Single-field inflation and the local ansatz: Distinguishability and consistency

Roland de Putter,[1] Olivier Doré,[2,1] Daniel Green,[3,4] and Joel Meyers[4]
[1]*California Institute of Technology, Pasadena, California 91125, USA*
[2]*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA*
[3]*University of California, Berkeley, California 94720, USA*
[4]*Canadian Institute for Theoretical Astrophysics, Toronto, Ontario M5S 3H8, Canada*
(Received 6 January 2017; published 3 March 2017)

The single-field consistency conditions and the local ansatz have played separate but important roles in characterizing the non-Gaussian signatures of single- and multifield inflation respectively. We explore the precise relationship between these two approaches and their predictions. We demonstrate that the predictions of the single-field consistency conditions can never be satisfied by a general local ansatz with deviations necessarily arising at order $(n_s - 1)^2$. This implies that there is, in principle, a minimum difference between single- and (fully local) multifield inflation in observables sensitive to the squeezed limit such as scale-dependent halo bias. We also explore some potential observational implications of the consistency conditions and its relationship to the local ansatz. In particular, we propose a new scheme to test the consistency relations. In analogy with delensing of the cosmic microwave background, one can deproject the coupling of the long wavelength modes with the short wavelength modes and test for residual anomalous coupling.

DOI: 10.1103/PhysRevD.95.063501

I. INTRODUCTION

Understanding the origin of the initial conditions for the Universe is one of the primary goals of modern cosmology. Most ambitiously, we hope to test fundamental principles behind the origin of structure, independently of any framework. For example, one might hope to distinguish whether the initial seeds are the result of quantum or classical fluctuations [1]. Even within the context of inflation, we would like to test the nature of inflation, including whether inflation was single- or multifield [2] or if inflation is a weakly or strongly coupled phenomenon [3,4]. Significant progress has been made in identifying possible observational targets [5], often in the context of deviations from Gaussianity. Still, many of these targets are qualitative in nature and more work remains to connect them to fundamental principles [4].

Perhaps the most quantitative tools for testing inflation are the single-field consistency conditions [2,6]. They state that when inflation is driven by a single field (or clock), the coupling of short and long modes is completely specified: $(N+1)$-point correlation functions involving short and long modes can be specified in terms of lower order correlation functions. These relations are testable observationally.

The basic reason underlying these conditions is that, to leading order in gradients, the long mode metric fluctuation ζ_L is locally a constant that is equivalent to a reparametrization of the clock. This logic has been extended to show the long mode has no *local* physical effects up to quadratic order in gradients [2,7,8]. As such, the statement of the consistency conditions is essentially that, modulo gradients of the long mode, the short modes cannot measure the presence of the long mode physically. The leading order effect of the long mode that *can* be measured locally is a perturbation to the local curvature, which is suppressed by k_L^2, where k_L is the wave number of ζ_L.

Whereas these consistency conditions were initially introduced by Maldacena to explain the properties of inflationary correlation functions [6], they have since been found to have very general consequences to cosmology [2], even at much later times. The essence of these consistency conditions was understood much earlier in the context of the separate universe approach (see e.g. [9,10]). Weinberg [11] later understood that these are all consequences of a large gauge transformation that may be implemented at any time (not just during inflation), which has ultimately made a number of powerful applications possible. In particular, it was shown to be straightforward to predict the implications of the consistency conditions for any observable and thus look for deviations [12–14].

Since the full set of consistency conditions strongly constrains the statistics of the initial conditions, it is natural to compare these constraints to those stemming from a common prescription for the initial conditions, namely the local ansatz. The local ansatz simply assumes that there exists some Gaussian random field $\zeta_g(\mathbf{x}, t)$ such that the initial conditions for the adiabatic mode are generated *locally* in this Gaussian field:

$$\zeta(\mathbf{x}, t_i) = \sum_n c_n \zeta_g(\mathbf{x}, t_i)^n = \zeta_g + \frac{3}{5} f_{\mathrm{NL}}^{\mathrm{local}} \zeta_g^2 + \cdots. \quad (1.1)$$

Data from the Planck satellite currently constrain $f_{\mathrm{NL}}^{\mathrm{local}} = 0.8 \pm 5.0$ [15] but future observations have the potential to reach $\sigma(f_{\rm NL}) < 1$ [16–19]. This is particularly interesting as $|f_{\rm NL}| > 1$ is a common feature of models that reproduce the local ansatz [5,20–22].

The idea that some nonlinear but local physics generated the initial conditions is very plausible and is indeed found to arise in many multifield models of inflation and alternatives to inflation. Nevertheless, the origin of the local ansatz in physical examples is qualitatively different from the single-field consistency conditions. The local ansatz is usually the consequence of local nonlinear evolution at times when all the observable modes are outside the horizon. Since there are no physical scales larger than the horizon, long and short wavelength modes are treated on the same footing. While local interactions also govern the single-field consistency conditions, only the long wavelength modes are outside the horizon and therefore the short and long modes are physically distinguishable in the resulting statistics.

Given the differences in the physics, it is natural to ask at what level one expects to find deviations in predictions made by the local ansatz and single-field inflation. This is particularly important when testing observationally the nature of inflation. The statement that the consistency conditions imply that $f_{\rm NL} = -\frac{5}{12}(n_s - 1)$ would seem to suggest that single-field inflation is equivalent to a local ansatz with specific coefficients. As we will show explicitly, this statement is not correct. First of all, the single-field consistency conditions are really an infinite set of constraints rather than just a statement of a single statistic [23] and matching the above relation would only confirm one from this infinite set. Second, as will be discussed further below, the relation between $f_{\rm NL}$ and n_s involves statistically average quantities whereas the consistency conditions should hold for any realization and not just statistically. This suggests that mapping the single-field consistency conditions onto parameters predicted by the local ansatz mischaracterizes the relevant physical effects.

Another motivation for this work is to further clarify the observability of the single-field consistency conditions. As has been emphasized by a number of authors, the consistency conditions physically imply that the short modes are statistically independent of the long mode, in physical coordinates. In this sense, single-field inflation predicts "zero mode coupling" which suggests there is no natural target for local non-Gaussianity, even in principle [14,24,25]. Nevertheless, as we will show, the local ansatz can never reproduce this prediction; it leaves a nonzero mode coupling at least of order $(n_s - 1)^2$ in any such observable and therefore sets a natural target (although unobservable in practice). For example, the local ansatz will always lead to scale-dependent bias[1] while single-field inflation does not [13,14,25].

[1]Here, scale-dependent bias refers to any term in the bias expansion which is not consistent with locality in space. This includes terms like $\zeta_L^{n>2}$ which are nonlocal and also nonlinear.

In this paper, we will explore the relationship between the consistency conditions and the local ansatz. In Sec. II, we will show that the local ansatz cannot reproduce the consistency conditions for any choice of parameters. In Sec. III, we describe how the local ansatz needs to be modified to be consistent with Weinberg's derivation of the consistency conditions. In Sec. IV, we will demonstrate how the mode coupling induced in single-field inflation can be deprojected from the observed statistics in direct analogy with weak lensing of the cosmic microwave background.

II. VIOLATING THE SINGLE-FIELD CONSISTENCY CONDITIONS

In this section, we will show that the local ansatz, $\zeta(\mathbf{x}) = \sum_n c_n \zeta_g(\mathbf{x})^n$, cannot satisfy the single-field consistency conditions for any choice of c_n. It will be important that the coefficients c_n cannot depend on the location in space because we are assuming that only $\zeta_g(\mathbf{x})$ breaks homogeneity. Therefore, c_n is a list of numbers rather than functions.

