

Cosmological information in perturbative forward modeling

Giovanni Cabass,¹ Marko Simonović²,[✉] and Matias Zaldarriaga¹

¹*School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540, USA*

²*Theoretical Physics Department, CERN, 1 Esplanade des Particules, Geneva 23, CH-1211, Switzerland*



(Received 9 November 2023; accepted 29 January 2024; published 20 February 2024)

We study how well field-level inference with a perturbative forward model can constrain cosmological parameters compared to conventional analyses. We exploit the fact that in perturbation theory the field-level posterior can be computed analytically in the limit of small noise. In the idealized case where the only relevant parameter for the nonlinear evolution is the nonlinear scale, we argue that information content in this posterior is the same as in the n -point correlation functions computed at the same perturbative order. In the real universe other parameters can be important, and there are possibly enhanced effects due to nonlinear interactions of long- and short-wavelength fluctuations that can either degrade the signal or increase covariance matrices. We identify several different parameters that control these enhancements and show that for some shapes of the linear power spectrum they can be large. This leads to degradation of constraints in the standard analyses, even though the effects are not dramatic for a Λ CDM-like cosmology. The aforementioned long-short couplings do not affect the field-level inference which remains optimal. Finally, we show how in these examples calculation of the perturbative posterior motivates new estimators that are easier to implement in practice than the full forward modeling but lead to nearly optimal constraints on cosmological parameters. These results generalize to any perturbative forward model, including galaxies in redshift space.

DOI: [10.1103/PhysRevD.109.043526](https://doi.org/10.1103/PhysRevD.109.043526)

I. INTRODUCTION

Traditional methods to infer cosmological parameters from large-scale structure galaxy surveys are based on measuring and analyzing the n -point correlation functions. The two-point function or its Fourier transform, the power spectrum, is the most commonly used statistics in the data analysis. Higher-order statistics, such as the three-point function or bispectrum, are employed in the data analysis less often. There are several reasons why tackling the higher-order n -point functions is difficult. They are harder to estimate from the data and predict theoretically, particularly taking into account geometry of the survey, selection functions, and other observational effects. Furthermore, the length of the data vector grows very fast with each new n -point function added, making the estimates of the covariance matrices computationally unfeasible. While some progress has been made in order to circumvent some of these issues (see e.g. [1–27]), more work has to be done before the higher-order statistics can be routinely used in the data analysis.

One alternative to using n -point functions is forward modeling [28–33].¹ In forward modeling one predicts the

full nonlinear density field of matter and galaxies, given some cosmological parameters and a realization of the initial conditions. Such a nonlinear field should be directly compared to the data, without using any summary statistics. This is nontrivial since the correct form of the likelihood is in general unknown, which motivates the use of the so-called likelihood-free inference (see for example [34–36]). In order to constrain cosmological models, one has to vary not only all cosmological and nuisance parameters but also all amplitudes and phases of the initial density field. This is clearly a formidable task, despite the recent progress in efficient sampling of large-dimensional parameter space [37–41]. However, setting these important technical challenges aside, forward modeling is argued to have a number of advantages. Most importantly, it includes all available information from the nonlinear density field, which corresponds to the information from all n -point correlation functions combined. It also allows for much easier combination of various types of data, as well as inclusion of all observational effects such as masks and selection functions, or various systematic effects, all of which can be forward modeled. In a nutshell, forward modeling is guaranteed to give the optimal constraints on cosmological parameters with an optimal combination of different datasets, assuming technical difficulties related to sampling can be overcome.

The full nonlinear forward modeling, which aims at describing distribution and properties of galaxies on very

¹While the term forward modeling is usually intended as the procedure to obtain a dark matter or biased tracer field from the linear density field, in this paper we use it interchangeably with field-level inference.

small scales, needs very advanced hydrodynamical cosmological simulations with realistic subgrid models that have to be run for each point in the immense parameter space. While this may be the ultimate way to interpret cosmological observations in the future, such data analysis is not feasible at present. A more modest but doable alternative is perturbative forward modeling. In this approach, given some realization of the initial conditions, the nonlinear density field is calculated in perturbation theory. Using the standard effective field theory methods applied to a large-scale structure [42–59], this can be done for the dark matter field and biased tracers in real or redshift space [60–70].² One important advantage of perturbative forward modeling is that it allows for rigorous definition of the field-level likelihood [71–74], which can be made arbitrarily precise on large enough scales. A lot of recent effort was dedicated to clarifying all technical aspects of this approach [75–77] and making first applications to mock data from simulations [78–83], leading to encouraging proof-of-principle results. On the other hand, perturbative forward modeling has a major limitation—it is applicable only on large scales, where perturbation theory provides a good description of the nonlinear density field. Since the importance of nonlinearities in the perturbative regime is expected to be small by construction, it is interesting to ask whether perturbative forward modeling is different from analyses based on a few leading n -point correlation functions (for some earlier work on this, see Ref. [71]).

A sharp answer can be given in a simple, Eulerian-like perturbative setup in which the only relevant parameter is the nonlinear scale k_{NL} and the variance of the density field approximated by a positive power of k/k_{NL} controls the nonlinear expansion. At a given order in perturbation theory the same nonlinear terms are used to make predictions both at the field level and for the n -point functions. It is then easy to argue that the exact same information is contained in both, provided that they are calculated at the same order. We will confirm this basic expectation using explicit calculations. However, the true nonlinear evolution in our universe is more complicated and there are effects that can change this simple picture. For example, loops in perturbation theory can lead to the appearance of new, potentially large parameters, not controlled by the variance of the density field. Such large parameters usually appear due to the averaging over the couplings of long-wavelength and short-wavelength modes, and it is in such cases that the

field-level inference can be superior to the analysis based on correlation functions.

The most well-known example are large displacements generated by the long-wavelength density fluctuations, where the large parameter is the velocity dispersion. The long modes in this example displace the short-scale fluctuations by a large amount from their initial Lagrangian positions, significantly affecting the nonlinear density field. Averaging over these large displacements when computing the two-point function famously leads to the broadening of the baryon acoustic oscillations (BAO) peak [84–86]. This can substantially degrade the measurement of the BAO scale, one of the most important cosmological parameters inferred from galaxy surveys. On the other hand, assuming that the effects of large displacements are properly taken into account, the forward modeling can recover the full linear theory information about the position of the BAO peak, as demonstrated in [80]. This is not a surprise, since the same knowledge of the long-short couplings at the map level is exploited in the BAO reconstruction algorithms, used in practice to sharpen the BAO peak and improve measurement of its position [85,87,88].

In this paper we explore less-known examples of a different type where averaging over the long-short interactions leads to degradation of errors rather than dilution of signal. Also in such cases the field-level inference can be more optimal than the conventional analysis based on the n -point functions. We identify new large parameters associated with each of these examples and show that they are related to the variance of powers of the density contrast δ . For instance, depending on the shape of the linear power spectrum, the variance of δ^2 on large scales can be very large, even if the variance of δ is small. In the analyses using n -point functions, these large parameters can lead to sizable contributions to the covariance matrices, impacting the inference of all cosmological parameters. For example, long modes can modulate the short-scale fluctuations, which can lead to a large scatter in the power spectrum on small scales once the average over the long modes is taken [2,89–92]. As we will see, this contribution to the small-scale covariance matrix is exactly controlled by the variance of δ^2 . The long-short couplings can impact the covariance on large scales as well. The short modes can couple (through the nonlinear bias) to produce the long-wavelength field of biased tracers³ with flat power spectrum. Once the short modes are averaged over, their contribution to the power spectrum and the covariance matrix (controlled again by the parameter related to the

²Note that even some hybrid forward models which rely on N -body simulations are essentially perturbative, as long as the expansion in density fluctuations is assumed in any one of the steps of producing the final galaxy density field. One typical example is models where the usual bias expansion is applied to the full nonlinear matter fields obtained from simulations; see for example [68]. All the conclusions of this work apply to such models as well.

³Note that this is not true for dark matter, since the mass and momentum conservation imply that the long modes produced by interactions of the short modes are suppressed by k^2/q^2 , where \mathbf{k} is a wave number of the long mode and \mathbf{q} a typical wave number of the short modes. This can be checked explicitly in perturbation theory. For instance, in the limit $\mathbf{q}_1 \approx -\mathbf{q}_2$ the $F_2(\mathbf{q}_1, \mathbf{q}_2)$ kernels scale as k^2/q_1^2 , where $\mathbf{k} = \mathbf{q}_1 + \mathbf{q}_2$.

variance of δ^2) is indistinguishable from the shot noise. However, depending on the observed sample, the amplitude of this noise can be much larger than the Poisson prediction, producing larger error bars than expected. Unlike in the standard analyses, none of these issues impact cosmological inference at the field level, where all relevant long-short couplings are automatically computed and taken into account.

One important feature of these examples is that unexpectedly large contributions to the covariance do not come from the fully nonlinear regime, but rather from couplings of long and short modes. Such interactions are usually easier to compute in perturbation theory or measure in simulations. Given this, instead of doing the full forward modeling, one can look for simple modified estimators for cosmological parameters which have nearly optimal variance. This is similar to the case of local primordial non-Gaussianity in the cosmic microwave background (CMB) where the optimal estimators for $f_{\text{NL}}^{\text{local}}$ have to take into account the realization of the long modes in the survey [93]. In many aspects we follow the logic of [93] and adopt it to relevant cases in a large-scale structure. One important result of our analysis is that the perturbatively calculated posterior can be used to motivate the form of optimal estimators, and we will show a couple of examples that illustrate this point.

The paper is structured as follows: in Sec. II we derive perturbative expressions for errors on cosmological parameters at the field level (with and without marginalization over linear bias); in Sec. III we compare the power spectrum plus bispectrum to forward modeling; in Sec. IV we discuss three cases where additional parameters beside the variance of the density field are present in the game, and consequently the field-level analysis can be different from standard ones; we conclude in Sec. V. Appendix A collects some nonperturbative results beyond what we derive in Sec. II; Appendix B shows how to perturbatively include a finite shot noise at the field level.

II. FROM LIKELIHOOD TO POSTERIOR- PERTURBATIVE INVERSION

In this section we show how to arrive at the expression for the full posterior and the Fisher matrix for cosmological parameters at the field level, in the limit of small noise. We apply this formula to the case where the only parameter in the theory is the nonlinear scale k_{NL} and derive explicit expressions for the posterior and the Fisher matrix at the one-loop order.

A. Posterior in the limit of small noise

Let us imagine that the forward model for the nonlinear galaxy field δ_g is known in terms of the linear density field δ . We collect cosmological and bias parameters in θ . The forward model is then given by

$$\delta_g = \delta_g[\delta, \theta] + \epsilon_g \quad (1)$$

for a given noise field ϵ_g . On large scales we can approximate the likelihood of the galaxy density field as a Gaussian

$$\begin{aligned} \mathcal{L}[\hat{\delta}_g|\delta, \theta] &= \text{normalization} \\ &\times \exp\left(-\frac{1}{2} \int_{\mathbf{k}} \frac{|\hat{\delta}_g(\mathbf{k}) - \delta_g[\delta, \theta](\mathbf{k})|^2}{P_\epsilon}\right), \end{aligned} \quad (2)$$

where P_ϵ is the noise power spectrum. This form of the likelihood can be rigorously justified in the perturbative framework that we are going to use throughout the paper. The fiducial galaxy field $\hat{\delta}_g$ is given in terms of fiducial values of the initial field $\hat{\delta}$, fiducial values of parameters $\hat{\theta}$, and a fiducial noise realization $\hat{\epsilon}_g$ by

$$\hat{\delta}_g = \delta_g[\hat{\delta}, \hat{\theta}] + \hat{\epsilon}_g. \quad (3)$$

Note that δ is an independent variable that does not depend on θ , even though it is drawn from a Gaussian distribution with the variance $P(k, \theta)$. The normalization in Eq. (2) is such that

$$\int \mathcal{D}\hat{\delta}_g \mathcal{L}[\hat{\delta}_g|\delta, \theta] = 1. \quad (4)$$

