1, \theta_2} = \frac{1}{\Omega} \sum_l C_l e^{-il(\theta_1 - \theta_2)}.
$$

(A.2)
Analogously, $C^{-1}$ can be expressed in real space as:

$$C^{-1}_{\theta_1\theta_2} = \frac{\Omega}{N_{\text{pix}}^2} \sum_l \frac{1}{C_l} e^{-\imath l \cdot (\theta_1 - \theta_2)}. \quad (A.3)$$

Let us compute the variance of the general term in equation (A.1). This will be the sum of terms of the form:

$$\sum_{\theta_1\theta_2\theta_3\theta_4} C_{\theta_1\theta_2}^{-1} C_{\theta_3\theta_4}^{-1} (C_{\theta_1\theta_2})^\alpha (C_{\theta_3\theta_4})^\beta (C_{\theta_2\theta_3})^\gamma (C_{\theta_2\theta_3})^\delta (C_{\theta_1\theta_2})^\Sigma (C_{\theta_3\theta_4})^\rho, \quad (A.4)$$

with the positive integers $\alpha, \beta, \gamma, \delta, \Sigma, \rho$ constrained to satisfy $\alpha + \beta + \gamma + \delta + \Sigma + \rho = (N + M)/2$.

We can now express each of the $C$s and $C^{-1}$s in Fourier space with the relations (A.2) and (A.3). After this, the summation over the angles $\theta_1, \theta_2, \theta_3$ and $\theta_4$ becomes trivial, each of these giving a Kronecker delta. In particular, the summations over $\theta_1$ and $\theta_2$ give the constraints:

$$l_1 + l_\alpha + \cdots + l_\alpha + l_\beta + \cdots + l_\beta + l_\gamma + \cdots + l_\gamma = 0,$$

$$-l_1 - l_\alpha - \cdots - l_\alpha + l_\beta + \cdots + l_\beta + l_\Sigma = 0 \quad (A.5)$$

where $l_1$ is the $l$ associated with the Fourier transform of $C_{\theta_1\theta_2}$, $l_\alpha$, with $i = 1, \ldots, \alpha$, are the $l$s associated with the Fourier transform of $(C_{\theta_1\theta_2})^\alpha$, and analogously for the other $C$s. These two constraints can be usefully rewritten as:

$$l_1 = -l_\alpha - \cdots - l_\alpha - l_\beta - l_\gamma - \cdots - l_\gamma,$$

$$l_\beta + \cdots + l_\beta + l_\gamma + \cdots + l_\beta + l_\Sigma = 0. \quad (A.6)$$

From the summation over $\theta_3$ and $\theta_4$, we obtain two analogous constraints that can be written as:

$$l_2 = l_1 + \cdots + l_\beta + l_\gamma + \cdots + l_\beta + l_\Sigma = 0,$$

$$l_\beta + \cdots + l_\beta + l_\beta + \cdots + l_\beta + l_\Sigma = 0 \quad (A.7)$$

where $l_2$ is the $l$ associated with the Fourier transform of $C_{\theta_3\theta_4}^{-1}$. We see that the second of these constraints is equivalent to the one in equation (A.8). The presence of a redundant constraint is a manifestation of the fact that the term we started with in equation (A.1) was rotationally invariant.

After the summation over the angles, we are left with summations only over the $l$s:

$$\sum_{l_1 l_2} \sum_{l_1^\prime \cdots l_\rho^\prime} \cdots \sum_{l_\alpha l_\beta l_\gamma} \frac{(C_{l_1} \cdots C_{l_\alpha}) \cdots (C_{l_\beta} \cdots C_{l_\rho})}{C_{l_1} C_{l_2}}, \quad (A.11)$$

subject to the three independent constraints we have found. Notice that the $N_{\text{pix}}$ factors in (A.3) exactly cancel with the four ones from the sums over $\theta_i$. Now, the first and third constraints can be used to eliminate the summations over $l_1$ and $l_2$, leaving us with:

$$\sum_{l_1 l_2} \sum_{l_1^\prime \cdots l_\rho^\prime} \frac{(C_{l_1} \cdots C_{l_\alpha}) \cdots (C_{l_\beta} \cdots C_{l_\rho})}{C_{l_1} C_{l_2}}, \quad (A.12)$$
with $l_1$ and $l_2$ given by equations (A.7) and (A.9). The only remaining constraint is equation (A.8). After applying it, the variance is in the form such that we can quickly apply our thumb rule, and understand its level of enhancement. Naively given that the number of $C_l$s is $(N + M)/2 - 2$ and the number of summations is $(N + M)/2 - 1$ we get a behaviour $\sim N_{\text{pix}}$. But it can happen that the last constraint makes two $C_l$s at numerator equal. In this case the sum goes as $N_{\text{pix}}^2$. No further enhancement is possible.

Appendix B. Higher order corrections in the definition of local non-Gaussianities

In Fourier space equation (35) reads:

$$\Phi_l = g_l + f_{\text{NL}} \tilde{\chi}_l + \alpha f_{\text{NL}}^2 \tilde{\eta}_l.$$  

(B.1)

In this appendix we want to prove that, although $\alpha$ enters in the Likelihood at order $f_{\text{NL}}$, this does not imply that data are sensitive to this parameter, unless it is huge compared to the naive estimate $\alpha \sim O(1)$. The new term gives the following contribution to the Likelihood at order $f_{\text{NL}}^2$:

$$L_\alpha = \alpha f_{\text{NL}}^2 \left( \frac{1}{\Omega C_l} (g_l \tilde{\eta}_{l-1} - \langle g_l \tilde{\eta}_{l-1} \rangle) + 3 \frac{N_{\text{pix}}}{\Omega} \tilde{\chi}_{l=0} \right),$$  

(B.2)

where we have neglected terms which are independent of $\tilde{f}_{\text{NL}}$. We notice that both terms above have zero average. However this is not enough to prove that there is no relevant dependence on $\alpha$ because both terms have enhanced variance, so that they converge to zero very slowly. Their importance with respect to the other terms in the Likelihood decreases as $1/\ln^2 N_{\text{pix}}$. What we are now going to prove is that, although both terms have large variance, they are strongly correlated and their contributions in equation (B.2) cancel up to terms suppressed by $1/N_{\text{pix}}$. Therefore the dependence on $\alpha$ is extremely small as expected on physical grounds.

Defining:

$$S_1 = \frac{N_{\text{pix}}}{\Omega^2} \sum_l g_l g_{l-1}, \quad S_2 = \sum_l \frac{1}{\Omega C_l} g_l \tilde{\eta}_{l-1},$$  

(B.3)

we can write equation (B.2) as:

$$L_\alpha = \alpha f_{\text{NL}}^2 (-\Delta S_2 + 3\Delta S_1).$$  

(B.4)

We can now compute the correlation functions of $S_1$ and $S_2$:

$$\langle S_1 \rangle = \frac{N_{\text{pix}}}{\Omega^2} \sum_l \langle g_l g_{l-1} \rangle = \frac{N_{\text{pix}}}{\Omega^2} \sum_l C_l,$$  

(B.5)

$$\langle \Delta S_1^2 \rangle = \frac{N_{\text{pix}}^2}{\Omega^2} \sum_{ll} \langle g_l g_{l-1} g_l g_{l-1} \rangle - \langle S_1 \rangle^2 = 2 \frac{N_{\text{pix}}^2}{\Omega^2} \sum_l C_l^2,$$  