The qualitative reason these two models do not agree can be understood as follows. The local ansatz cannot distinguish long and short modes (as required by locality), and therefore a given coefficient predicts that a number of different mode couplings are related. This is particularly important for $c_{n>2}$ as there is more than one long-short coupling per coefficient. If the local ansatz is to match the single-field consistency conditions, these nontrivial relations must also arise in single-field inflation. However, single-field inflation distinguishes long and short modes and there is no reason to expect the same relations to hold. The essence of this section is to check that this expected difference cannot be eliminated by carefully choosing the coefficients of the local expansion.

We first need to be clear about how the consistency conditions act on correlation functions of short modes.[2] Let us start with a metric without a long mode such that

$$d\tilde{s}^2 = -dt^2 + a(t)^2 e^{2\tilde{\zeta}_S(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}^2. \tag{2.1}$$

Now we introduce the long mode through the transformation $\mathbf{x} = e^{-\zeta_L}\tilde{\mathbf{x}}$, which implies

$$\begin{aligned} ds^2 = d\tilde{s}^2 &= -dt^2 + a(t)^2 e^{2\tilde{\zeta}_S(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}^2 \\ &= -dt^2 + a^2(t) e^{2\tilde{\zeta}_S(e^{\zeta_L}\mathbf{x}) + 2\zeta_L} d\mathbf{x}^2, \end{aligned} \tag{2.2}$$

[2]The exact separation between short and long modes is not always precise. Very conservatively, requiring $k_L/k_S < \mathcal{O}(100)$ should guarantee that we are in the squeezed limit where the consistency conditions apply [26]. For many single-field models, a small hierarchy is sufficient.

where $\tilde{\zeta}_S$ is the original short perturbation that is independent of ζ_L. Throughout, we will ignore all gradients[3] of ζ_L and keep only the leading order behavior in the limit of vanishing wave number, $k_L \to 0$. The resulting transformation of the short mode is

$$\zeta_S(\mathbf{x}) = \tilde{\zeta}_S(e^{\zeta_L}\mathbf{x}) = \tilde{\zeta}_S(\mathbf{x}) + \zeta_L \mathbf{x}\cdot\nabla\tilde{\zeta}_S(\mathbf{x}) + \cdots. \tag{2.3}$$

Thus, in the presence of a long mode ζ_L, all "local" statistics of ζ, i.e. N-point functions purely of the short modes ζ_S, can be obtained by evaluating the same quantities in the absence of the long mode, but at a different scale, $\mathbf{x} \to \mathbf{x}e^{\zeta_L}$, or $\mathbf{k} \to \mathbf{k}e^{-\zeta_L}$.

We will compare this to the local ansatz, which we will define as

$$\zeta(\mathbf{x}, t_i) = \sum_n c_n \zeta_g(\mathbf{x}, t_i)^n = \zeta_g + \frac{3}{5} f_{\rm NL}^{\rm local} \zeta_g^2 + \cdots \tag{2.4}$$

where from here on, we will drop the dependence on the initial time t_i. Here ζ_g is assumed to satisfy Gaussian statistics and therefore $\zeta_{g,S}$ and $\zeta_{g,L}$ are statistically independent.[4] The local ansatz thus leads to a mode coupling,

$$\begin{aligned}\zeta_S = {} & c_1\zeta_{g,S} + c_2\zeta_{g,S}^2 + c_3\zeta_{g,S}^3 + \cdots \\ & + \zeta_{g,L}(2c_2\zeta_{g,S} + 3c_3\zeta_{g,S}^2 + \cdots) \\ & + \zeta_{g,L}^2(3c_3\zeta_{g,S} + \cdots).\end{aligned} \tag{2.5}$$

While this series extends to arbitrary orders in $\zeta_{g,S}$ and $\zeta_{g,L}$ it is important that the modulation of a connected $(N+1)$-point correlation function of short modes at $\mathcal{O}(\zeta_{g,L}^q)$ is determined by $c_{n\le N+q}$ (ignoring loop-suppressed corrections).

Now, let us examine what the single-field consistency conditions predict for the behavior of the two-point statistics of the short modes. Up to second order in the long mode, we find

$$\begin{aligned}\langle\zeta_S(\mathbf{k})\zeta_S(\mathbf{k}')\rangle' &= \langle\tilde{\zeta}_S(\mathbf{k}e^{-\zeta_L})\tilde{\zeta}_S(\mathbf{k}'e^{-\zeta_L})\rangle' \\ &= e^{-(n_s-1)\zeta_L}\langle\tilde{\zeta}_S(\mathbf{k})\tilde{\zeta}_S(\mathbf{k}')\rangle' \\ &= P_S(k) - (n_s-1)P_S(k)\zeta_L \\ &\quad + \frac{1}{2}(n_s-1)^2 P_S(k)\zeta_L^2 + \mathcal{O}(\zeta_L^3).\end{aligned} \tag{2.6}$$

The primes in the first line indicate that we suppress the usual factor $(2\pi)^3\delta^{(D)}(\sum_i \mathbf{k}_i)$ from the expectation value. We have taken n_s to be constant, since if it depended on scale, the local ansatz would fail to match the predictions of the single-field consistency conditions. Thus, a first (trivial) requirement for the local ansatz to reproduce the predictions of the consistency conditions is that the spectral index is scale independent.

Up to second order in the long mode, we find for the local ansatz

$$\begin{aligned}\langle\zeta_S(\mathbf{k})\zeta_S(\mathbf{k}')\rangle' &= (c_1^2 + 4c_1c_2\zeta_L + (4c_2^2 + 6c_1c_3)\zeta_L^2) \\ &\quad \times \langle\zeta_{g,S}(\mathbf{k})\zeta_{g,S}(\mathbf{k}')\rangle' + \mathcal{O}(\zeta_L^3) \\ &= c_1^2 P_S(k) + 4c_1c_2 P_S(k)\zeta_L \\ &\quad + (4c_2^2 + 6c_1c_3)P_S(k)\zeta_L^2 + \mathcal{O}(\zeta_L^3).\end{aligned} \tag{2.7}$$

Matching[5] the two-point predictions of the single-field consistency conditions order by order requires that we have $c_1 = 1$, $c_2 = -\frac{1}{4}(n_s-1)$ [i.e. the familiar $f_{\rm NL} = -\frac{5}{12}(n_s-1)$], and $c_3 = \frac{1}{24}(n_s-1)^2$. This choice of coefficients then dictates the behavior of the three-point function of short modes for the local ansatz

$$\begin{aligned}\langle\zeta_S\zeta_S\zeta_S\rangle' &= 6c_1^2c_2P_S(k)^2 + (18c_1^2c_3 + 24c_1c_2^2)P_S(k)^2\zeta_L \\ &\quad + \mathcal{O}(\zeta_L^2) \\ &= -\frac{3}{2}(n_s-1)P_S(k)^2 + \frac{9}{4}(n_s-1)^2P_S(k)^2\zeta_L \\ &\quad + \mathcal{O}(\zeta_L^2).\end{aligned} \tag{2.8}$$

Returning to the predictions of single-field inflation, we are free to choose the form of the bispectrum containing only short modes since that correlation is unconstrained by symmetries (although it would be very constraining if the only way to reconcile the local ansatz with single-field inflation is for this exact form of the local bispectrum). Once this choice is made, however, the scaling of the bispectrum with long modes is completely determined by the single-field consistency conditions

$$\begin{aligned}\langle\zeta_S\zeta_S\zeta_S\rangle' &= -\frac{3}{2}(n_s-1)P_S(k)^2 e^{-2(n_s-1)\zeta_L} \\ &= -\frac{3}{2}(n_s-1)P_S(k)^2 + 3(n_s-1)^2P_S(k)^2\zeta_L \\ &\quad + \mathcal{O}(\zeta_L^2).\end{aligned} \tag{2.9}$$

Comparing Eqs. (2.8) and (2.9), we see that if the coefficients of the local ansatz are chosen to make the

[3]We can extend these results to linear order in gradients using the conformal consistency conditions [8].