Ultimately, we are interested in the posterior for cosmological parameters given some realization of the observed galaxy density field. This posterior is obtained by integrating the likelihood multiplied by the prior on initial conditions δ . We will assume that δ has a Gaussian distribution with the variance given by the linear power spectrum $P(k)$. Defining $p(\theta)$ to be the prior on cosmological parameters, the posterior is expressed as

$$\begin{aligned} \mathcal{P}[\theta|\hat{\delta}_g] &= \text{normalization} \times \int \mathcal{D}\delta \exp\left(-\frac{1}{2} \int_{\mathbf{k}} \frac{|\delta(\mathbf{k})|^2}{P(k)}\right) \\ &\quad - \frac{1}{2} \int_{\mathbf{k}} \frac{|\hat{\delta}_g(\mathbf{k}) - \delta_g[\delta, \theta](\mathbf{k})|^2}{P_\epsilon} \times p(\theta), \end{aligned} \quad (5)$$

where we have now included the normalization of the prior in the overall factor. In what follows we will set $p(\theta) = 1$ for simplicity and having in mind situations in which constraints on cosmological parameters are dominated by the data. This choice does not change any of our conclusions, and if needed the effect of the prior $p(\theta)$ can be straightforwardly included in all our equations. Note that, contrary to our starting point, the power spectrum $P(k)$ does not depend on θ . It is easy to see why there is no loss of generality in making this assumption. Indeed, consider the case in which $P = P(k, \theta)$, which is the case of interest in cosmology. It is then possible to perform the following change of variables $\delta(\mathbf{k}) \rightarrow \delta(\mathbf{k})/P^{1/2}(k, \theta)$. In terms of the new integration variables the prior is a normalized Gaussian

with unit power spectrum, and all dependence of parameters is in the forward model. In the rest of the paper we will not need such a drastic change of variables. It is enough to use

$$\delta(\mathbf{k}) \rightarrow \tau(k, \boldsymbol{\theta})\delta(\mathbf{k}), \quad \text{with} \quad \tau(k, \boldsymbol{\theta}) = \frac{\mathcal{M}(k, \boldsymbol{\theta})}{\mathcal{M}(k, \hat{\boldsymbol{\theta}})}, \quad (6)$$

where $\mathcal{M}(k, \boldsymbol{\theta})$ is the linear transfer function which relates the primordial potential to δ [for the amplitude of the linear field A , we simply have $\tau(k, \boldsymbol{\theta}) = A/\hat{A}$]. In this way the prior for the new δ field is a normalized Gaussian with a power spectrum equal to the fiducial linear power spectrum, which exactly agrees with Eq. (5).

The main difficulty in calculating the posterior is to carry out the integral in Eq. (5). In general, this cannot be done analytically, and one has to rely on numerical sampling of the likelihood.⁴ In most of the paper we will focus on the limit of small noise, relevant for dense spectroscopic samples on large scales. In this case one can expand the posterior in P_ϵ which simplifies calculations significantly. For the time being we focus only on the leading order in the $P_\epsilon \rightarrow 0$ limit, in which the posterior becomes

$$\begin{aligned} \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] &= \text{normalization} \times \int \mathcal{D}\delta \exp\left(-\frac{1}{2} \int_k \frac{|\delta(\mathbf{k})|^2}{P(k)}\right) \\ &\times \delta_{\text{D}}^{(\infty)}(\hat{\delta}_g - \delta_g[\delta, \boldsymbol{\theta}]), \end{aligned} \quad (7)$$

where the normalization is now only that of the prior and, in the same way as $P(k)$, it does not depend on $\boldsymbol{\theta}$ (hence, we will drop it from now on to keep the notation as contained as possible). We leave the discussion of higher orders in P_ϵ for Sec. IV B. In order to exploit the delta function in the integrand of the posterior, we can do the following change of variables $\delta \rightarrow \delta_g$. The posterior can then be written as

$$\begin{aligned} \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] &= \int \mathcal{D}\delta_g \left| \frac{\partial\delta}{\partial\delta_g} \right| \exp\left(-\frac{1}{2} \int_k \frac{|\delta[\delta_g, \boldsymbol{\theta}](\mathbf{k})|^2}{P(k)}\right) \\ &\times \delta_{\text{D}}^{(\infty)}(\hat{\delta}_g - \delta_g) \equiv e^{\text{Tr} \ln J[\hat{\delta}_g, \boldsymbol{\theta}] - \frac{1}{2} \chi_{\text{prior}}^2[\hat{\delta}_g, \boldsymbol{\theta}]}, \end{aligned} \quad (8)$$

where the two terms in the final result, one coming from the prior and the other one coming from the Jacobian, are denoted by χ_{prior}^2 and J , respectively. More explicitly

⁴A simple analytical solution exists only if the forward model is linear in the initial conditions. In this case the integrand is Gaussian in δ and the integral can be solved to obtain a well-known expression for the posterior in linear theory. We also refer to [32] for a discussion of how to carry out the path integral by expanding around a saddle point found numerically. Reference [94] instead discusses the saddle point in the high noise limit.

$$\begin{aligned} \chi_{\text{prior}}^2[\hat{\delta}_g, \boldsymbol{\theta}] &\equiv \int_k \frac{|\delta[\hat{\delta}_g, \boldsymbol{\theta}](\mathbf{k})|^2}{P(k)}, \quad \text{and} \\ J[\hat{\delta}_g, \boldsymbol{\theta}] &\equiv \left| \frac{\partial\delta[\delta_g, \boldsymbol{\theta}]}{\partial\delta_g} \right|_{\delta_g=\hat{\delta}_g}. \end{aligned} \quad (9)$$

As expected, the final result depends on the realization of the galaxy density field $\hat{\delta}_g$ and parameters $\boldsymbol{\theta}$. The key ingredient needed to find the posterior $\mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g]$ is the inverse of the forward model $\delta[\delta_g, \boldsymbol{\theta}]$, which allows us to compute χ_{prior}^2 and J . Finding $\delta[\delta_g, \boldsymbol{\theta}]$ is in general a very difficult task. However, in perturbative forward modeling the inverse model is also perturbative and can be calculated analytically. The full posterior can then be consistently computed up to a given power of the variance of the density field, which resembles the more familiar loop expansion for correlation functions. Such a posterior can be used to do cosmological inference *without* the need to run Markov chain Monte Carlo, and on large scales it is guaranteed to lead to optimal constraints on cosmological parameters. We will show a few explicit perturbative examples throughout the paper.

Before proceeding, let us add two more comments about Eq. (8). First, we are assuming that the change of variables is one-to-one: in other words, we consider only the saddle point in the likelihood connected to linear theory. Other solutions will be present if the forward model is pushed to short scales (e.g. due to shell crossing; see Ref. [95] for a discussion), but on large scales the assumption of a single solution is correct. The second observation regards the change of variables itself: we invert $\delta = \delta_g^{-1}[\delta_g, \boldsymbol{\theta}]$ for varying $\boldsymbol{\theta}$. Indeed, the Dirac delta in Eq. (7) is a Dirac delta in a space of dimension equal to the number of Fourier modes of the linear field (as is clear from the expression of the Gaussian likelihood, which involves an integral in d^3k). Hence, even if $\delta_g = \hat{\delta}_g$ we do *not* obtain $\delta = \hat{\delta}$: we do so only if $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The procedure is the same as that done in Ref. [93], in which similar calculations were carried out in the context of constraints on local primordial non-Gaussianity from higher-order statistics of the CMB.

B. Posterior in the perturbative forward model

In order to derive an explicit expression for the posterior $\mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g]$, we will assume that the perturbative forward model can be written as

$$\begin{aligned} \delta_g(\mathbf{k}) &= \sum_{n=1}^{+\infty} \int_{\mathbf{p}_1, \dots, \mathbf{p}_n} (2\pi)^3 \delta_{\text{D}}^{(3)}(\mathbf{k} - \mathbf{p}_1 \dots \mathbf{p}_n) X_n \\ &\times (\boldsymbol{\theta}; \mathbf{p}_1, \dots, \mathbf{p}_n) \delta(\mathbf{p}_1) \dots \delta(\mathbf{p}_n) \equiv \sum_{n=1}^{+\infty} \delta_g^{(n)}(\mathbf{k}), \end{aligned} \quad (10)$$

where X_n are perturbation theory kernels. Note that the whole dependence on cosmological parameters is in the

kernels X_n . For example, this is the form of the nonlinear density field in Eulerian perturbation theory and for biased tracers. Since we assume that the only parameter in the theory is the variance of the density field, there are no large displacements in this ansatz. This can be achieved in practice by the appropriate choice of the power spectrum for which the velocity dispersion is small.

With our assumptions the forward model can be inverted perturbatively on large scales, i.e.

$$\begin{aligned} \delta(\mathbf{k}) &= \sum_{n=1}^{+\infty} \int_{\mathbf{p}_1, \dots, \mathbf{p}_n} (2\pi)^3 \delta_{\mathbf{D}}^{(3)}(\mathbf{k} - \mathbf{p}_1 \dots \mathbf{p}_n) Y_n \\ &\times (\boldsymbol{\theta}; \mathbf{p}_1, \dots, \mathbf{p}_n) \delta_g(\mathbf{p}_1) \cdots \delta_g(\mathbf{p}_n) \equiv \sum_{n=1}^{+\infty} \Delta_g^{[n]}(\mathbf{k}), \end{aligned} \quad (11)$$

where the Y_n kernels can be calculated in term of the original nonlinearities X_n . Up to cubic order, they are given by

$$Y_1(\boldsymbol{\theta}) = X_1^{-1}(\boldsymbol{\theta}), \quad (12a)$$

$$Y_2(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2) = -X_1^{-3}(\boldsymbol{\theta}) X_2(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2), \quad (12b)$$

$$\begin{aligned} Y_3(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3) &= \frac{2}{3} X_1^{-5}(\boldsymbol{\theta}) \left[X_2(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2 + \mathbf{p}_3) X_2(\boldsymbol{\theta}; \mathbf{p}_2, \mathbf{p}_3) \right. \\ &+ X_2(\boldsymbol{\theta}; \mathbf{p}_2, \mathbf{p}_1 + \mathbf{p}_3) X_2(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_3) \\ &+ X_2(\boldsymbol{\theta}; \mathbf{p}_3, \mathbf{p}_1 + \mathbf{p}_2) X_2(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2) \\ &\left. - \frac{3}{2} X_1(\boldsymbol{\theta}) X_3(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3) \right]. \end{aligned} \quad (12c)$$

Notice that in all the examples discussed in this paper we will consider only multiplicative parameters, for which the transfer functions, and hence, the kernels X_1 and Y_1 , are scale-independent. In more general cases, one has to keep track of the ratio of transfer functions as in Eq. (6) when deriving the inverse kernels Y_n . The inverse model defines simple operations on the galaxy density field (order by order in δ_g) which ensures the optimal combination of the data in order to recover the linear modes. At the few leading orders they are explicitly given by

$$\Delta_g^{[1]}(\mathbf{k}) = Y_1(\boldsymbol{\theta}) \delta_g(\mathbf{k}), \quad (13a)$$

$$\Delta_g^{[2]}(\mathbf{k}) = \int_{\mathbf{p}} Y_2(\boldsymbol{\theta}; \mathbf{k} - \mathbf{p}, \mathbf{p}) \delta_g(\mathbf{p}) \delta_g(\mathbf{k} - \mathbf{p}), \quad (13b)$$

$$\begin{aligned} \Delta_g^{[3]}(\mathbf{k}) &= \int_{\mathbf{p}_1, \mathbf{p}_2} Y_3(\boldsymbol{\theta}; \mathbf{k} - \mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_1, \mathbf{p}_2) \delta_g(\mathbf{p}_1) \\ &\times \delta_g(\mathbf{p}_2) \delta_g(\mathbf{k} - \mathbf{p}_1 - \mathbf{p}_2), \end{aligned} \quad (13c)$$

and they are simple convolutions of the data. Let us stress again that the whole dependence on parameters $\boldsymbol{\theta}$ is in the kernels Y_n .