(B.6)

$$\langle S_2 \rangle = 2 \frac{N_{\text{pix}}}{\Omega} \sum_l C_l,$$  

(B.7)
as computed before, and:
\[
\langle \Delta S_2^2 \rangle = \sum_{\ell} \frac{1}{\Omega C_{\ell}} \frac{1}{\Omega C_{\ell}'} \left< g_{\ell} \tilde{\eta}_{-\ell} g_{\ell} \tilde{\eta}_{-\ell} \right> - \langle S_2 \rangle^2
\]
\[
= \sum_{\ell} \frac{1}{\Omega^2 C_{\ell} C_{\ell}'} \left< g_{\ell} \left( \frac{1}{\Omega} \sum_{l''} g_{-l'-l''} \left( \frac{1}{\Omega} \sum_{l'''} g_{l''-l'''-l'''} - \Omega \sigma^2 \delta_{l',0} \right) \right) \right> \times g_{\ell} \left( \frac{1}{\Omega} \sum_{l''} g_{-l'-l''} \left( \frac{1}{\Omega} \sum_{l'''} g_{l''-l'''-l'''} - \Omega \sigma^2 \delta_{l',0} \right) \right) \right> - \langle S_2 \rangle^2. \tag{B.8}
\]

The summation is dominated by those terms which come from the contraction of the \( g_{\ell} \) in each \( S_2 \) with one of the three \( g_{\ell} \)'s contained inside the \( \tilde{\eta}_{-\ell} \) from the same \( S_2 \) term. These are enhanced by a factor of \( N_{\text{pix}} \) with respect to the other contributions. Neglecting subleading terms we obtain:
\[
\langle \Delta S_2^2 \rangle = 18 \frac{N_{\text{pix}}^2}{\Omega^2} \sum_{\ell} C_{\ell}^2. \tag{B.9}
\]

Finally for \( \langle \Delta S_1 \Delta S_2 \rangle = \langle S_1 S_2 \rangle - \langle S_1 \rangle \langle S_2 \rangle \), we obtain:
\[
\langle \Delta S_1 \Delta S_2 \rangle = 6 \frac{N_{\text{pix}}^2}{\Omega^2} \sum_{\ell} C_{\ell}^2, \tag{B.10}
\]
again keeping only leading terms. We see that \( S_1 \) and \( S_2 \) are fully correlated, up to correction \( \mathcal{O}(1/N_{\text{pix}}) \):
\[
\frac{\langle \Delta S_1 \Delta S_2 \rangle}{(\langle \Delta S_1^2 \rangle \langle \Delta S_2^2 \rangle)^{1/2}} = \frac{6(N_{\text{pix}}^2/\Omega^2) \sum_{\ell} C_{\ell}^2}{(2(N_{\text{pix}}^2/\Omega^2) \sum_{\ell} C_{\ell}^2 \times 18(N_{\text{pix}}^2/\Omega^2) \sum_{\ell} C_{\ell}^2)^{1/2}} = 1. \tag{B.11}
\]

On each realization we have:
\[
\Delta S_2 = \left( \frac{\langle \Delta S_2^2 \rangle}{\langle \Delta S_1^2 \rangle} \right)^{1/2} \Delta S_1 = 3 \Delta S_1, \tag{B.12}
\]
and therefore \( L_\alpha = 0 \) up terms suppressed by \( 1/N_{\text{pix}} \).

We conclude that terms which depend on \( \alpha \) are negligible in the Likelihood.

References

[1] Maldacena J, Non-Gaussian features of primordial fluctuations in single field inflationary models, 2003 J. High Energy Phys. JHEP05(2003)013 [SPIRES] [astro-ph/0210603]

[2] Acquaviva V, Bartolo N, Matarrese S and Riotto A, Second-order cosmological perturbations from inflation, 2003 Nucl. Phys. B 667 119 [SPIRES] [astro-ph/0209156]

[3] Land K and Magueijo J, The axis of evil, 2005 Phys. Rev. Lett. 95 071301 [SPIRES] [astro-ph/0502237]

[4] Cruz M, Cayon L, Martinez-Gonzalez E, Vielva P and Jin J, The non-Gaussian cold spot in the 3-year WMAP data, 2006 Preprint astro-ph/0603859

[5] McEwen J D, Hobson M P, Lasenby A N and Mortlock D J, A high-significance detection of non-Gaussianity in the WMAP 3-year data using directional spherical wavelets, 2006 Preprint astro-ph/0604305