[4]The perturbation $\tilde{\zeta}_S$ appearing in the consistency conditions is simply the small-scale perturbation in the absence of the long mode so that $\tilde{\zeta}_S$ is not in general equal to ζ_g because we have made no assumption about the statistics of $\tilde{\zeta}_S$.

[5]Note that our predictions are for the global statistics of ζ and may not match the observations in a given Hubble patch [27,28]. A similar argument could be applied instead to the statistics only in a specific Hubble region. The claim that the local anstaz cannot reproduce the single-field consistency conditions holds for both cases.

behavior of the two-point statistics of the short modes match the predictions of the consistency conditions, then the predictions for the bispectrum necessarily disagree at $\mathcal{O}((n_s-1)^2)$. Furthermore, we cannot correct this disagreement by introducing additional terms to the local ansatz with $c_{n>3}$ because no such terms contribute to the three-point statistics of the short modes at first order in ζ_L (except through loops which are highly suppressed).

The origin of this contradiction can be generalized to arbitrary orders in ζ_L. Suppose we truncate the local expansion at order ζ^N. In this case, once we make the split into long and short modes, we have

$$\zeta = \sum_{n=1}^{N} c_n(\zeta_{g,S}+\zeta_{g,L})^n \tag{2.10}$$

and we can always fix $c_1 = 1$ by definition. This means we have $N-1$ unknown coefficients to match $\langle\zeta_{g,S}^m\rangle$ to order ζ_L^{N-m+1} where $m = 2\ldots N$. We find that there are $\sum_{m=2}^{N}(N-m+1) = \sum_{i=1}^{N-1} i = N\times(N-1)/2$ different coefficients that we need to match using these $N-1$ unknown coefficients. This system is therefore overconstrained and it would thus be a miracle if the coefficients matched the consistency conditions.

We can see that the general pattern matches the explicit calculations including $c_{1,2,3}$. For $N=2$, we have one coefficient (c_2) but we only have to match one number, the squeezed limit of the bispectrum. At order $N=3$, we have 2 coefficients c_2, c_3 but now we have 3 different squeezed limits to match and we simply cannot pick c_2 and c_3 to make them all agree with the single-field consistency conditions. At order N we should find that floor($N/2$) consistency conditions cannot be satisfied by the local ansatz.

In summary, we have shown that it is impossible to exactly obey the single-field consistency conditions with the local ansatz. In that sense, testing the single- vs multifield nature of inflation by constraining $f_{\rm NL}$, etc., within the local ansatz is technically not correct, as no point in this parameter space is consistent with single-field inflation. However, the local ansatz is of course still very useful as a shorthand description for the squeezed limit behavior of the bispectrum and/or the collapsed limit trispectrum. These are also the quantities that determine the leading order signal of scale-dependent halo bias [29–32], which is one of the main ways in the near future to constrain primordial non-Gaussianity using large-scale structure [5]. This is how the local ansatz is most commonly used, and in this sense the single-field case is indeed equivalent to $f_{\rm NL} = -\frac{5}{12}(n_s-1)$. However, if one were to use the local form to also predict e.g. the modulation of the short-scale bispectrum, $\langle\zeta_L\zeta_S^3\rangle$, and higher order modulations in ζ_L^2 such as $\langle\zeta_L^2\zeta_S^2\rangle$, we have shown that one would inevitably make predictions inconsistent with single-field inflation. Of course, in practice, these deviations from the predictions of single-field inflation are too small to be detected with any near-term observations.

III. CONSISTENCY CONDITIONS FOR THE LOCAL ANSATZ

In the previous section, we found that the local ansatz can never match the predictions of the single-field consistency conditions. Physics is rarely discontinuous and therefore we expect that there is some generalization of the local ansatz that should allow us to interpolate between the two. This is also obvious from the point of view of model building, as we can certainly write models of inflation that interpolate between single- and multifield by varying the mass of the additional fields. However, if we take the local ansatz as our starting point, we want to know the minimal set of terms needed to reproduce both limits.

There are two generalizations of the local ansatz that could plausibly change our results: (1) multiple random fields and (2) "nonlocal" terms[6] in the expansion in the Gaussian random field(s). Given that the local ansatz is a prediction of multifield inflation, adding more random fields is an obvious choice. We will see that adding multiple fields is not a sufficient condition, but that both nonlocal terms and multiple fields are needed to interpolate between the consistency conditions and the local ansatz.

Let us consider a scenario with perturbations in two directions (this can be straightforwardly generalized to the case of more than two fields), ζ and σ, and let us assume that any shift in the perturbation with $\Delta\sigma = 0$ implies the shift is along the adiabatic direction. Varying σ at $\zeta = 0$ then of course describes an isocurvature fluctuation.[7]

The single-field consistency conditions in this more general context are really consistency conditions about the effects of an *adiabatic* shift in the long-mode fluctuation (see e.g. [12,24,33–35] for related discussions). Specifically, the generalization of the single-field consistency conditions, Eq. (2.3), is that under such a transformation,

$$\begin{aligned}\tilde\zeta(\mathbf{x}) &\to \zeta(\mathbf{x}) = \tilde\zeta_S(e^{\Delta\zeta_L}\mathbf{x}) + \tilde\zeta_L + \Delta\zeta_L,\\ \tilde\sigma(\mathbf{x}) &\to \sigma(\mathbf{x}) = \tilde\sigma_S(e^{\Delta\zeta_L}\mathbf{x}) + \tilde\sigma_L,\end{aligned} \tag{3.1}$$

[6] We remind the reader that *local* is taken in the sense of the local ansatz, i.e. functions of the form $\Phi(\mathbf{x}) = F(\{\phi_i(\mathbf{x})\})$. Nonlocal terms need not imply a violation of causality/locality in the dynamics of ϕ. Nonlocal terms can arise when statistics have memory of past evolution and/or when there is a scale, such as the horizon, that can distinguish the wavelengths of $\phi(\mathbf{x})$ (the local form necessarily treats all wavelengths on the same footing).

[7] As a simple example, in the case with two scalar fields $\phi = \bar\phi + \delta\phi$ and $\chi = \bar\chi + \delta\chi$, a commonly considered scenario is one where the curvature-isocurvature basis is approximately aligned with the $\delta\phi - \delta\chi$ basis, so that $\zeta \approx -\frac{H}{\dot{\bar\phi}}\delta\phi$, and $\sigma \approx \delta\chi$. This is typically the case for the initial conditions in models where χ is a spectator field during inflation.

where quantities with a tilde are the fields in the absence of the shift $\Delta\zeta_L$, which must be statistically independent of $\Delta\zeta_L$. If all we wanted was to express the consistency conditions in a multifield scenario, Eq. (3.1) would be sufficient. However, the above expression does not fully specify the statistics of the curvature perturbation (nor of σ), as it does not say anything about the statistics of $\tilde{\zeta}_S$ and $\tilde{\sigma}$, other than their independence of $\Delta\zeta_L$. In particular, we have not fully specified the response of ζ_S to long modes, because we have not specified the response to σ_L.