It is important to point out that in order to apply these formulas we will assume that the inverse model is valid up to the similar scale as the forward model. This is not a trivial assumption. To get some intuition about possible differences between these two scales, we can focus on a simple model of spherical collapse. In this case the perturbative forward model can be written as⁵

$$\delta_{\text{sc}} = \sum_{n=1}^{\infty} \nu_n \delta_{\ell}^n, \quad (14)$$

where ν_n are spherically averaged perturbation theory kernels and δ_{ℓ} is the size of the spherical linear overdensity or underdensity in real space evolved to the present time. It is well-known that this series converges to the exact solution for $|\delta_{\ell}| < \delta_{\text{cr}} \approx 1.68$. While for $0 < \delta_{\ell} < \delta_{\text{cr}}$ the nonlinear overdensities δ_{sc} can be arbitrarily large, for $\delta_{\ell} < 0$ this series converges to the true answer only for voids with $\delta_{\text{sc}}[-\delta_{\text{cr}}] \gtrsim -0.7$. Let us now turn to the inverse perturbative model, which can be written as

$$\delta_{\ell} = \sum_{n=1}^{\infty} c_n \delta_{\text{sc}}^n, \quad (15)$$

where the coefficients c_n can be derived from ν_n . This series converges to the true linear solution for $|\delta_{\text{sc}}| \lesssim 1$, which corresponds to any $\delta_{\ell} \lesssim 0.5$. Note that the inverse perturbative solution correctly predicts δ_{ℓ} even for very empty voids, which the forward model cannot describe. We can see that the convergence properties for the forward and the inverse models in this simple example are rather different.

However, in practice we are interested in the nonlinear model which has only a small number of terms in the perturbative expansion. The validity of such a model is even more restricted compared to the full perturbative series, by the requirement that the variance of the density field is small and of order $\langle \delta_{\ell}^2 \rangle \lesssim \mathcal{O}(0.1)$. Using the spherical collapse as a toy example, we can explicitly check that with this requirement the forward and inverse models perform similarly. Up to cubic order we can write

$$\delta_{\text{sc}}^{[3]} = \delta_{\ell} + \frac{17}{21} \delta_{\ell}^2 + \frac{341}{567} \delta_{\ell}^3. \quad (16)$$

Note that this is just the spherical average of our cubic forward model in Eq. (10), assuming the standard dark matter perturbation theory kernels instead of X_n . The inverse model is then given by

⁵The usual perturbative expansion of the exact spherical collapse solution is in powers of the growth factor. We convert it here into the expansion in the linear density field evaluated at late times, such that $\delta_{\ell} = a \delta_{\text{initial}}$. At the time of collapse, the critical linear overdensity is approximately $\delta_{\ell}(a_{\text{collapse}}) = \delta_{\text{cr}} \approx 1.68$.

$$\delta_\ell^{[3]} = \delta_{\text{sc}} - \frac{17}{21}\delta_{\text{sc}}^2 + \frac{2815}{3969}\delta_{\text{sc}}^3. \quad (17)$$

It can be verified explicitly that both forward and inverse models reproduce the correct answer to better than 5%, for $|\delta_\ell| \lesssim 0.4$ and $|\delta_{\text{sc}}| \lesssim 0.4$, respectively. We use this as an indication that the inversion is valid to the similar scale as the forward model. The full check of this claim with the realistic three-dimensional (3D) fields can be done only using numerical simulations. While the details can be different due to the mode coupling, we do not expect the conclusions to change dramatically, as long as the field-level analysis is performed on scales $k_{\text{max}} \lesssim 0.1 h/\text{Mpc}$.

With the forward and inverse models at hand, we can compute the prior and Jacobian contributions to the full posterior. In this paper we want to work at an order equivalent to the one-loop calculations of the two-point function, and therefore we have to keep all nonlinearities up to cubic order. Note that this is equivalent to working with the one-loop power spectrum and the tree-level bispectrum in the conventional correlation functions approach.

Let us start from the prior term χ_{prior}^2 . Using the definitions from the previous section, it immediately follows that χ_{prior}^2 can be written as the integral over various cross spectra of $\hat{\Delta}_g^{[n]}$. Up to cubic order in the observed galaxy density field we get

$$\chi_{\text{prior}}^2 = \int_{\mathbf{k}} \frac{\hat{\Delta}_g^{[1]}(\mathbf{k})\hat{\Delta}_g^{[1]}(-\mathbf{k}) + 2\hat{\Delta}_g^{[2]}(\mathbf{k})\hat{\Delta}_g^{[1]}(-\mathbf{k}) + \hat{\Delta}_g^{[2]}(\mathbf{k})\hat{\Delta}_g^{[2]}(-\mathbf{k}) + 2\hat{\Delta}_g^{[3]}(\mathbf{k})\hat{\Delta}_g^{[1]}(-\mathbf{k})}{P(k)}. \quad (18)$$

As expected, this is a function of the data $\hat{\delta}_g$ and parameters $\boldsymbol{\theta}$ through the kernels Y_n . In order to calculate the Jacobian we have to take the derivative of the inverse model with respect to δ_g first and then set $\delta_g = \hat{\delta}_g$. In our setup this is simply given by

$$J(\mathbf{k}, \mathbf{k}') = \underbrace{Y_1(\boldsymbol{\theta})(2\pi)^3 \delta_{\text{D}}^{(3)}(\mathbf{k} - \mathbf{k}')}_{\equiv J^{(0)}(\mathbf{k}, \mathbf{k}')} + \sum_{n=1}^{+\infty} (n+1) \int_{\mathbf{p}_1, \dots, \mathbf{p}_n} (2\pi)^3 \delta_{\text{D}}^{(3)}(\mathbf{k} - \mathbf{k}' - \mathbf{p}_1 - \dots - \mathbf{p}_n) Y_{n+1}(\boldsymbol{\theta}; \mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{k}') \hat{\delta}_g(\mathbf{p}_1) \dots \hat{\delta}_g(\mathbf{p}_n). \quad (19)$$

Note that in the expression for χ_{prior}^2 [cf. Eq. (18)], we have $P(k)$ in the denominator. This tells us that we have to keep only terms up to second order in $\hat{\delta}_g$ in the Jacobian to have correct expressions at one-loop order. Ultimately, we are interested in $\text{Tr} \ln J$ since this quantity appears in the log-posterior. Taking the logarithm and keeping all terms at second order, we find

$$\begin{aligned} [\ln J(\mathbf{k}, \mathbf{k}')]^{1\text{-loop}} &= \ln J^{(0)}(\mathbf{k}, \mathbf{k}') + 2 \int_{\mathbf{p}} (2\pi)^3 \delta_{\text{D}}^{(3)}(\mathbf{k} - \mathbf{k}' - \mathbf{p}) Y_1^{-1}(\boldsymbol{\theta}) Y_2(\boldsymbol{\theta}; \mathbf{p}, \mathbf{k}') \hat{\delta}_g(\mathbf{p}) \\ &+ 3 \int_{\mathbf{p}_1, \mathbf{p}_2} (2\pi)^3 \delta_{\text{D}}^{(3)}(\mathbf{k} - \mathbf{k}' - \mathbf{p}_1 - \mathbf{p}_2) Y_1^{-1}(\boldsymbol{\theta}) Y_3(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2, \mathbf{k}') \hat{\delta}_g(\mathbf{p}_1) \hat{\delta}_g(\mathbf{p}_2) \\ &- 2 \int_{\mathbf{p}_1, \mathbf{p}_2, \mathbf{k}''} (2\pi)^3 \delta_{\text{D}}^{(3)}(\mathbf{k} - \mathbf{k}'' - \mathbf{p}_1) (2\pi)^3 \delta_{\text{D}}^{(3)}(\mathbf{k}'' - \mathbf{k}' - \mathbf{p}_2) Y_1^{-2}(\boldsymbol{\theta}) Y_2(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{k}'') Y_2(\boldsymbol{\theta}; \mathbf{p}_2, \mathbf{k}') \hat{\delta}_g(\mathbf{p}_1) \hat{\delta}_g(\mathbf{p}_2). \end{aligned} \quad (20)$$

We use the convention in which the ‘‘one-loop’’ label indicates that all terms up to one-loop order are taken into account. Finally, taking the trace leads to

$$\begin{aligned} [\text{Tr} \ln J[\hat{\delta}_g, \boldsymbol{\theta}]]^{1\text{-loop}} &= N_{\text{pix}} \ln Y_1(\boldsymbol{\theta}) + 3 Y_1^{-1}(\boldsymbol{\theta}) \int_{\mathbf{k}, \mathbf{p}} Y_3(\boldsymbol{\theta}; \mathbf{p}, -\mathbf{p}, \mathbf{k}) \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(-\mathbf{p}) \\ &- 2 Y_1^{-2}(\boldsymbol{\theta}) \int_{\mathbf{k}, \mathbf{p}} Y_2(\boldsymbol{\theta}; \mathbf{p}, \mathbf{k} - \mathbf{p}) Y_2(\boldsymbol{\theta}; -\mathbf{p}, \mathbf{k}) \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(-\mathbf{p}), \end{aligned} \quad (21)$$

where we have defined the number of pixels⁶ as $N_{\text{pix}} \equiv V \int_{\mathbf{k}}$.

⁶Notice that this is not exactly equal to $k_{\text{max}}^3/k_{\text{min}}^3$ as it would be in a box: we have $N_{\text{pix}} = (4\pi/3)(k_{\text{max}}^3/k_{\text{min}}^3) \approx 4k_{\text{max}}^3/k_{\text{min}}^3$.

In summary, we have derived the negative log-posterior for the perturbative forward model, in the limit of small noise and analytically marginalizing over the initial field δ . This can be written as

$$-\log \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] = \frac{1}{2}\chi_{\text{prior}}^2[\hat{\delta}_g, \boldsymbol{\theta}] - \text{Tr} \ln J[\hat{\delta}_g, \boldsymbol{\theta}], \quad (22)$$

where the prior and the Jacobian terms are given by Eq. (18) and Eq. (21), respectively. Our final result depends only on the observed galaxy density field $\hat{\delta}_g$, convolved with the kernels of the inverse model Y_n . In our convention, these kernels contain the *entire* dependence on cosmological and nuisance parameters $\boldsymbol{\theta}$, as well as the nonlinear dynamics. This posterior can be used for data analysis or to derive optimal estimators for a given cosmological parameter, and we will show several examples of this in the rest of the paper. Finally, one can also calculate the *averaged*

log-posterior and the corresponding Fisher matrix, given some fiducial galaxy density field $\hat{\delta}_g$. We turn to that in the next section.

C. Fisher matrix in the perturbative forward model

In order to derive the Fisher matrix for the perturbative forward modeling, we first have to compute the averaged negative log-posterior

$$\langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \rangle = \frac{1}{2}\langle \chi_{\text{prior}}^2[\hat{\delta}_g, \boldsymbol{\theta}] \rangle - \langle \text{Tr} \ln J[\hat{\delta}_g, \boldsymbol{\theta}] \rangle, \quad (23)$$

where the average is done over the fiducial initial conditions. We average the log-posterior assuming that the error on cosmological parameters does not vary a lot between realizations of the initial conditions. In the language of statistics, we are looking at the ‘‘Cram er-Rao’’ bound. The average of χ_{prior}^2 can be written as

$$\langle \chi_{\text{prior}}^2 \rangle = V \int_{\mathbf{k}} \frac{Y_1^2(\boldsymbol{\theta})\hat{P}_g(k) + P_{g,12}(\boldsymbol{\theta}; k) + P_{g,22}(\boldsymbol{\theta}; k) + P_{g,13}(\boldsymbol{\theta}; k)}{P(k)}, \quad (24)$$

where we have used $V = (2\pi)^3 \delta_D^{(3)}(\mathbf{0})$, $\hat{P}_g(k)$ is the measured galaxy power spectrum, and $P_{g,12}(\boldsymbol{\theta}; k)$, $P_{g,22}(\boldsymbol{\theta}; k)$, and $P_{g,13}(\boldsymbol{\theta}; k)$ are defined in terms of higher-order correlation functions in the following way:

$$P_{g,12}(\boldsymbol{\theta}; k) \equiv 2Y_1(\boldsymbol{\theta}) \int_{\mathbf{p}} Y_2(\boldsymbol{\theta}; \mathbf{p}, \mathbf{k} - \mathbf{p}) \langle \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(\mathbf{k}') \hat{\delta}_g(\mathbf{k} - \mathbf{p}) \rangle', \quad (25a)$$

$$P_{g,22}(\boldsymbol{\theta}; k) \equiv \int_{\mathbf{p}_1, \mathbf{p}_2} Y_2(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{k} - \mathbf{p}_1) Y_2(\boldsymbol{\theta}; \mathbf{p}_2, -\mathbf{k} - \mathbf{p}_2) \langle \hat{\delta}_g(\mathbf{p}_1) \hat{\delta}_g(\mathbf{k} - \mathbf{p}_1) \hat{\delta}_g(\mathbf{p}_2) \hat{\delta}_g(\mathbf{k}' - \mathbf{p}_2) \rangle', \quad (25b)$$

$$P_{g,13}(\boldsymbol{\theta}; k) \equiv 2Y_1(\boldsymbol{\theta}) \int_{\mathbf{p}_1, \mathbf{p}_2} Y_3(\boldsymbol{\theta}; \mathbf{p}_1, \mathbf{p}_2, \mathbf{k} - \mathbf{p}_1 - \mathbf{p}_2) \langle \hat{\delta}_g(\mathbf{p}_1) \hat{\delta}_g(\mathbf{p}_2) \hat{\delta}_g(\mathbf{k} - \mathbf{p}_1 - \mathbf{p}_2) \hat{\delta}_g(\mathbf{k}') \rangle'. \quad (25c)$$

In these expressions, the prime on the n -point functions of the observed galaxy density field $\langle \hat{\delta}_g \cdots \hat{\delta}_g \rangle'$ indicates that the overall factor of $(2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}')$ should be removed from the final result. Note that these correlation functions include the disconnected pieces. In the Fisher matrix calculation we will assume that the observed galaxy n -point functions are evaluated at the fiducial initial conditions and the fiducial parameters $\hat{\boldsymbol{\theta}}$. Already from this part of the averaged

log-posterior we can see that the Fisher matrix will depend on the higher-order n -point functions, appropriately combined with the kernels of the inverse model.