[6] Creminelli P and Zaldarriaga M, CMB 3-point functions generated by non-linearities at recombination, 2004 Phys. Rev. D 70 083532 [SPIRES] [astro-ph/0405428]
Estimators for local non-Gaussianities

[7] Bartolo N, Matarrese S and Riotto A, *CMB anisotropies at second order I*, 2006 Preprint astro-ph/0604416 and references therein

[8] Babich D, Creminelli P and Zaldarriaga M, *The shape of non-Gaussianities*, 2004 J. Cosmol. Astropart. Phys. JCAP08(2004)009 [SPIRES] [astro-ph/0405356]

[9] Lyth D H, Uggarelli C and Wands D, *The primordial density perturbation in the curvaton scenario*, 2003 Phys. Rev. D 67 023503 [SPIRES] [astro-ph/0208055]

[10] Zaldarriaga M, *Non-Gaussianities in models with a varying inflaton decay rate*, 2004 Phys. Rev. D 69 043508 [SPIRES] [astro-ph/0306006]

[11] Creminelli P, *On non-Gaussianities in single-field inflation*, 2003 J. Cosmol. Astropart. Phys. JCAP10(2003)003 [SPIRES] [astro-ph/0306122]

[12] Arkani-Hamed N, Creminelli P, Mukohyama S and Zaldarriaga M, *Ghost inflation*, 2004 J. Cosmol. Astropart. Phys. JCAP04(2004)001 [SPIRES] [hep-th/0312100]

[13] Alishahiha M, Silverstein E and Tong D, *DBI in the sky*, 2004 Phys. Rev. D 70 123505 [SPIRES] [hep-th/0404084]

[14] Senatore L, *Tilted ghost inflation*, 2005 Phys. Rev. D 71 043512 [SPIRES] [astro-ph/0406187]

[15] Chen X, Huang M x, Kachru S and Shin G, *Observational signatures and non-Gaussianities of general single field inflation*, 2006 Preprint hep-th/0605045

[16] Creminelli P, Nicolis A, Senatore L, Tegmark M and Zaldarriaga M, *Limits on non-Gaussianities from WMAP data*, 2006 J. Cosmol. Astropart. Phys. JCAP05(2006)004 [SPIRES] [astro-ph/0509029]

[17] Komatsu E et al, *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: tests of Gaussianity*, 2003 Astrophys. J. Suppl. 148 119 [astro-ph/0302223]

[18] Spergel D N et al, *Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology*, 2006 Preprint astro-ph/0603449

[19] Komatsu E, Spergel D N and Wandelt B D, *Measuring primordial non-Gaussianity in the cosmic microwave background*, 2005 Astrophys. J. 634 14 [SPIRES] [astro-ph/0305189]

[20] Babich D, *Optimal estimation of non-Gaussianity*, 2005 Phys. Rev. D 72 043003 [SPIRES] [astro-ph/0503375]

[21] Okamoto T and Hu W, *The angular trispectra of CMB temperature and polarization*, 2002 Phys. Rev. D 66 063008 [SPIRES] [astro-ph/0206155]

[22] Kogo N and Komatsu E, *Angular trispectrum of CMB temperature anisotropy from primordial non-Gaussianity with the full radiation transfer function*, 2006 Phys. Rev. D 73 083007 [SPIRES] [astro-ph/0602099]

[23] Komatsu E, *The pursuit of non-Gaussian fluctuations in the cosmic microwave background*, 2002 Preprint astro-ph/0206039

[24] Komatsu E and Spergel D N, *Acoustic signatures in the primary microwave background bispectrum*, 2001 Phys. Rev. D 63 063002 [SPIRES] [astro-ph/0005036]

[25] Babich D and Zaldarriaga M, *Primordial bispectrum information from CMB polarization*, 2004 Phys. Rev. D 70 083005 [SPIRES] [astro-ph/0408455]