The usual *local ansatz*, Eq. (2.4), fixes the *full* statistics of the curvature perturbation by expressing ζ as a local function of a *Gaussian* field ζ_g. We would like to do the same here, but using the presence of σ (or in general of multiple fields) to remain in agreement with the consistency conditions. Specifically, we would like to express the perturbations in terms of two Gaussian fields, $\zeta_{\rm ad}$ (an adiabatic fluctuation) and σ_g, to fully specify the statistics.[8] Based on Eq. (3.1), a minimal consistent ansatz we could write is

$$\begin{aligned}\zeta(\mathbf{x}) &= \zeta_{{\rm ad},S}(e^{\zeta_{{\rm ad},L}}\mathbf{x}) + \zeta_{{\rm ad},L},\\ \sigma(\mathbf{x}) &= \sigma_g(e^{\zeta_{{\rm ad},L}}\mathbf{x}),\end{aligned}\tag{3.2}$$

with the component fields Gaussian. Now, the most general[9] local transformation of this ansatz that still respects Eq. (3.1) is

$$\begin{aligned}\zeta &\to \zeta = \zeta + f(\sigma),\\ \sigma &\to \sigma = g(\sigma),\end{aligned}\tag{3.3}$$

leading to the generalized local ansatz,

$$\begin{aligned}\zeta(\mathbf{x}) &= [\zeta_{{\rm ad},S} + f(\sigma_g)](e^{\zeta_{{\rm ad},L}}\mathbf{x}) + \zeta_{{\rm ad},L},\\ \sigma(\mathbf{x}) &= [g(\sigma_g)](e^{\zeta_{{\rm ad},L}}\mathbf{x}).\end{aligned}\tag{3.4}$$

Finally, expressing the generalized local ansatz for ζ to second order, and separating short and long modes, gives

$$\begin{aligned}\zeta &= [\zeta_{{\rm ad},S} + \sigma_{g,S}](e^{\zeta_{{\rm ad},L}}\mathbf{x}) + c_2\sigma_g^2 + \zeta_{{\rm ad},L} + \sigma_{g,L} + \cdots\\ &= \zeta_{\rm ad} + \sigma_g + \zeta_{{\rm ad},L}\mathbf{x}\cdot\nabla[\zeta_{{\rm ad},S} + \sigma_{g,S}] + c_2\sigma_g^2 + \cdots\\ &\quad (\text{generalized local ansatz})\end{aligned}\tag{3.5}$$

where we have Taylor expanded f in powers of σ_g and then absorbed the coefficients $\partial_\sigma f$ and $\partial_\sigma^2 f$ into σ_g and c_2. The statistics of ζ, and in particular the mode coupling, are now fully determined by Eq. (3.5) as soon as the variance of $\zeta_{\rm ad}$ and σ_g are specified. We choose them to be uncorrelated[10] $\langle\zeta_{\rm ad}\sigma_g\rangle = 0$. Clearly, the restriction placed on Eq. (3.3) by the adiabatic consistency conditions means our final form can only have significant local-type non-Gaussianity due to the presence of the second field, σ_g.

Finally, in cases where there is more than one nonadiabatic mode (more than two fields), one can without loss of generality define $\sigma_g \equiv \sigma_{g,1}$ to be the linear combination contributing linearly to ζ, generalizing Eq. (3.5) so that only the quadratic term is modified,

$$c_2\sigma_g^2 \to \sum_{ij} c_{2,ij}\sigma_{g,i}\sigma_{g,j},\tag{3.6}$$

where the sum is over all nonadiabatic modes $\sigma_{g,i}$, with $\langle\sigma_{g,i}\sigma_{g,j}\rangle = 0$ for $i \neq j$. The modes $\sigma_{g,i}$ with $i > 1$ exclusively contribute to stochastic non-Gaussianity because they are by definition uncorrelated with ζ at linear order.

Equation (3.5) is *not* intended to be the most general form for non-Gaussianity in multifield inflation. It is merely an ansatz that, loosely speaking, minimally satisfies the consistency conditions, and allows for all local (in the sense discussed in the beginning of this section) terms that do not violate them. However, we have not addressed how this ansatz can arise physically. There are two implicit assumptions about the dynamics that are crucial:

(i) The mode coupling at horizon crossing is trivial. The horizon sets a natural scale that allows for terms that are not of the local form. Most significantly, this would allow for terms of the form $(\zeta_S)^m(\sigma_L)^n$ that are allowed by the consistency conditions.

(ii) The fluctuations in ζ at constant σ correspond to the adiabatic mode that is constant in time outside of the horizon. It is this mode that can be removed by a coordinate transformation. This is an assumption about having reached the inflationary attractor solution.

We can make these points more concrete by considering a simple multifield inflation scenario. The discussion below closely resembles the “derivation” above of the generalized ansatz. We can decompose field perturbations in terms of curvature and isocurvature fluctuations. For instance, in a 2-field model with separable potential $W(\phi,\chi) = U(\phi) + V(\chi)$, and assuming slow roll for simplicity, the curvature perturbation is, to first order,

[8]The fields technically do not have to be Gaussian. To specify the mode coupling, we really only need to demand that the short-mode components of $\zeta_{\rm ad}$ and σ_g are independent of the long-mode components.

[9]It is straightforward to check that any other local term is not allowed in Eq. (3.3). The contributions $\zeta \to F(\zeta) = F(\zeta_{{\rm ad},S}(e^{\zeta_{{\rm ad},L}}\mathbf{x}) + \zeta_{{\rm ad},L})$ or $\sigma \to G(\zeta) = G(\zeta_{{\rm ad},S}(e^{\zeta_{{\rm ad},L}}\mathbf{x}) + \zeta_{{\rm ad},L})$ will not obey the transformation in Eq. (3.1) unless $F(x) = x$ and $G(x) = 0$.

[10]If Eq. (3.5) holds, but $\zeta_{\rm ad}$ and σ_g are *a priori not* independent, we can always apply a redefinition $\zeta_{\rm ad} \to \zeta'_{\rm ad} \equiv \zeta_{\rm ad} + A\sigma_g$ such that $\zeta'_{\rm ad}$ and σ_g *are* independent. However, after the redefinition, the mode coupling would have a slightly more general form (dropping the prime and reabsorbing some coefficients into σ_g and c_2), $\zeta = [\zeta_{{\rm ad},S} + \sigma_{g,S}](e^{\zeta_{{\rm ad},L} - \alpha\sigma_{g,L}}\mathbf{x}) + c_2\sigma_g^2 + \zeta_{{\rm ad},L} + \sigma_{g,L} + \cdots$.

$$\zeta = \frac{WU_\phi}{U_\phi^2 + V_\chi^2}\delta\phi + \frac{WV_\chi}{U_\phi^2 + V_\chi^2}\delta\chi + \mathcal{O}(\delta\phi^2, \delta\chi\delta\phi, \delta\chi^2). \tag{3.7}$$

While we have only included the linear order terms, ζ is defined to all orders in the fluctuations. We can choose

$$\sigma \propto \frac{\delta\phi}{U_\phi} - \frac{\delta\chi}{V_\chi}, \tag{3.8}$$

so that $\sigma = 0$ corresponds to an adiabatic fluctuation (to first order).