It is instructive to write down explicitly contributions to the prior at leading order in perturbation theory, which is sufficient for the one-loop forward model. Keeping the tree-level bispectrum and disconnected parts of the four-point function and using kernels for the forward model, we get

$$\hat{P}_g^{1\text{-loop}}(k) = X_1^2(\hat{\boldsymbol{\theta}})P(k) + 2 \int_{\mathbf{p}} X_2^2(\hat{\boldsymbol{\theta}}; \mathbf{p}, \mathbf{k} - \mathbf{p})P(p)P(|\mathbf{k} - \mathbf{p}|) + 6X_1(\hat{\boldsymbol{\theta}})P(k) \int_{\mathbf{p}} X_3(\hat{\boldsymbol{\theta}}; \mathbf{p}, -\mathbf{p}, \mathbf{k})P(p), \quad (26a)$$

$$P_{g,12}^{1\text{-loop}}(\boldsymbol{\theta}; k) = 4Y_1(\boldsymbol{\theta})X_1^2(\hat{\boldsymbol{\theta}}) \int_{\mathbf{p}} Y_2(\boldsymbol{\theta}; \mathbf{p}, \mathbf{k} - \mathbf{p})(2X_2(\hat{\boldsymbol{\theta}}; \mathbf{p}, -\mathbf{k})P(p)P(k) + X_2(\hat{\boldsymbol{\theta}}; \mathbf{p}, \mathbf{k} - \mathbf{p})P(p)P(|\mathbf{k} - \mathbf{p}|)), \quad (26b)$$

$$P_{g,22}^{1\text{-loop}}(\boldsymbol{\theta}; k) = 2X_1^4(\hat{\boldsymbol{\theta}}) \int_{\mathbf{p}} Y_2^2(\boldsymbol{\theta}; \mathbf{p}, \mathbf{k} - \mathbf{p})P(p)P(|\mathbf{k} - \mathbf{p}|), \quad (26c)$$

$$P_{g,13}^{1\text{-loop}}(\boldsymbol{\theta}; k) = 6Y_1(\boldsymbol{\theta})X_1^4(\hat{\boldsymbol{\theta}})P(k) \int_{\mathbf{p}} Y_3(\boldsymbol{\theta}; \mathbf{p}, -\mathbf{p}, \mathbf{k})P(p). \quad (26d)$$

Note that we do not include those contractions that lead to the results proportional to $(2\pi)^3 \delta_D^{(3)}(\mathbf{k})$, which contribute only to the unobservable zero mode. At leading order in perturbation theory the fiducial galaxy power spectrum was evaluated as $\hat{P}_g(k) = X_1^2(\hat{\theta})P(k)$. We are going to use these

equations to evaluate the Fisher matrix for the forward model for biased tracers in real space, up to one-loop order in perturbation theory. We next turn to evaluating $\langle \ln \det J \rangle = \langle \text{Tr} \ln J \rangle$. Using results of the previous section, we get

$$\begin{aligned} \langle \text{Tr} \ln J \rangle^{1\text{-loop}} &= N_{\text{pix}} \ln Y_1(\boldsymbol{\theta}) + 3V X_1^2(\hat{\boldsymbol{\theta}}) Y_1^{-1}(\boldsymbol{\theta}) \int_{k,p} Y_3(\boldsymbol{\theta}; \mathbf{p}, -\mathbf{p}, k) P(p) \\ &\quad - 2V X_1^2(\hat{\boldsymbol{\theta}}) Y_1^{-2}(\boldsymbol{\theta}) \int_{k,p} Y_2(\boldsymbol{\theta}; \mathbf{p}, \mathbf{k} - \mathbf{p}) Y_2(\boldsymbol{\theta}; -\mathbf{p}, \mathbf{k}) P(p). \end{aligned} \quad (27)$$

This completes our derivation of the averaged log-posterior at one-loop, and one can proceed by calculating the Fisher matrix. Before we do that, let us make two comments. First, note that a nontrivial check of the previous

equations is unbiasedness. If the formulas are correct, then the derivative of the averaged log-posterior at the fiducial values of parameters must be zero. Calculating this derivative explicitly and using Eq. (12a) to Eq. (12c), we find

$$\frac{\partial}{\partial \boldsymbol{\theta}} \langle -\ln \mathcal{P} \rangle^{1\text{-loop}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{2V}{Y_1^4(\hat{\boldsymbol{\theta}})} \int_{k,p} \left(Y_2(\hat{\boldsymbol{\theta}}; \mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{\partial}{\partial \boldsymbol{\theta}} Y_2(\boldsymbol{\theta}; -\mathbf{p}, \mathbf{k}) - Y_2(\hat{\boldsymbol{\theta}}; -\mathbf{p}, \mathbf{k}) \frac{\partial}{\partial \boldsymbol{\theta}} Y_2(\boldsymbol{\theta}; \mathbf{p}, \mathbf{k} - \mathbf{p}) \right) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} P(p). \quad (28)$$

After a simple change of variables $\mathbf{k} - \mathbf{p} \rightarrow -\mathbf{k}$ in one of the terms under the integral, the right-hand side vanishes. This implies that the estimate of the cosmological parameters at the field level is indeed unbiased. Note that in order to be able to do the change of variables, it is crucial that both integrals run over all possible values of the momenta. On the other hand, we always have some maximum wave number up to which we can trust perturbation theory. In order to preserve the unbiasedness and implement this cutoff in practice, one can always apply the window function on the power spectrum, rather than changing the limits of integration.

The second comment is about higher orders in perturbation theory. So far we have discussed only the leading nonlinearities, but it is important to stress how a clear loop expansion in $\langle -\ln \mathcal{P} \rangle$ arises if one wants to go further. The negative log-posterior is expressed in terms of the fiducial initial conditions via a combination of Eq. (10) and Eq. (11). Schematically, the solution of $\delta = \delta_g^{-1}[\delta_g, \boldsymbol{\theta}]$ can be written in terms of the fiducial linear field $\hat{\delta}$ as follows:

$$\delta = \sum_{m=1}^{+\infty} Y_m(\boldsymbol{\theta}) \left(\sum_{n=1}^{+\infty} X_n(\hat{\boldsymbol{\theta}}) \hat{\delta}^n \right)^m = \sum_{k=1}^{+\infty} Z_k(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) \hat{\delta}^k, \quad (29)$$

where, by construction, the new kernels $Z_k(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ satisfy the following properties:

$$Z_1(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) = 1 \quad \text{and} \quad Z_{k>1}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) = 0. \quad (30)$$

In other words, evaluated at the fiducial values of parameters, the linear field in the inverse model must be equal to the

fiducial linear field. One can explicitly check that this is indeed the case for the inverse kernels derived above (at one-loop order). Using Eq. (19), a similar expression in terms of $\hat{\delta}$ can be found for the Jacobian as well. Therefore, the negative log-posterior is naturally organized as a perturbative series in the fiducial linear density field $\hat{\delta}$. It follows that the expectation value $\langle -\ln \mathcal{P} \rangle$ has a clear loop expansion, the same one as in the standard perturbation theory. To evaluate the averaged log-posterior at the given order in the loop expansion, only a finite number of terms in the forward and inverse models have to be kept. Note, however, that due to the linear power spectrum in the denominator in the expression for $\langle \chi_{\text{prior}}^2 \rangle$ and the functional derivative in the Jacobian, the leading order term in $\langle -\ln \mathcal{P} \rangle$ starts at zeroth order in $P(k)$. This is expected, since in the linear theory, as we will see shortly, the posterior depends only on the number of pixels.

With all these results at hand, it is straightforward to calculate the Fisher matrix F , given some fiducial galaxy density field $\hat{\delta}_g$. In practice, we evaluate

$$\begin{aligned} F_{\alpha\beta}^{1\text{-loop}} &= \frac{\partial^2}{\partial \theta_\alpha \partial \theta_\beta} \left(\frac{1}{2} \langle \chi_{\text{prior}}^2[\hat{\delta}_g, \boldsymbol{\theta}] \rangle^{1\text{-loop}} \right. \\ &\quad \left. - \langle \text{Tr} \ln J[\hat{\delta}_g, \boldsymbol{\theta}] \rangle^{1\text{-loop}} \right) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \end{aligned} \quad (31)$$

where the first term is given by Eq. (24) and the second term by Eq. (27), both evaluated up to one-loop order, and the expectation value refers to averaging over fiducial initial conditions as discussed under Eq. (23). It is important to

point out that this formula makes the connection between perturbative forward modeling and the standard analyses manifest. As long as the variance of the density field is the only relevant parameter, the two approaches are equivalent, order by order in perturbation theory. Before considering more interesting situations in which new, potentially large parameters play an important role, we show how the field-level Fisher matrix works in practice in several examples of interest.

III. APPLICATIONS OF THE FIELD-LEVEL POSTERIOR AND FISHER MATRIX

In this section we apply Eq. (22) and Eq. (31) and show explicitly in some simple setups that the field-level analysis leads to the same errors on cosmological parameters as the standard analysis based on the n -point correlation functions, as long as the only relevant parameter in the theory is the nonlinear scale k_{NL} and the only expansion parameter is the variance of the density field.

A. Simple examples of the field-level Fisher matrix: Linear and nonlinear dark matter

Let us begin with two very simple examples: linear theory and nonlinear dark matter field. In order to keep expressions as clear as possible, we will focus on the amplitude of the linear density field A as the only cosmological parameter of interest.