Now consider initial conditions at some time when all modes of interest have just exited the horizon, indicated by a $*$ subscript. In scenarios with two light fields, $\delta\phi_*$ and $\delta\chi_*$ are typically close to Gaussian and independent. Writing only the minimal mode coupling required to satisfy the consistency conditions, we can then express the initial fluctuations in the $\zeta - \sigma$ basis in terms of truly independent Gaussian fields (which we will again write as $\zeta_{\rm ad}$ and σ_g) as

$$\zeta_* = \zeta_{{\rm ad},S}(e^{\zeta_{{\rm ad},L}}\mathbf{x}) + \zeta_{{\rm ad},L}, \tag{3.9}$$

$$\sigma_* = \sigma_{g,S}(e^{\zeta_{{\rm ad},L}}\mathbf{x}) + \sigma_{g,L}. \tag{3.10}$$

In essence, we are assuming that the physics of horizon crossing is trivial (in local coordinates) and all subsequent evolution can be treated classically from these initial conditions.[11] After all modes have exited the horizon, one can then describe the evolution of perturbations in terms of the separate Universe picture/δN formalism, where evolution is classical and local (in the sense discussed above). The initial adiabatic perturbations are then nonlinearly conserved, but the entropy perturbation can be transferred into ζ at both linear and nonlinear order. Moreover, a purely adiabatic perturbation ($\sigma_* = 0$) remains adiabatic. In other words, evolution gives

$$\zeta_* \to \zeta = \zeta_* + f(\sigma_*) = \zeta_* + N_{\sigma_*}\sigma_* + \frac{1}{2}N_{\sigma_*\sigma_*}\sigma_*^2 + \cdots,$$
$$\sigma_* \to \sigma = g(\sigma_*), \tag{3.11}$$

where N_{σ_*} and $N_{\sigma_*\sigma_*}$ refer to the fact that in the δN formalism, the effect of the initial isocurvature perturbation can be computed as the response of the number of e-foldings of expansion up to a constant-density hypersurface. Thus, in this scenario, we end up with exactly our generalized local ansatz (3.5), where $\zeta_{\rm ad}$ and σ_g now have the physical interpretation of (Gaussian components of) the initial curvature and isocurvature perturbations at horizon exit.

We can understand from this example where our implicit assumptions are necessary. The critical simplification is that we reduced the problem from four real solutions down to two, the growing modes $\zeta_\star$ and $\sigma_\star$. If we set $\sigma_\star = 0$, then we are by definition in the adiabatic attractor solution and, by definition, we must reproduce all the predictions of the single-field consistency conditions. This is what forces $\zeta|_{\sigma_\star=0} = \zeta_\star$. Furthermore, having truncated the number of solutions, the second solution can always be rewritten in terms of the initial condition for the isocurvature mode, $\sigma_\star$. If we allow for nontrivial mode coupling at horizon crossing, but retain the truncation of the superhorizon solutions, we can generate mode coupling of the from $(\zeta_S)^m(\sigma_L)^n$, but no coupling to ζ_L beyond those in (3.9). Although the consistency conditions allow mode coupling between ζ_S and $\dot\zeta_L$, the evolution requires that $\dot\zeta \propto f(\sigma_\star)$, and we can always rewrite the result in terms of the isocurvature mode.

The more dramatic modification to the local ansatz occurs when the "decaying" modes are no longer negligible. It remains generally true that when we set $\sigma(\mathbf{x}, t) = 0$, we must reproduce all the predictions of single-field inflation; yet, a more general model allows higher order mixing between σ and ζ, like those appearing in the EFT of multifield inflation [36]. In deriving Eqs. (3.5) and (3.11), we were able to forbid all such terms by symmetry. However, in doing this, we were assuming that $\zeta_{\rm ad}(\mathbf{x}, t)$ is the solution that is constant outside the horizon. Of course, there is always a second solution that violates this assumption, but typically decays as a^{-3} and plays no role in the dynamics. However, with sufficiently rapid time dependence, sharp turns in field space, or other nontrivial dynamics, the decaying modes may not be negligible at some time during inflation and may generate nontrivial mode couplings.[12] In fact, if we allow for nonattractor solutions (i.e. the constant mode is the decaying mode), we may violate the consistency conditions even in single-field inflation [26,37–40].

Now that we have covered the physical interpretation of the generalized local ansatz, let us briefly consider its implications. Although $\zeta_{\rm ad}$ must always be present to maintain diffeomorphism invariance, when $P_{\zeta_{\rm ad}} \ll P_\sigma$ we

[11] We could even allow for significant initial non-Gaussianity in σ by adding a term $\mathcal{O}(\sigma_g^2)$ to Eq. (3.10). This would leave the final form of the statistics unchanged. In models with multiple light fields, deviations from these initial statistics are typically slow-roll suppressed.

[12] One may wonder how such contributions can arise without violating the symmetries in Eq. (3.3). Because the decaying mode necessarily depends on time, one can include terms of the form $\int dt' \dot\zeta_{{\rm ad},L}(t')$ that are manifestly invariant under (3.3) but are proportional only to the decaying mode. These terms are nonlocal in time in our ansatz, but are perfectly consistent with local time evolution. This is simply a reflection that the statistics have a memory of the past evolution (which is the same reason they encode information about inflation when we measure them much later).

can effectively neglect $\zeta_{\rm ad}$ for the purpose of computing statistics. In this limit, we will reproduce the results of the standard local ansatz. More generally, one should include both terms. For example, if we compute $f_{\rm NL}$ using Eq. (3.5) we have

$$f_{\rm NL} = \frac{5}{12}\frac{\langle\zeta_L\zeta_S\zeta_S\rangle'}{\langle\zeta_L^2\rangle'\langle\zeta_S^2\rangle'} = \frac{5}{12}\frac{[-P_{\zeta_{\mathrm{ad},L}}(3+\frac{\partial}{\partial\ln k})(P_{\zeta_{\mathrm{ad},S}}+P_{\sigma_S})+4c_2P_{\sigma_L}P_{\sigma_S}]}{(P_{\zeta_{\mathrm{ad},L}}+P_{\sigma_L})(P_{\zeta_{\mathrm{ad},S}}+P_{\sigma_S})}. \tag{3.12}$$

It is easy to see that the first term is a statement of the consistency conditions in the presence of σ. Furthermore, the contribution to $f_{\rm NL}$ from each term is suppressed by the relative contribution σ_L or $\zeta_{\mathrm{ad},L}$ makes to ζ_L. Now if we take the limit $P_{\sigma_L} \gg P_{\zeta_{\mathrm{ad},L}}$ or $P_{\sigma_L} \ll P_{\zeta_{\mathrm{ad},L}}$ we effectively return to the local ansatz or the single-field consistency conditions respectively.

For higher N-point functions, the presence of $\zeta_{\rm ad}$ and σ with $\langle\zeta_{\rm ad}\sigma\rangle = 0$ will also lead to stochastic non-Gaussianity (and scale-dependent stochastic bias [41]). Specifically, the collapsed limits of higher N-point functions will be enhanced relative to the expectation from lower N-point functions. For example if $c_2 \gg (n_s-1)$, $\tau_{\rm NL}$ is given by

$$\begin{aligned}\tau_{\rm NL} &= \frac{1}{4}\frac{1}{P_{\zeta_L}P_{\zeta_S}^2}\langle\zeta(\mathbf{k}_S-\mathbf{k}_L)\zeta(-\mathbf{k}_S)\zeta(\mathbf{k}'_S+\mathbf{k}_L)\zeta(-\mathbf{k}'_S)\rangle' \\ &\approx 4c_2^2\frac{P_{\sigma_L}P_{\sigma_S}^2}{P_{\zeta_L}P_{\zeta_S}^2} \\ &\approx \left(\frac{6}{5}f_{\rm NL}\right)^2\frac{(P_{\zeta_{\mathrm{ad}L}}+P_{\sigma_L})}{P_{\sigma_L}},\end{aligned} \tag{3.13}$$

where the last line follows from Eq. (3.12) and $P_{\zeta_L} \equiv P_{\zeta_{\mathrm{ad}L}} + P_{\sigma_L}$. We see that the amplitude is enhanced by $\frac{(P_{\zeta_{\mathrm{ad}L}}+P_{\sigma_L})}{P_{\sigma_L}} \geq 1$ relative to the expectation from local ansatz with a single field,[13] namely $\tau_{\rm NL} = (\frac{6}{5}f_{\rm NL})^2$. The reason is that the noncollapsed N-point functions are suppressed by the correlation coefficient of σ with ζ because we do not observe σ_L directly. This additional suppression does not arise in collapsed configurations where we do not need to directly measure σ_L to be sensitive to its mode coupling. It is the same reason that one finds scale-dependent stochastic bias in these models [41]; halos are biased with respect to σ_L which is not fully correlated with the linear density field.

[13]The Suyama-Yamaguchi inequality [42], $\tau_{\rm NL} \geq (\frac{6}{5}f_{\rm NL})^2$, must always be satisfied [43,44] but is saturated for a single degree of freedom (up to loop corrections [45]).

IV. DEPROJECTING THE LONG MODE

We showed in Sec. II that the single-field consistency conditions are more than just statements about the squeezed limit bispectrum, but instead dictate the response of the full short-wavelength statistics to a long mode. Specifically, in terms of the statistically independent fluctuation, $\tilde{\zeta}_S(\mathbf{x})$, it is a remapping of coordinates by the long mode,

$$\zeta_S(\mathbf{x}) = \tilde{\zeta}_S(e^{\zeta_L(\mathbf{x})}\mathbf{x}). \tag{4.1}$$

One way of testing this condition in all its richness is to study various N-point functions, correlating the long mode with powers of the short mode, e.g. $\langle\zeta_L\zeta_S^n\rangle$. An intriguing alternative follows from the realization that the remapping in Eq. (4.1) is reminiscent of the effect of the gravitational lensing deflection field on cosmic microwave background (CMB) fluctuations (see e.g. [46] for review). For example, lensing of CMB temperature is given by

$$T(\mathbf{x}) = \tilde{T}(\mathbf{x}+\boldsymbol{\alpha}(\mathbf{x})), \tag{4.2}$$

where $T(\mathbf{x})$ is the lensed CMB temperature, $\tilde{T}$ the unlensed temperature, and $\alpha(\mathbf{x})$ is the deflection field. In the CMB, given a measurement of the lensed temperature map $T(\mathbf{x})$, it is well known that one can reconstruct the actual *realization* of the lensing deflection field and then "delens" the CMB fluctuations to obtain $\tilde{T}$ (see e.g. [47–49]). It should therefore be possible, in principle, to do the same in the present context, i.e. use an estimate of the long mode, $\hat{\zeta}_L$ (we will use hats to denote estimators), to locally map ζ_S back to $\tilde{\zeta}_S$, assuming the consistency conditions,

$$\hat{\tilde{\zeta}}_S(\mathbf{x}) \equiv \hat{\zeta}_S(e^{-\hat{\zeta}_L}\mathbf{x}). \tag{4.3}$$

Assuming $\hat{\zeta}_L$ is unbiased, the resulting "deprojected" short mode thus gives the fluctuations in a local unperturbed coordinate system, i.e. the fluctuations as they would appear to a local observer.[14] If the consistency conditions indeed hold, these local fluctuations should be completely independent of the long mode,

$$\hat{\tilde{\zeta}}_S(\mathbf{x}) \to \tilde{\zeta}_S(\mathbf{x}). \tag{4.4}$$

Technically speaking, the procedure defined in Eq. (4.3) does not perfectly deproject the long mode, due to the position dependence of the long mode, but this procedure

[14]In general, one can test the consistency conditions by considering any local observable and testing if it depends on the long mode. Another good example is halo number density, which can only depend on local physics. If the consistency conditions hold, this quantity can not be modulated by ζ_L (modulo gradients of ζ_L) so that the $\propto k^{-2}$ scale-dependent bias has to be exactly zero [13,14,25].

can be promoted to an exact inversion along the same lines as delensing in the CMB.

Thus, one can test the consistency conditions by comparing the local statistics of the deprojected short mode in different spatial patches, and checking that they are independent of ζ_L. These local statistics can be N-point functions of $\hat{\hat{\zeta}}_S$ or histograms of the mode amplitudes, or another statistic. The point is that the consistency conditions predict that any local statistic will have to be independent of the long mode.

For the estimate of the long mode $\hat{\zeta}_L$, there are two scenarios. First, one could imagine measuring it directly from large-scale structure. Second, one could take the CMB lensing analogy further, and reconstruct the realization of the long mode directly from the statistics of the short modes assuming the consistency conditions. By analogy with the quadratic estimator for lensing reconstruction, we have

$$\hat{\zeta}_L^{\text{q.e.}}(\mathbf{k}) = N(\mathbf{k}) \int d^3\mathbf{k}' \zeta_S(\mathbf{k}') \zeta_S(\mathbf{k}-\mathbf{k}') g(\mathbf{k}',\mathbf{k}). \tag{4.5}$$

If we assume that the consistency conditions hold, we can make our estimator unbiased at first order in ζ_L by requiring that

$$\begin{aligned} \zeta_L(\mathbf{k}) &= \langle \hat{\zeta}_L^{\text{q.e.}}(\mathbf{k}) \rangle_{\zeta_S}{}' \\ &= N(\mathbf{k}) \int d^3\mathbf{k}' \langle \zeta_S(\mathbf{k}') \zeta_S(\mathbf{k}-\mathbf{k}') \rangle' g(\mathbf{k}',\mathbf{k}) \\ &\approx -N(\mathbf{k}) \int d^3\mathbf{k}' (n_s - 1) P_S(k') \zeta_L(\mathbf{k}) g(\mathbf{k}',\mathbf{k}), \end{aligned} \tag{4.6}$$

where we have used Eq. (2.6) in the second line. This then fixes our choice of $N(\mathbf{k})$ to be

$$N(\mathbf{k})^{-1} = -(n_s - 1) \int d^3\mathbf{k}' P_S(k') g(\mathbf{k}',\mathbf{k}). \tag{4.7}$$

One could go on to define the weights $g(\mathbf{k}',\mathbf{k})$ which minimize the variance of the estimator for a particular set of observations of the short modes, but that will not be necessary here.

Note, however, that if the long mode is estimated via "lensing" reconstruction, Eq. (4.5), $\hat{\zeta}_L$ will be biased if the consistency conditions are violated. To leading order in $n_s - 1$ and ζ_L, we can estimate this bias by

$$\begin{aligned} \hat{\zeta}_L^{\text{q.e.}}(\mathbf{k}) \approx{}& \frac{\int d^3\mathbf{k}' \langle \zeta_S(\mathbf{k}') \zeta_S(\mathbf{k}-\mathbf{k}') \zeta_L(\mathbf{k}) \rangle' P_L^{-1}(k) g(\mathbf{k}',\mathbf{k})}{-(n_s - 1) \int d^3k' P_S(k') g(\mathbf{k}',\mathbf{k})} \\ &\times \zeta_L(\mathbf{k}). \end{aligned} \tag{4.8}$$

We see that the leading bias is determined by the squeezed limit of the three-point function. However, if we do not have an independent measure of ζ_L we cannot see this bias directly. Furthermore, for the local ansatz we would also find that the variance of $\hat{\hat{\zeta}}_S(\mathbf{x}) = \hat{\zeta}_S(e^{-\hat{\zeta}_L^{\text{q.e.}}}\mathbf{x})$ is independent of ζ_L *despite* the consistency conditions being violated,

$$\begin{aligned} \langle \hat{\hat{\zeta}}_S \hat{\hat{\zeta}}_S \rangle' &= 4c_2 P_S(k) \zeta_L + (n_s - 1) P_S(k) \hat{\zeta}_L^{\text{q.e.}} + \mathcal{O}(\zeta_L^2) \\ &= \mathcal{O}(\zeta_L^2). \end{aligned} \tag{4.9}$$

Since our quadratic estimator is only unbiased at linear order in ζ_L when the consistency conditions apply, we will see no visible mode coupling in the power spectrum to the expected level of accuracy. Nevertheless, violations would show up in higher order correlation functions

$$\begin{aligned} \langle \hat{\hat{\zeta}}_S \hat{\hat{\zeta}}_S \hat{\hat{\zeta}}_S \hat{\hat{\zeta}}_S \rangle' &= 6c_2 P_S(k)^2 + 24 c_2^2 P_S(k)^2 \zeta_L \\ &\quad + 12 c_2 (n_s - 1) P_S(k)^2 \hat{\zeta}_L^{\text{q.e.}} + \mathcal{O}(\zeta_L^2) \\ &= 6c_2 P_S(k)^2 - 24 c_2^2 P_S(k)^2 \zeta_L + \mathcal{O}(\zeta_L^2), \end{aligned} \tag{4.10}$$

where we set $c_3 = 0$ for simplicity. Since we are only able to check mode coupling to linear order in ζ_L, this mode coupling can be made to vanish with an appropriate choice of c_3.