Let us first consider the simplest possible scenario, in which the evolution of density fluctuations is linear. In this example the kernels of the forward model are trivial

$$X_1 = A \quad \text{and} \quad X_{n>1} = 0, \quad (32)$$

which implies

$$Y_1 = 1/A \quad \text{and} \quad Y_{n>1} = 0. \quad (33)$$

This simplifies the form of the posterior significantly, and we can write (note that we keep using δ_g for the nonlinear field even though we are not considering galaxies here)

$$-\ln \mathcal{P}[A|\hat{\delta}_g] = \frac{1}{2A^2} \int_{\mathbf{k}} \frac{\hat{\delta}_g(\mathbf{k})\hat{\delta}_g(-\mathbf{k})}{P(k)} + N_{\text{pix}} \ln A. \quad (34)$$

Finding the maximum of the posterior, we get the optimal estimator for A ,

$$\mathcal{E} = \frac{1}{N_{\text{pix}}} \int_{\mathbf{k}} \frac{\hat{\delta}_g(\mathbf{k})\hat{\delta}_g(-\mathbf{k})}{P(k)}. \quad (35)$$

Not surprisingly, the optimal estimator for the amplitude of fluctuations in linear theory coincides with the estimator of the power spectrum. The averaged negative log-posterior is given by

$$\langle -\ln \mathcal{P} \rangle = \frac{N_{\text{pix}}}{2A^2} + N_{\text{pix}} \ln A, \quad (36)$$

where we have used $\hat{P}_g(k) = \hat{A}^2 P(k)$ and the fiducial value of the amplitude is set to one, $\hat{A} = 1$. In the absence of the nonlinear evolution, $\langle -\ln \mathcal{P} \rangle$ depends only on the number of pixels N_{pix} . Using this equation one can explicitly show that the estimate of the amplitude is unbiased and that the error is given by the well-known formula for the linear theory⁷

$$\frac{1}{\sigma_A^2} = \left. \frac{\partial^2 \langle -\ln \mathcal{P} \rangle}{\partial A^2} \right|_{A=\hat{A}} = 2N_{\text{pix}}. \quad (38)$$

A slightly more nontrivial example is the nonlinear evolution, where the amplitude of the linear field A is still the only unknown parameter. It is easy to see that in this case the kernels are given by

$$X_n(A) = \frac{A^n}{\hat{A}^n} X_n(\hat{A}) \quad \text{and} \quad Y_n(A) = \frac{\hat{A}}{A} Y_n(\hat{A}). \quad (39)$$

Note that all inverse kernels scale as $1/A$. This has two important consequences. One is that χ^2_{prior} in Eq. (18) scales exactly as $1/A^2$. The other is that only the first term in Eq. (21) depends on A . The full posterior can then be written as (setting $\hat{A} = 1$)

$$-\ln \mathcal{P}[A|\hat{\delta}_g] = \frac{1}{2A^2} \int_{\mathbf{k}} \frac{|\delta[\hat{\delta}_g, \hat{A} = 1](\mathbf{k})|^2}{P(k)} + N_{\text{pix}} \ln A + (A\text{-independent terms}). \quad (40)$$

The optimal estimator for the amplitude of density fluctuations is

$$\mathcal{E} = \frac{1}{N_{\text{pix}}} \int_{\mathbf{k}} \frac{|\delta[\hat{\delta}_g, \hat{A} = 1](\mathbf{k})|^2}{P(k)}. \quad (41)$$

We can see that in this case the estimator for the amplitude does not depend only on the nonlinear power spectrum, but also various other combinations of data which enter the numerator and which are explicitly written in Eq. (18). However, by definition of the inverse model, when evaluated at the fiducial values of parameters, it gives the initial Fourier modes of $\hat{\delta}$ given the observed nonlinear density

⁷The reader is perhaps more familiar with the formula for the error of the amplitude of the power spectrum A_s . Changing variables from A to $A_s \equiv A^2$ we find

$$\sigma_{A_s}^2 = \frac{2}{N_{\text{pix}}} \quad (37)$$

as expected.

field. This means that the amplitude of the matter fluctuations can still be optimally measured from the power spectrum only, but also of the initial field reconstructed from $\hat{\delta}_g$. Given this, we expect the error on A to be the same as in the linear theory. Indeed, the averaged log-posterior is given by

$$\langle -\ln \mathcal{P} \rangle = \frac{N_{\text{pix}}}{2A^2} + N_{\text{pix}} \ln A + (A\text{-independent terms}), \quad (42)$$

which implies the same error as before

$$\frac{1}{\sigma_A^2} = \left. \frac{\partial^2 \langle -\ln \mathcal{P} \rangle}{\partial A^2} \right|_{A=\hat{A}} = 2N_{\text{pix}}. \quad (43)$$

Such a result makes sense since for a fixed nonlinear model, and in the absence of noise the field-level posterior should contain all available information on cosmological parameters from the data optimally combined. In such an ideal setup, this amount of information saturates the bound given by the linear theory. While it is difficult to prove this statement for generic cosmological parameters following the approach of this section, this can be done using the nonperturbative formulation of the posterior at the field level. We defer this general analysis to Appendix A.

Let us make a comment about the derivation above. In the case of an imperfect inverse model, we can imagine that the Y_n kernels are different from their expression of Eq. (12), but with the same overall scaling with A intact. The argument leading to Eq. (40) behaving as $1/A^2$ still holds, with the crucial difference that now all the nonlinear terms in the numerator under the integral will *not* combine to the reconstruction of the linear field. As a consequence, the momentum integral in Eq. (24) multiplied by the volume will not lead to N_{pix} , and the error bar on A would be different than in the linear theory, as expected if the inversion is wrong.⁸

B. Including the linear bias

Going beyond these idealized examples, things become more complicated. In the real universe, our ability to measure cosmological parameters depends on their degeneracies in the linear power spectrum, peculiarities of the nonlinear evolution, and complexity of galaxy formation which is on large scales encoded in a number of nuisance parameters one must marginalize over. It is interesting to show how the field-level inference boils down to the conventional analysis with n -point functions even in this case.

In the simplest and observationally most relevant setup in which one can still gain some intuition in an analytically tractable way, we allow two parameters: the amplitude A of the linear density field and the linear bias b_1 . We keep the product Ab_1 fixed (for simplicity, we assume that \hat{A} and \hat{b}_1 are both equal to 1). This is inspired by the fact that the overall amplitude of the power spectrum of biased tracers is usually very well measured on large scales and such an assumption does not affect our conclusions. As it is well-known, since this particular combination multiplies the linear power spectrum, all information on the amplitude A in this example must come from the nonlinearities. Here we compute the posterior and the Fisher matrix for A in forward modeling. For $Ab_1 = 1$, perturbation theory kernels scale as

$$X_1(A) = 1 \quad \text{and} \quad X_n(A; \mathbf{p}_1, \dots, \mathbf{p}_n) \equiv A^{n-1} X_n(\mathbf{p}_1, \dots, \mathbf{p}_n), \quad (44)$$

and

$$Y_1(A) = 1 \quad \text{and} \quad Y_n(A; \mathbf{p}_1, \dots, \mathbf{p}_n) \equiv A^{n-1} Y_n(\mathbf{p}_1, \dots, \mathbf{p}_n). \quad (45)$$

Note that here (and the rest of the paper) the kernels without the explicit dependence on A are evaluated at the fiducial value $\hat{A} = 1$. The explicit form of the posterior for A is a bit more complicated in this example. The prior part is given by

$$\begin{aligned} \frac{1}{2} \chi_{\text{prior}}^2 &= \frac{1}{2} \int_{\mathbf{k}} \frac{\hat{\delta}_g(\mathbf{k}) \hat{\delta}_g(-\mathbf{k})}{P(k)} + A \int_{\mathbf{k}, \mathbf{p}} Y_2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{\hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(\mathbf{k} - \mathbf{p}) \hat{\delta}_g(-\mathbf{k})}{P(k)} \\ &+ \frac{1}{2} A^2 \int_{\mathbf{k}, \mathbf{p}_1, \mathbf{p}_2} Y_2(\mathbf{k} - \mathbf{p}_1, \mathbf{p}_1) Y_2(-\mathbf{k} - \mathbf{p}_2, \mathbf{p}_2) \frac{\hat{\delta}_g(\mathbf{p}_1) \hat{\delta}_g(\mathbf{k} - \mathbf{p}_1) \hat{\delta}_g(\mathbf{p}_2) \hat{\delta}_g(-\mathbf{k} - \mathbf{p}_2)}{P(k)} \\ &+ A^2 \int_{\mathbf{k}, \mathbf{p}_1, \mathbf{p}_2} Y_3(\mathbf{k} - \mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_1, \mathbf{p}_2) \frac{\hat{\delta}_g(-\mathbf{k}) \hat{\delta}_g(\mathbf{p}_1) \hat{\delta}_g(\mathbf{p}_2) \hat{\delta}_g(\mathbf{k} - \mathbf{p}_1 - \mathbf{p}_2)}{P(k)}. \end{aligned} \quad (46)$$

⁸Interestingly, things are different for the Jacobian part of $\langle -\ln \mathcal{P} \rangle$. Even if the Y_n kernels are wrong but their overall scaling with A is correct, $-\text{Tr} \ln J = N_{\text{pix}} \ln A + (A\text{-independent terms})$ will continue to hold. This is a coincidence for the amplitude of the density fluctuations, and it is not true in general for other cosmological parameters.

The Jacobian term is simpler, and it has the following form:

$$\text{Tr} \ln J = 3A^2 \int_{\mathbf{k}, \mathbf{p}} Y_3(\mathbf{p}, -\mathbf{p}, \mathbf{k}) \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(-\mathbf{p}) - 2A^2 \int_{\mathbf{k}, \mathbf{p}} Y_2(\mathbf{p}, \mathbf{k} - \mathbf{p}) Y_2(-\mathbf{p}, \mathbf{k}) \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(-\mathbf{p}). \quad (47)$$

Note that the leading terms that correspond to the linear theory do not depend on A , in agreement with the expectation that all information on the amplitude of δ comes from the nonlinearities. Given the simple dependence of the log-posterior on A , it is possible to explicitly write down the optimal estimator

$$\mathcal{E} = -\frac{1}{2 \langle \text{quadr} \rangle} \int_{\mathbf{k}, \mathbf{p}} Y_2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{\hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(\mathbf{k} - \mathbf{p}) \hat{\delta}_g(-\mathbf{k})}{P(k)}. \quad (48)$$

The numerator is the only term in the log-posterior linear in A , while in the denominator we collect all other quadratic and quartic combinations of data $\hat{\delta}_g$ which are all proportional to A^2 . More precisely, we write

$$(-\ln \mathcal{P}) = (-\ln \mathcal{P})_{(0)} + A(-\ln \mathcal{P})_{(1)} + A^2 \underbrace{(-\ln \mathcal{P})_{(2)}}_{\equiv \text{quadr}} + \dots \quad (49)$$

Given the realization of data both the numerator and the denominator can be computed easily since they are just numbers. This means that in practice one can use the exact posterior to find the constraints on cosmological parameters.

However, in order to get a better understanding of the optimal estimator and simplify equations, it is convenient to replace the denominator by its average, assuming that it does not vary significantly between the different realizations of data. The modified estimator is given by

$$\tilde{\mathcal{E}} = -\frac{1}{2 \langle \text{quadr} \rangle} \int_{\mathbf{k}, \mathbf{p}} Y_2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{\hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(\mathbf{k} - \mathbf{p}) \hat{\delta}_g(-\mathbf{k})}{P(k)}. \quad (50)$$

From now on, we will use the tilde to denote such ‘‘simplified’’ estimators. In our example, the explicit form of the denominator evaluated at leading order in perturbation theory is

$$\langle \text{quadr} \rangle = V \int_{\mathbf{k}, \mathbf{p}} \left[Y_2^2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{P(p)P(|\mathbf{k} - \mathbf{p}|)}{P(k)} + 2Y_2(\mathbf{p}, \mathbf{k} - \mathbf{p}) Y_2(-\mathbf{p}, \mathbf{k}) P(p) \right]. \quad (51)$$

Note that the contributions to prior and Jacobian with cubic kernels Y_3 exactly cancel when taking the average. It is then easy to explicitly check that the estimator is unbiased, $\langle \tilde{\mathcal{E}} \rangle = 1$, calculating the tree-level galaxy bispectrum and

remembering that $X_2 = -Y_2$ in this example. Finally, computing the variance of $\tilde{\mathcal{E}}$ or using the one-loop expression for the Fisher matrix from the previous section, we find that the error on A is given by

$$\frac{1}{\sigma_A^2} = 2V \int_{\mathbf{k}, \mathbf{p}} \left[X_2^2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{P(p)P(|\mathbf{k} - \mathbf{p}|)}{P(k)} + 2X_2(\mathbf{p}, \mathbf{k} - \mathbf{p}) X_2(-\mathbf{p}, \mathbf{k}) P(p) \right]. \quad (52)$$

Note that $1/\sigma_A^2 = 2 \langle \text{quadr} \rangle$. This is expected since $\langle \text{quadr} \rangle$ is the expectation value of all the terms in the negative log-posterior which are proportional to A^2 and therefore equal to the Fisher matrix for the amplitude A .

Three comments are in order. First, it is clear from the expression for the error that the degeneracy between b_1 and A is broken only by nonlinearities. The right hand side of Eq. (52) has the typical size of $V \int_{\mathbf{k}} P^{1\text{-loop}}(k)/P(k)$. This is an explicit example in which we can see that the loop counting in forward modeling works the same way as in the conventional analyses, as discussed above. Since we are always working at the one-loop order, we expect σ_A^2 to be exactly the same as the error in the standard joint one-loop power spectrum and tree-level bispectrum analysis, as we will demonstrate soon.