Ultimately, the analogy with CMB lensing is limited because we want to define a procedure that works to all orders in ζ_L rather than just linear order, as defined by the quadratic estimator. Fortunately, we can measure ζ_L directly rather than inferring it through mode coupling. With such a measurement, one can directly check the bias of the quadratic estimator as a test of the consistency conditions. A direct measurement of ζ_L can also be used to deproject ζ_S to all orders in ζ_L when the consistency conditions are satisfied. If ζ is determined by the local ansatz, then we will find that for some n, m with $m \geq 1$ and $n + m \leq 4$, such that $\langle \hat{\hat{\zeta}}_S^n \hat{\zeta}_L^m \rangle \neq 0$. Since the consistency conditions require that $\hat{\hat{\zeta}}_S(\mathbf{x})$ is statistically independent of $\hat{\zeta}_L$, the presence of any nonzero contribution defines the violation of the consistency conditions when using the deprojected modes.

The description here is an idealized description of deprojection and is more challenging to implement on real observables. In reality, we do not have the luxury of observing $\zeta(\mathbf{x})$ directly, but instead see projection effects due to redshifts, lensing, recombination, etc. [50–55]. One may hope to separate the three-dimensional projections from the consistency conditions for these other projections. Showing that this procedure can be implemented in practice is beyond the scope of this work. From a conceptual point of view, this method of deprojection highlights that the single-field consistency conditions are a statement about about the Universe for every realization of ζ_L, rather than just its statistics, and can therefore be removed realization by realization.

V. DISCUSSION

Local non-Gaussianity as parametrized by the local ansatz is a natural consequence of many scenarios that convert isocurvature fluctuations into curvature perturbations at late times. Such situations arise frequently in both multifield inflation and alternatives to inflation and are therefore compelling targets for current and future observations. Meanwhile, single-field inflation makes a very specific set of predictions for the same correlation functions that are predicted by the local ansatz. Thus, a common way of observationally distinguishing between single-field inflation and its alternatives is by measuring local non-Gaussianity parameters. For instance, the consistency conditions predict a squeezed limit bispectrum corresponding to $f_{\mathrm{NL}} = -\frac{5}{12}(n_s - 1)$ in the local ansatz and any deviation from this points to a clear violation of single-field inflation.[15]

On the other hand, the local ansatz makes statements of a fundamentally different nature than the consistency conditions, and it is not *a priori* clear that constraining local non-Gaussianity is equivalent to testing the single-field consistency conditions. In this article, we have attempted to clarify the relation between these two approaches.

First, we have shown that, while the local ansatz can reproduce, e.g., the single-field prediction for the squeezed limit bispectrum, it is impossible to agree with the consistency conditions to all orders, so that the local ansatz is in general inconsistent with single-field inflation. Thus, in principle, precision measurements of the correlation functions validating the consistency relations can rule out the local ansatz and confirm the single-field consistency conditions. This is nontrivial in the sense that by choosing coefficients carefully, the local ansatz can match the prediction of single-field inflation for any *one* correlation function. However, we have showed that there is no choice of coefficients that may satisfy *all* the conditions simultaneously. Violations must appear which are at least of order $(n_s - 1)^2$.

Secondly, we have noted that, even in multifield inflation, a weaker version of the consistency conditions persists, namely the fact that small-scale statistics should be independent of an *adiabatic* shift in the long mode. This means that, technically, the usual local ansatz is inconsistent even with multifield inflation. However, the local ansatz can be generalized in a simple way, by explicitly adding a second field (loosely identified with the isocurvature fluctuation), to make it explicitly consistent with these consistency conditions. This generalized form reduces to the usual local form in the limit where the final curvature fluctuations are dominated by the second field, and reduces to the single-field prediction in the limit where the second field is negligible.

Finally, we have suggested a novel way of testing the consistency conditions. Instead of studying a hierarchy of N-point functions, one could follow an approach analogous to delensing of the cosmic microwave background, i.e. remove the effect of the long mode from the short modes assuming the consistency conditions, and then check that the short-wavelength statistics are indeed independent of the long mode.

In practice, the minimal deviation of the local ansatz from the single-field consistency conditions is unobservably small. Nevertheless, understanding the precise predictions of these models provides an important framework for future tests of inflation and its alternatives. It is often argued that measuring $f_{\mathrm{NL}} = -\frac{5}{12}(n_s - 1)$ would confirm single-field inflation. This view has been challenged on the ground that this prediction does not require inflation but only that the short wavelength modes are statistically independent of the long wavelength modes in physical coordinates [14,24,25]. In this work, we showed that even if the mode coupling underlying this relation is "trivial" in physical coordinates, it can never be reproduced locally in space after inflation. As a consequence, any physical observable, such as scale-dependent bias, should therefore show a minimum violation of the consistency conditions in a universe governed by the local ansatz.

[15]Violations within single-field inflation are possible by violating some of the technical assumptions discussed in Sec. III [26,37–40].

ACKNOWLEDGMENTS

We thank Daniel Baumann, Rafael Porto, and Alex van Engelen for helpful discussions. D. G. was supported by an NSERC Discovery Grant and the Canadian Institute for Advanced Research. R. d. P. and O. D. acknowledge support by the Heising-Simons foundation. J. M. was supported by the Vincent and Beatrice Tremaine Fellowship. Part of the research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

[1] J. Maldacena, A model with cosmological Bell inequalities, Fortschr. Phys. **64**, 10 (2016).
[2] P. Creminelli and M. Zaldarriaga, Single field consistency relation for the 3-point function, J. Cosmol. Astropart. Phys. 10 (2004) 006.
[3] D. Baumann, D. Green, and R. A. Porto, B-modes and the nature of inflation, J. Cosmol. Astropart. Phys. 01 (2015) 016.
[4] D. Baumann, D. Green, H. Lee, and R. A. Porto, Signs of analyticity in single-field inflation, Phys. Rev. D **93**, 023523 (2016).
[5] M. Alvarez *et al.*, Testing inflation with large scale structure: Connecting hopes with reality, arXiv:1412.4671.
[6] J. M. Maldacena, Non-Gaussian features of primordial fluctuations in single field inflationary models, J. High Energy Phys. 05 (2003) 013.
[7] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, Galaxy bias and non-linear structure formation in general relativity, J. Cosmol. Astropart. Phys. 10 (2011) 031.
[8] P. Creminelli, J. Noreña, and M. Simonović, Conformal consistency relations for single-field inflation, J. Cosmol. Astropart. Phys. 07 (2012) 052.
[9] D. H. Lyth, Large-scale energy-density perturbations and inflation, Phys. Rev. D **31**, 1792 (1985).
[10] D. S. Salopek and J. R. Bond, Nonlinear evolution of long wavelength metric fluctuations in inflationary models, Phys. Rev. D **42**, 3936 (1990).
[11] S. Weinberg, Adiabatic modes in cosmology, Phys. Rev. D **67**, 123504 (2003).
[12] P. Creminelli, J. Noreña, M. Simonović, and F. Vernizzi, Single-field consistency relations of large scale structure, J. Cosmol. Astropart. Phys. 12 (2013) 025.
[13] R. de Putter, O. Doré, and D. Green, Is there scale-dependent bias in single-field inflation?, J. Cosmol. Astropart. Phys. 10 (2015) 024.
[14] L. Dai, E. Pajer, and F. Schmidt, On separate universes, J. Cosmol. Astropart. Phys. 10 (2015) 059.
[15] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XVII. Constraints on primordial non-Gaussianity, Astron. Astrophys. **594**, A17 (2016).
[16] R. de Putter and O. Doré, Designing an inflation galaxy survey: How to measure $\sigma(f_{\rm NL}) \sim 1$ using scale-dependent galaxy bias, arXiv:1412.3854.
[17] O. Doré *et al.*, Cosmology with the SPHEREX All-Sky Spectral Survey, arXiv:1412.4872.
[18] S. Ferraro and K. M. Smith, Using large scale structure to measure $f_{\rm NL}$, $g_{\rm NL}$ and $\tau_{\rm NL}$, Phys. Rev. D **91**, 043506 (2015).
[19] D. Yamauchi, K. Takahashi, and M. Oguri, Constraining primordial non-Gaussianity via a multitracer technique with surveys by Euclid and the Square Kilometre Array, Phys. Rev. D **90**, 083520 (2014).
[20] D. H. Lyth, C. Ungarelli, and D. Wands, Primordial density perturbation in the curvaton scenario, Phys. Rev. D **67**, 023503 (2003).
[21] M. Zaldarriaga, Non-Gaussianities in models with a varying inflaton decay rate, Phys. Rev. D **69**, 043508 (2004).
[22] R. de Putter, J. Gleyzes, and O. Doré, The next non-Gaussianity frontier: what can a measurement with $\sigma(f_{\rm NL}) \lesssim 1$ tell us about multifield inflation?, arXiv:1612.05248.
[23] M. Li and Y. Wang, Consistency relations for non-Gaussianity, J. Cosmol. Astropart. Phys. 09 (2008) 018.
[24] L. Senatore and M. Zaldarriaga, A note on the consistency condition of primordial fluctuations, J. Cosmol. Astropart. Phys. 08 (2012) 001.
[25] E. Pajer, F. Schmidt, and M. Zaldarriaga, The observed squeezed limit of cosmological three-point functions, Phys. Rev. D **88**, 083502 (2013).
[26] R. Flauger, D. Green, and R. A. Porto, On squeezed limits in single-field inflation. Part I, J. Cosmol. Astropart. Phys. 08 (2013) 032.
[27] E. Nelson and S. Shandera, Statistical Naturalness and Non-Gaussianity in a Finite Universe, Phys. Rev. Lett. **110**, 131301 (2013).
[28] M. LoVerde, E. Nelson, and S. Shandera, Non-Gaussian mode coupling and the statistical cosmological principle, J. Cosmol. Astropart. Phys. 06 (2013) 024.
[29] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects, Phys. Rev. D **77**, 123514 (2008).
[30] S. Matarrese and L. Verde, The effect of primordial non-Gaussianity on halo bias, Astrophys. J. Lett. **677**, L77 (2008).
[31] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, Constraints on local primordial non-Gaussianity from large scale structure, J. Cosmol. Astropart. Phys. 8 (2008) 031.
[32] V. Desjacques and U. Seljak, Primordial non-Gaussianity from the large-scale structure, Classical Quantum Gravity **27**, 124011 (2010).
[33] D. Lopez Nacir, R. A. Porto, and M. Zaldarriaga, The consistency condition for the three-point function in dissipative single-clock inflation, J. Cosmol. Astropart. Phys. 09 (2012) 004.
[34] W. D. Goldberger, L. Hui, and A. Nicolis, One-particle-irreducible consistency relations for cosmological perturbations, Phys. Rev. D **87**, 103520 (2013).
[35] M. Mirbabayi and M. Simonović, Effective theory of squeezed correlation functions, J. Cosmol. Astropart. Phys. 03 (2016) 056.
[36] L. Senatore and M. Zaldarriaga, The effective field theory of multifield inflation, J. High Energy Phys. 04 (2012) 024.
[37] M. H. Namjoo, H. Firouzjahi, and M. Sasaki, Violation of non-Gaussianity consistency relation in a single field inflationary model, Europhys. Lett. **101**, 39001 (2013).
[38] V. Assassi, D. Baumann, and D. Green, Symmetries and loops in inflation, J. High Energy Phys. 02 (2013) 151.
[39] X. Chen, H. Firouzjahi, M. H. Namjoo, and M. Sasaki, A single field inflation model with large local non-Gaussianity, Europhys. Lett. **102**, 59001 (2013).
[40] S. Mooij and G. A. Palma, Consistently violating the non-Gaussian consistency relation, J. Cosmol. Astropart. Phys. 11 (2015) 025.
[41] D. Baumann, S. Ferraro, D. Green, and K. M. Smith, Stochastic bias from non-Gaussian initial conditions, J. Cosmol. Astropart. Phys. 05 (2013) 001.
[42] T. Suyama and M. Yamaguchi, Non-Gaussianity in the modulated reheating scenario, Phys. Rev. D **77**, 023505 (2008).

[43] K. M. Smith, M. LoVerde, and M. Zaldarriaga, Universal Bound on N-Point Correlations from Inflation, Phys. Rev. Lett. **107**, 191301 (2011).
[44] V. Assassi, D. Baumann, and D. Green, On soft limits of inflationary correlation functions, J. Cosmol. Astropart. Phys. 11 (2012) 047.
[45] G. Tasinato, C. T. Byrnes, S. Nurmi, and D. Wands, Loop corrections and a new test of inflation, Phys. Rev. D **87**, 043512 (2013).
[46] A. Lewis and A. Challinor, Weak gravitational lensing of the CMB, Phys. Rep. **429**, 1 (2006).
[47] C. M. Hirata and U. Seljak, Reconstruction of lensing from the cosmic microwave background polarization, Phys. Rev. D **68**, 083002 (2003).
[48] K. M. Smith, D. Hanson, M. LoVerde, C. M. Hirata, and O. Zahn, Delensing CMB polarization with external datasets, J. Cosmol. Astropart. Phys. 06 (2012) 014.
[49] D. Green, J. Meyers, and A. van Engelen, CMB delensing beyond the B modes, arXiv:1609.08143.
[50] J. Yoo, A. L. Fitzpatrick, and M. Zaldarriaga, New perspective on galaxy clustering as a cosmological probe: General relativistic effects, Phys. Rev. D **80**, 083514 (2009).
[51] C. Bonvin and R. Durrer, What galaxy surveys really measure, Phys. Rev. D **84**, 063505 (2011).
[52] A. Challinor and A. Lewis, Linear power spectrum of observed source number counts, Phys. Rev. D **84**, 043516 (2011).
[53] M. Bruni, R. Crittenden, K. Koyama, R. Maartens, C. Pitrou, and D. Wands, Disentangling non-Gaussianity, bias and GR effects in the galaxy distribution, Phys. Rev. D **85**, 041301 (2012).
[54] D. Jeong, F. Schmidt, and C. M. Hirata, Large-scale clustering of galaxies in general relativity, Phys. Rev. D **85**, 023504 (2012).
[55] J. Yoo, N. Hamaus, U. Seljak, and M. Zaldarriaga, Testing general relativity on horizon scales and the primordial non-Gaussianity, Phys. Rev. D **86**, 063514 (2012).