Second, the X_2 kernel in this setup is equal to the F_2 kernel of standard perturbation theory. It follows that the first of the two terms in Eq. (52) is nothing but the usual $P_{22}(k)$ diagram [divided by $P(k)$ and integrated in $d^3k/(2\pi)^3$]. It is well-known that $P_{22}(k)$ has very large contributions from soft loops at high k . This is due to the shifts in F_2 , whose variance is large in a Λ CDM-like cosmology. However, the same shifts would be present even in a more generic example, where other cosmological or biased parameters are considered, and the conclusion below would not change in such more general setup. Naively, large contributions to P_{22} from soft loops would lead to a very small error bar on A . In the standard calculation of the one-loop power spectrum these large contributions are usually canceled by the P_{13} diagram. For the field-level posterior we expect similar cancellation to happen. In our case, this is ensured by the second term in Eq. (52). To see this explicitly, let us remember that the two infrared contributions ($p \rightarrow 0$ and $|\mathbf{k} - \mathbf{p}| \rightarrow 0$) from $P_{22}(k)$ give

$$F_2^2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{P(p)P(|\mathbf{k} - \mathbf{p}|)}{P(k)} \rightarrow 2 \frac{(\mathbf{p} \cdot \mathbf{k})^2}{4p^4} P(p), \quad p \ll k. \quad (53)$$

The only infrared contribution from the Jacobian part in Eq. (52), instead, is

$$2F_2(\mathbf{p}, \mathbf{k} - \mathbf{p})X_2(-\mathbf{p}, \mathbf{k})P(p) \rightarrow 2 \frac{(\mathbf{p} \cdot \mathbf{k})}{2p^2} \frac{(-\mathbf{p} \cdot \mathbf{k})}{2p^2} P(p) = -2 \frac{(\mathbf{p} \cdot \mathbf{k})^2}{4p^4} P(p), \quad p \ll k, \quad (54)$$

which exactly cancels Eq. (53). One important lesson that we learn from this result is that one cannot use large displacements to break degeneracies between cosmological and nuisance parameters, e.g. linear bias and A , even though at the map level large displacements produce a large effect and by the equivalence principle they are proportional only to the amplitude of the fluctuations for any type of tracer. The reason is that the posterior is expressed as a combination of data which on average are equal to the summary statistics of the observed nonlinear field. Therefore, the effects off the displacements cancel in the final result for the error on cosmological parameters in the same way they do in equal-time correlators [96–101]. On the other hand, if the initial conditions are fixed and known, as is the case in simulations, the amplitude of the linear density field can be indeed measured from the displacements [72]. Note that this is not only the consequence of the cosmic variance cancellation for the known initial conditions. Even if the biases are unknown, the nontrivial dynamics involving large shifts is what allows one to measure A .

The third comment regards the positivity of $1/\sigma_A^2$. In Fig. 1 we show that the square bracket in Eq. (52) is very similar to the one-loop matter power spectrum, $P_{1\text{-loop}}(k)$. This quantity is negative at low k , so one might worry that this affects the sign of $1/\sigma_A^2$. However, we must recall

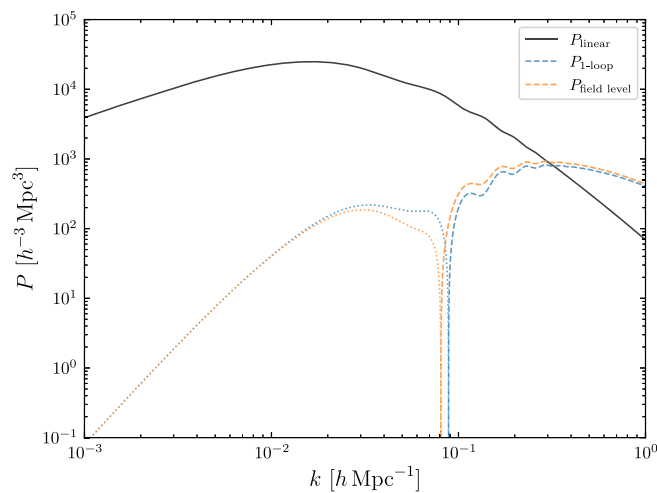


FIG. 1. Integrand in the square brackets of Eq. (52) (“ $P_{\text{field level}}$ ”) compared with the one-loop standard perturbation theory power spectrum for a standard Λ CDM cosmology.

that all our derivations in the previous section assume integration over all the modes. Therefore, we only need to show that the large- k limit of the square bracket in Eq. (52) is positive, whatever the form of the power spectrum. If this is true, then $1/\sigma_A^2$ is positive-definite. The discussion about infrared safety helps us to confirm this. Let us expand the square bracket at next-to-leading order in $1/k$, since the leading order vanishes. More precisely, let us consider only the (angle-averaged) expansion of the kernels. We have

$$4 \int d\mu F_2^2(\mathbf{p}, \mathbf{k} - \mathbf{p}) = \frac{2}{3} \frac{k^2}{q^2} + \frac{1138}{735} + \mathcal{O}\left(\frac{q^2}{k^2}\right), \quad (55a)$$

$$4 \int d\mu F_2(\mathbf{p}, \mathbf{k} - \mathbf{p})F_2(-\mathbf{p}, \mathbf{k}) = -\frac{2}{3} \frac{k^2}{q^2} + \frac{152}{147} + \mathcal{O}\left(\frac{q^2}{k^2}\right), \quad (55b)$$

where $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{p}}$. Hence, we see that the contribution to the square bracket at large k takes the form

$$\frac{1898}{735} P(k) \int \frac{dpp^2}{(2\pi)^2} P(p), \quad (56)$$

which is manifestly positive for any cosmology.

In practice, one always has to use some finite cutoff k_{max} . However, in that case the expression in square brackets in Eq. (52) is also modified, since the same cutoff has to be applied in each integration over momenta. One can show that the error remains positive for any choice of k_{max} as long as the cutoff is implemented consistently.

C. Comparison to the joint power spectrum and bispectrum analysis

It is instructive to compare results of the previous section with the conventional joint power spectrum and bispectrum analysis and show explicitly that the likelihood and the Fisher matrix for cosmological parameters are the same. To that end, we can compute the posterior and the Fisher matrix for the standard analysis and compare it with the field-level results in Eq. (50) and Eq. (52). Let us begin with the power spectrum. In the setup where $Ab_1 = 1$, the galaxy power spectrum at leading order in perturbation theory can be schematically written as

$$P_g(k) = P(k) + A^2 P_{1\text{-loop}}(k). \quad (57)$$

As expected, the only information on the amplitude of the density fluctuations comes from the nonlinearities. How well can we measure A this way? We can estimate this using the Fisher matrix. Using the Gaussian covariance, it is given by the well-known expression

$$\begin{aligned} \left(\frac{1}{\sigma_A^2}\right)_P &= \frac{V}{2} \int_k \left(\frac{\partial P_g(k)}{\partial A} \frac{1}{P^2(k)} \frac{\partial P_g(k)}{\partial A} \right) \Big|_{A=1} \\ &= 2V \int_k \left(\frac{P_{1\text{-loop}}(k)}{P(k)} \right)^2. \end{aligned} \quad (58)$$

This result is rather different from the error in Eq. (52), which schematically looks like

$$\frac{1}{\sigma_A^2} \approx 2V \int_k \frac{P_{1\text{-loop}}(k)}{P(k)}. \quad (59)$$

Clearly, the signal in the power spectrum analysis is suppressed by the variance of the density field compared to the field-level result. In the perturbative setup this means that the one-loop power spectrum does not carry significant information about the amplitude of the density fluctuations once we marginalize over b_1 . This can be checked numerically choosing k_{max} to be in the mildly nonlinear regime and computing the Fisher matrix. As a consequence, the leading information in this setup is expected to be only in the bispectrum.

In order to confirm this expectation, we can explicitly calculate the bispectrum likelihood. Using the Gaussian covariance we can write⁹

$$\chi_B^2 = \frac{V}{6} \int_{\mathbf{q}_1, \mathbf{q}_2} \frac{[\hat{B}_g(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_1 - \mathbf{q}_2) - B_g(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_1 - \mathbf{q}_2)]^2}{P(q_1)P(q_2)P(|\mathbf{q}_1 + \mathbf{q}_2|)}, \quad (62)$$

where the bispectrum estimator is

$$\hat{B}_g(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_1 - \mathbf{q}_2) = \frac{1}{V} \hat{\delta}_g(\mathbf{q}_1) \hat{\delta}_g(\mathbf{q}_2) \hat{\delta}_g(-\mathbf{q}_1 - \mathbf{q}_2), \quad (63)$$

and the theoretical model (remembering that $Ab_1 = 1$) is

$$B_g(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = 2AX_2(\mathbf{q}_1, \mathbf{q}_2)P(q_1)P(q_2) + 2 \text{ perms.} \quad (64)$$

⁹Note that introducing the finite size bins of width Δk and replacing the integrals with the sum over triangles, we can rewrite the likelihood in a more familiar form

$$\chi_B^2 = \sum_T \frac{(\hat{B}_g(T) - B_g(T))^2}{C_B(T)}, \quad (60)$$

where the sum runs over all triangles $T = \{k_1, k_2, k_3\}$ such that $k_1 \geq k_2 \geq k_3$. The Gaussian covariance for the bispectrum is

$$C_B(k_1, k_2, k_3) = \frac{(2\pi)^6 S_{\text{shape}}}{VV_{123}} P(k_1)P(k_2)P(k_3), \quad (61)$$

where $V_{123} = 8\pi^2 k_1 k_2 k_3 \Delta k^3$ and $S_{\text{shape}} = 6, 2, 1$ for equilateral, isosceles and scalene triangles, respectively.

Note that the bispectrum likelihood is Gaussian in A , with the variance

$$\left(\frac{1}{\sigma_A^2}\right)_B = \frac{V}{6} \int_{\mathbf{q}_1, \mathbf{q}_2} \frac{B_g^2(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_1 - \mathbf{q}_2)}{P(q_1)P(q_2)P(|\mathbf{q}_1 + \mathbf{q}_2|)}. \quad (65)$$

Using the explicit form of the theoretical model and keeping all the permutations, it is easy to show that

$$\begin{aligned} \left(\frac{1}{\sigma_A^2}\right)_B &= 2V \int_{\mathbf{k}, \mathbf{p}} \left[X_2^2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{P(p)P(|\mathbf{k} - \mathbf{p}|)}{P(k)} \right. \\ &\quad \left. + 2X_2(\mathbf{p}, \mathbf{k} - \mathbf{p})X_2(-\mathbf{p}, \mathbf{k})P(p) \right]. \end{aligned} \quad (66)$$

This precisely agrees with Eq. (52), confirming the expectation that the error bar in the field analysis is the same as in the (power spectrum and) bispectrum analysis at leading order in perturbation theory. Furthermore, this equivalence can be checked for the value of the best fit parameter as well. Using the bispectrum likelihood we get

$$\begin{aligned} \mathcal{E}_B &= \sigma_A^2 \cdot \frac{V}{6} \int_{\mathbf{q}_1, \mathbf{q}_2} \frac{\hat{B}_g(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_1 - \mathbf{q}_2) \cdot B_g(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_1 - \mathbf{q}_2)}{P(q_1)P(q_2)P(|\mathbf{q}_1 + \mathbf{q}_2|)} \\ &= \sigma_A^2 \cdot \int_{\mathbf{k}, \mathbf{p}} X_2(\mathbf{p}, \mathbf{k} - \mathbf{p}) \frac{\hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(\mathbf{k} - \mathbf{p}) \hat{\delta}_g(-\mathbf{k})}{P(k)}, \end{aligned} \quad (67)$$

which is again identical to the field-level result for the modified estimator $\tilde{\mathcal{E}}$ in Eq. (50). In conclusion, we have shown that the perturbative forward modeling recovers the same information about the amplitude of density fluctuations as the few leading correlation functions, computed at the same order in perturbation theory.

Even though this result was derived in a simple and analytically tractable case of $Ab_1 = 1$, our conclusions hold more generally. If the variance of the density field is the only expansion parameter in the theory, one can show, order by order in perturbation theory, that the field-level and the n -point function based inferences of cosmological parameters are equivalent for any cosmological parameter, although the explicit demonstration is more involved. If other, potentially large parameters are present in the theory, this simple picture can change. We turn to these more interesting situations next.

IV. BEYOND THE SIMPLE PERTURBATIVE MODEL

So far we have focused on a universe in which the only relevant scale for the nonlinear evolution is the nonlinear scale k_{NL} and the only small expansion parameter is the variance of the density field. However, the real universe can be more complicated, and depending on the shape of the linear power spectrum other scales can play an important role in the nonlinear dynamics. The most well-known example is the parameter related to the velocity dispersion

which is responsible for the broadening of the BAO peak. This parameter is given by [102,103]

$$\Sigma^2 = \frac{1}{6\pi^2} \int_0^{k_{\text{NL}}} dq P(q) [1 - j_0(q\ell_{\text{BAO}}) + 2j_2(q\ell_{\text{BAO}})], \quad (68)$$

where ℓ_{BAO} is the BAO scale. Note that the combination of spherical Bessel functions in the square brackets is such that it scales as q^2 in the limit $q \ll \ell_{\text{BAO}}^{-1}$. Therefore, neglecting the contribution given by the variance of the density field on the BAO scale, we can approximate the previous expression as

$$\Sigma^2 \approx \frac{1}{6\pi^2} \int_{\ell_{\text{BAO}}^{-1}}^{k_{\text{NL}}} dq P(q). \quad (69)$$

The integral is dominated by the peak of the power spectrum, which in a Λ CDM-like cosmology is at the equality scale k_{eq} . However, given that for our universe $\ell_{\text{BAO}}^{-1} \approx k_{\text{eq}}$, we will keep the BAO scale as the lower boundary of the integral, having in mind a more general power spectrum for which the maximum in principle can be at much smaller k . Approximating the linear power spectrum as a power law such that $P(k) \approx P(k_{\text{NL}})(k/k_{\text{NL}})^{-n}$, we can estimate Σ^2 as

$$\Sigma^2 \approx \frac{1}{k_{\text{NL}}^2} \frac{1}{3} (k_{\text{NL}} \ell_{\text{BAO}})^{n-1}. \quad (70)$$

Depending on the slope of the power spectrum and position of the BAO peak, the enhancement $(k_{\text{NL}} \ell_{\text{BAO}})^{n-1}$ can be large. In Λ CDM this is not dramatic, since we have $n \approx 1.8$ and therefore $\Sigma \approx 2k_{\text{NL}}^{-1}$ at redshift zero. However, note that

$$\frac{1}{3} (k_{\text{NL}} \ell_{\text{BAO}})^{n-1} \approx 4 \left(\frac{D^2(z)}{D^2(0)} \right)^{\frac{1-n}{3-n}}, \quad (71)$$

such that Σ compared to k_{NL}^{-1} grows at larger redshifts for $1 < n < 3$.

The existence of a large parameter controlled by some infrared scale implies a possible breakdown of the simple Eulerian-like forward model that we discussed in the previous section. Indeed, a simple one-loop calculation is known to poorly describe the shape of the BAO peak in the nonlinear two-point correlation function. It is well-understood that in order to circumvent this issue one has to either employ Lagrangian perturbation theory [45,53,55,60,104] or modify predictions of the Eulerian perturbation theory through the so-called infrared resummation [102,103,105–108]. Therefore, the measurement of the BAO scale is a well-known counterexample to our statement in the previous section. The field-level inference of ℓ_{BAO} is indeed more optimal compared to the measurement from the two-point correlation function [80].

However, we will not further discuss this example here for two reasons. First, it is rather special, since large displacements affect only features in the two-point correlation function and the only parameter that is impacted is ℓ_{BAO} . All other cosmological parameters are unaffected and our general conclusions still apply. More explicitly, the average log-posterior is expressed in terms of the n -point functions of the data, whose smooth part is not impacted by the large displacements [96–101]. The second reason is that the simple BAO reconstruction schemes [85,87,88] recover almost optimal information on ℓ_{BAO} , making the full forward modeling unnecessary. One may still do the forward modeling of the reconstructed field, but since in this case the significant fraction of large displacements is canceled, this is much closer to the regime that we discussed in the previous section and our conclusions remain valid. It would be interesting to check this explicitly in numerical simulations, and we leave it for future work.

For the rest of this section we will focus instead on different types of situations in which forward modeling can be more optimal. Unlike the BAO where the long-short interactions dilute the signal, in these examples these interactions increase the error. More precisely, they lead to large covariance matrices for the n -point functions, making the standard analyses suboptimal. Importantly, this affects *all* cosmological parameters. We will show how these situations arise and what are the new large parameters associated with them.

A. Large covariance matrix from long-wavelength fluctuations

In order to see how the standard analysis can be suboptimal, we can already use the simplest example of nonlinear dark matter field. We have shown in the previous section that in this case the optimal estimator for A is

$$\mathcal{E} = \frac{1}{N_{\text{pix}}} \int_{\mathbf{k}} \frac{|\delta[\hat{\delta}_g, \hat{A} = 1](\mathbf{k})|^2}{P(k)}, \quad (72)$$

where the numerator has various combinations of the data $\hat{\delta}_g$ dictated by the inverse model. In order to highlight new relevant parameters, in this section we will assume that the nonlinearities controlled by the variance of the density field are very small. In this limit the inverse model is well approximated by the linear term $\delta = Y_1 \delta_g$ and the higher-order loop contributions are expected to be small. The approximate estimator valid in such a regime is given by

$$\tilde{\mathcal{E}} = \frac{1}{N_{\text{pix}}} \int_{\mathbf{k}} \frac{|\hat{\delta}_g(\mathbf{k})|^2}{P(k)}, \quad (73)$$

assuming $Y_1 = 1$. Such a result is not surprising. This is the estimator of the nonlinear power spectrum, which in the limit of small nonlinearities gives the correct estimate of A .

While such a simplified estimator (which we will interchangeably call “naive” and “simplified” in the following) may lead to correct amplitude of the linear power spectrum, its variance may be large. We can easily compute that

$$\text{var}(\tilde{\mathcal{E}}) = \frac{2V}{N_{\text{pix}}^2} \int_k \frac{P_g^2(k)}{P^2(k)} + \frac{V}{N_{\text{pix}}^2} \int_{k,k'} \frac{T_g(\mathbf{k}, -\mathbf{k}, \mathbf{k}', -\mathbf{k}')}{P(k)P(k')}, \quad (74)$$

where T_g is a connected four-point function of the data (in the following, we will drop the hat on correlation functions of the data for simplicity of notation). Going beyond the leading result given by the linear theory, a simple estimate of the one-loop power spectrum in the first term and tree-level trispectrum in the second term both lead to

$$\text{var}(\tilde{\mathcal{E}}) = \frac{2}{N_{\text{pix}}} [1 + \mathcal{O}(\Delta^2(k_{\text{max}}))], \quad (75)$$

where $\Delta^2(k_{\text{max}})$ is the variance of the density field at k_{max} used in the analysis. As one may expect, corrections to the linear theory error bars are small at leading order in perturbation theory. However, going to the one-loop trispectrum in $\text{var}(\tilde{\mathcal{E}})$ something unexpected happens. Due to the particular momentum configuration there are one-loop contributions that schematically look like

$$T_g(\mathbf{k}, -\mathbf{k}, \mathbf{k}', -\mathbf{k}') \supset P(k)P(k') \times \int_q P^2(q) \text{ in the limit } q \ll k, k'. \quad (76)$$

This is a well-known result that can be explicitly derived in perturbation theory [2,89–92] and holds even in the nonlinear regime if one uses nonlinear responses to compute the covariance matrix [92,109]. Importantly, such contributions to the covariance matrix are *not* controlled by the variance of the density field as one may naively expect. Instead, they are proportional to the new parameter—the variance of δ^2 . In general, we can define the following dimensionless quantity:

$$\sigma_{n,-}^2 \equiv \left(\frac{k_{\text{NL}}^3}{2\pi^2} \right)^{n-1} \int_{q < k_{\text{NL}}} P^n(q), \quad (77)$$

such that the ratio of the one-loop and the tree-level trispectrum contributions to the covariance is given by

$$\frac{T_g^{1\text{-loop}}}{T_g^{\text{tree}}} \approx \sigma_{2,-}^2, \quad (78)$$

for $k \sim k_{\text{NL}}$. The minus sign in our notation for $\sigma_{n,-}^2$ indicates that momenta which $P^n(k)$ is integrated over are smaller than some scale, which we chose to be k_{NL} .

The existence of a new parameter that controls the loop expansion is indeed surprising, but it is the consequence of

a specific momentum configuration of the trispectrum that contributes to the covariance matrix. If all momenta in an n -point function are different, these parameters never appear. Importantly, $\sigma_{2,-}^2$ can be very large, even when the variance of the density field is small. To see this explicitly, let us consider a power-law universe with the IR cutoff k_* , which can be given by the size of the survey or can mimic the equality scale in a Λ CDM-like universe. In this simplified setup, the power spectrum is given by

$$P(k) = \frac{2\pi^2(3-n)}{k_{\text{NL}}^3} \left(\frac{k}{k_{\text{NL}}} \right)^{-n} \theta(k - k_*), \quad (79)$$

with $3/2 < n < 3$.

Note that the slope of the power spectrum in the Λ CDM cosmology at the nonlinear scale is approximately $n \approx 2$, which is in the range we consider. The variance of the density field is given by

$$\Delta^2(k) = \frac{2\pi^2(3-n)}{k_{\text{NL}}^3} \frac{1}{2\pi^2} \int_{k_{\text{min}}}^k dq q^2 \left(\frac{q}{k_{\text{NL}}} \right)^{-n} \approx \left(\frac{k}{k_{\text{NL}}} \right)^{3-n}, \quad (80)$$

where for $3/2 < n < 3$ we have neglected the lower bound of the integral. As usual, the power spectrum is normalized such that $\Delta^2(k_{\text{NL}}) = 1$. Also, the variance is smaller than 1 on perturbative scales, and it is growing with k for our choice of n . We can now explicitly compute $\sigma_{2,-}^2$ and find

$$\sigma_{2,-}^2 = \frac{(3-n)^2}{2n-3} \left(\frac{k_*}{k_{\text{NL}}} \right)^{3-2n}. \quad (81)$$

Note that for our choice $3/2 < n < 3$ the integral in $\sigma_{2,-}^2$ is dominated in the infrared, and in this estimate we neglected the upper bound. For $n \approx 2$ we have $\sigma_{2,-}^2 \approx k_{\text{NL}}/k_*$, which can be much larger than 1. This large parameter can significantly modify the variance of the estimator for A . Following Eq. (74) we have

$$\begin{aligned} \text{var}(\tilde{\mathcal{E}}) &= \frac{2}{N_{\text{pix}}} \left(1 + \frac{N_{\text{pix}}}{2V} \int_q P^2(q) \right) \\ &= \frac{2}{N_{\text{pix}}} \left(1 + \pi^2 \frac{N_{\text{pix}}}{V k_{\text{NL}}^3} \sigma_{2,-}^2 \right). \end{aligned} \quad (82)$$

Using $N_{\text{pix}} = V k_{\text{max}}^3 / 6\pi^2$, we can estimate the correction to the linear theory error as

$$\frac{N_{\text{pix}}}{2V} \int_q P^2(q) = \pi^2 \frac{N_{\text{pix}}}{V k_{\text{NL}}^3} \sigma_{2,-}^2 = \frac{1}{6} \left(\frac{k_{\text{max}}}{k_{\text{NL}}} \right)^3 \sigma_{2,-}^2. \quad (83)$$

For a Λ CDM-like cosmology where $k_* \approx k_{\text{eq}} \approx 0.02 h/\text{Mpc}$ and $k_{\text{NL}} \approx 0.3 h/\text{Mpc}$, this number at redshift zero is

$$\frac{N_{\text{pix}}}{2V} \int_q P^2(q) = \mathcal{O}(1) \times \left(\frac{k_{\text{max}}}{0.3 \text{ h/Mpc}} \right)^3, \quad (84)$$

assuming $n \approx 2$. We can see that for a Λ CDM-like universe the correction is not large on perturbative scales and in a realistic data analysis would be subdominant to other sources of error, such as marginalization over nuisance parameters [110]. However, it is easy to imagine a universe where the situation is very different. For instance, setting k_* to be the fundamental mode of the survey, we can write the correction to the linear theory errors as

$$\frac{N_{\text{pix}}}{2V} \int_q P^2(q) = \frac{(3-n)^2}{6(2n-3)} \left(\frac{k_{\text{max}}}{k_{\text{NL}}} \right)^{6-2n} \left(\frac{3N_{\text{pix}}}{4\pi} \right)^{\frac{2n}{3}-1}. \quad (85)$$

If the power-law power spectrum extends to arbitrarily large scales, for any k_{max} there is a volume big enough when the number of Fourier modes makes this number bigger than one. Setting $n \approx 2$ and $k_{\text{max}} \approx 0.1 \text{ h/Mpc}$ and $k_{\text{NL}} \approx 0.3 \text{ h/Mpc}$ we find

$$\text{var}(\tilde{\mathcal{E}}) = \frac{2}{N_{\text{pix}}} (1 + \mathcal{O}(0.01) \times N_{\text{pix}}^{1/3}). \quad (86)$$

For a hypothetical survey with more than one million pixels, the second term would dominate the covariance. The errors

would then scale only as $N_{\text{pix}}^{2/3}$. In the extreme case of the nearly scale-invariant power spectrum with $n \approx 3$, the error bars would improve only logarithmically as in the case of local non-Gaussianities [93]. This clearly indicates that the naive estimator is suboptimal. Let us point out that correcting the naive estimator is not necessary in order to get a detection of A : for $n \approx 2$, the variance of $\tilde{\mathcal{E}}$ is much less than one for the typical number of pixels in modern galaxy surveys. However, we see that by using $\tilde{\mathcal{E}}$ instead of \mathcal{E} we could be paying a very high price on the error bars. In contrast, as we have shown in the previous section, the full field-level analysis leads to the expected optimal result.

The appearance of large parameters $\sigma_{n,-}^2$ in the variance of “naive” estimators, where one takes a simplified version of the optimal estimators, is quite generic. For example, let us consider a case of a linearly biased tracer with unknown linear bias b_1 discussed in Sec. III B. The simplified estimator is given by Eq. (50):

$$\tilde{\mathcal{E}} = -\sigma_A^2 \int_{k,p} Y_2(\mathbf{p}, -\mathbf{k} - \mathbf{p}) \frac{\hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(-\mathbf{k} - \mathbf{p}) \hat{\delta}_g(\mathbf{k})}{P(k)}, \quad (87)$$

where we have used Eqs. (51) and (52) and the simple relation between Y_2 and X_2 . Computing $\langle \tilde{\mathcal{E}}^2 - 1 \rangle$, we see that the variance of $\tilde{\mathcal{E}}$ contains a piece

$$\text{var}(\tilde{\mathcal{E}}) \supset (\sigma_A^2)^2 V \int_{k,k',p} \frac{P(p)}{P(k)P(k')} T_g(\mathbf{k}, \mathbf{k}', -\mathbf{k} - \mathbf{p}, -\mathbf{k}' + \mathbf{p}) X_2(-\mathbf{k} - \mathbf{p}, \mathbf{p}) X_2(-\mathbf{k}' + \mathbf{p}, -\mathbf{p}). \quad (88)$$

The two relevant contributions in the trispectrum are

$$T_g(\mathbf{k}, \mathbf{k}', -\mathbf{k} - \mathbf{p}, -\mathbf{k}' + \mathbf{p}) \supset P(|\mathbf{k} + \mathbf{p}|) P(|\mathbf{k}' - \mathbf{p}|) P(p) X_2(\mathbf{k} + \mathbf{p}, \mathbf{p}) X_2(\mathbf{k}' - \mathbf{p}, \mathbf{p}) \quad (89)$$

and

$$T_g(\mathbf{k}, \mathbf{k}', -\mathbf{k} - \mathbf{p}, -\mathbf{k}' + \mathbf{p}) \supset P(k) P(k') P(p) X_2(-\mathbf{k}, -\mathbf{p}) X_2(-\mathbf{k}', \mathbf{p}). \quad (90)$$

We see that both, in the limit $p \ll k, k'$, give rise to the large parameter $\sigma_{2,-}^2$, which can make the variance of the naive estimator much larger than expected.

It is instructive to discuss another possible contribution to the variance of the naive estimator, i.e. the one coming from the fully connected six-point function

$$\text{var}(\tilde{\mathcal{E}}) \supset (\sigma_A^2)^2 \int_{k,k',p,p'} \frac{V \langle \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(\mathbf{p}') \hat{\delta}_g(-\mathbf{k} - \mathbf{p}) \hat{\delta}_g(-\mathbf{k}' - \mathbf{p}') \hat{\delta}_g(\mathbf{k}) \hat{\delta}_g(\mathbf{k}') \rangle'}{P(k)P(k')}. \quad (91)$$

For a particular momentum configuration where the pairs of momenta are opposite ($\mathbf{p}' = -\mathbf{p}$ and $\mathbf{k}' = -\mathbf{k}$), we expect that the one-loop contribution to the six-point function is controlled by $\sigma_{3,-}^2$. In the universe where large parameters $\sigma_{n,-}^2$ are infrared-dominated, we expect that $\sigma_{3,-}^2 \gg \sigma_{2,-}^2$ and this would be the leading contribution to the variance. This expectation is, however, wrong since

$\sigma_{3,-}^2$ appears only for special arrangements of momenta, unlike in the case of the estimator in Eq. (73) where the large parameter $\sigma_{2,-}^2$ exists for *all* momentum configurations of the four-point function. We can check this by explicitly evaluating the six-point function. Taking $X_2 = 1$ for simplicity (this does not change our conclusions), we get

$$\begin{aligned}
 \text{var}(\tilde{\mathcal{E}}) &\supset \frac{(\sigma_A^2)^2}{V} \int_{\mathbf{k}, \mathbf{k}', \mathbf{p}, \mathbf{p}'} \frac{(2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') (2\pi)^3 \delta_D^{(3)}(\mathbf{p} + \mathbf{p}')}{P(k)P(k')} P(p)P(|\mathbf{k} + \mathbf{p}|)P(k) \int_q P^3(q) \\
 &\approx \frac{(\sigma_A^2)^2}{V} \underbrace{\int_{\mathbf{k}, \mathbf{p}} \frac{P(p)P(|\mathbf{k} + \mathbf{p}|)}{P(k)}}_{\approx \frac{1}{V\sigma_A^2}} \int_q P^3(q) \approx \frac{\sigma_A^2}{V^2} \int_q P^3(q), \tag{92}
 \end{aligned}$$

where we have isolated the part of the six-point function that can give rise to $\sigma_{3,-}^2$ with two Dirac delta functions $(2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}')/V$ and $(2\pi)^3 \delta_D^{(3)}(\mathbf{p} + \mathbf{p}')/V$. Crucially, the presence of the factor $1/V^2$ from these two Dirac deltas guarantees that even in the worst-case scenario of a scale-invariant power spectrum with $n = -3$, this contribution to the variance is never large: indeed, we would have $\sigma_{3,-}^2 \sim V^2$ so that the variance would scale as $1/N_{\text{pix}}$, i.e. much faster than the $\sigma_{2,-}^2$ contribution discussed above; see e.g. Eq. (85).

While this argument shows that $\sigma_{3,-}^2$ does not produce the leading contribution to the variance of $\tilde{\mathcal{E}}$, it also suggests to us that it will enter its skewness, where the connected six-point function automatically appears in the configuration where all momenta are ‘‘pinched’’ two-by-two. We can see this already for the simple estimator of Eq. (73) for the nonlinear dark matter, i.e.

$$\tilde{\mathcal{E}} = \frac{1}{N_{\text{pix}}} \int_{\mathbf{k}} \frac{|\hat{\delta}_g(\mathbf{k})|^2}{P(k)}. \tag{93}$$

The skewness of the estimator is given by

$$\text{skewness}(\tilde{\mathcal{E}}) = \frac{\langle (\tilde{\mathcal{E}} - 1)^3 \rangle}{\sigma_A^3}, \tag{94}$$

where

$$\begin{aligned}
 \langle (\tilde{\mathcal{E}} - 1)^3 \rangle &= \frac{8}{N_{\text{pix}}^2} + \frac{1}{N_{\text{pix}}^3} \int_{\mathbf{k}, \mathbf{k}', \mathbf{k}''} \\
 &\times \frac{\langle \hat{\delta}_g(\mathbf{k}) \hat{\delta}_g(-\mathbf{k}) \hat{\delta}_g(\mathbf{k}') \hat{\delta}_g(-\mathbf{k}') \hat{\delta}_g(\mathbf{k}'') \hat{\delta}_g(-\mathbf{k}'') \rangle_c}{P(k)P(k')P(k'')}. \tag{95}
 \end{aligned}$$

Note that in the second term we have to keep only the connected six-point function. Naively, this contribution is suppressed by $\Delta^4(k_{\text{max}})$, but similar to the case of the variance, there is a one-loop contribution to the connected six-point function which is much bigger and can be estimated as

$$\begin{aligned}
 \langle (\tilde{\mathcal{E}} - 1)^3 \rangle &= \frac{8}{N_{\text{pix}}^2} \left(1 + \frac{1}{8} \int_{\mathbf{k}, \mathbf{k}'} \int_q P^3(q) \right) \\
 &= \frac{8}{N_{\text{pix}}^2} \left(1 + \frac{1}{72} \left(\frac{k_{\text{max}}}{k_{\text{NL}}} \right)^6 \sigma_{3,-}^2 \right). \tag{96}
 \end{aligned}$$

This is the analog of Eq. (82) for the variance of $\tilde{\mathcal{E}}$. We can see that the new large parameter

$$\sigma_{3,-}^2 = \frac{(3-n)^3}{3n-3} \left(\frac{k_*}{k_{\text{NL}}} \right)^{3-3n} \tag{97}$$

appears, making the second contribution potentially much bigger than the first one. Assuming the perfect power-law power spectrum and setting k_* to be the fundamental mode of the survey, we can estimate the skewness as

$$\text{skewness}(\tilde{\mathcal{E}}) = \frac{2\sqrt{2}}{\sqrt{N_{\text{pix}}}} (1 + \mathcal{O}(10^{-5}) \times N_{\text{pix}}), \tag{98}$$

for typical numbers we used before: $n \approx 2$ and $k_{\text{max}} \approx 0.1 h/\text{Mpc}$ and $k_{\text{NL}} \approx 0.3 h/\text{Mpc}$. In this hypothetical universe, even for $\mathcal{O}(10^5)$ pixels, corrections to the naive result become significant. For steeper power spectra, the corrections are even more pronounced. However, as before, in ΛCDM -like cosmology, these effects are never large on perturbative scales.

It is worth emphasizing that so far we have considered the case where $k_{\text{max}} < k_{\text{NL}}$, such that for ΛCDM -like cosmologies the impact of $\sigma_{2,-}^2$ or $\sigma_{3,-}^2$ is not dramatic. However, if $k_{\text{max}} > k_{\text{NL}}$, even in ΛCDM there can be large contributions to the variance of the power spectrum estimator. Weak lensing is the natural place where we would expect this effect to show up.¹⁰ In that case it is worth exploring alternative, more optimal estimators. We will come back to this in Sec. IV C.

B. Large noise from mildly nonlinear scales

In the previous section we have shown how long-wavelength fluctuations can lead to a possibly large variance of naive estimators on small scales. It is interesting

¹⁰Notice that here we are disregarding other effects that could prove to dominate the error budget in realistic scenarios, e.g. the supersample covariance.

to ask whether the opposite can happen. For dark matter the answer is no, since due to momentum conservation the impact of small-scale fluctuations on large scales is strongly suppressed. However, for biased tracers the situation is different. Let us consider a galaxy density field where nonlinearities induced by biasing and the shot noise P_e are very small. In this limit the variance of the naive estimator of Eq. (74) for the amplitude of the linear power spectrum can be written as

$$\text{var}(\tilde{\mathcal{E}}) = \frac{2V}{N_{\text{pix}}^2} \int_k \frac{P_g^2(k)}{P^2(k)}, \quad (99)$$

where we are neglecting the trispectrum contribution assuming that the power spectrum is such that $\sigma_{2,-}^2$ is not a large number. The galaxy power spectrum is given by (assuming fiducial values of the amplitude and b_1 to be 1)

$$P_g(k) = P(k) + P_e + \frac{b_2^2}{2} \int_q P(q)P(|\mathbf{k} - \mathbf{q}|) + \dots, \quad (100)$$

where the ellipses denote the other one-loop terms that are all suppressed by $\Delta^2(k_{\text{max}})$. Even in the limit $P_e \rightarrow 0$ that we consider in this paper, there is an effective noise at large scales given by the third term in the previous equation. Defining

$$\mathcal{I}_{\delta^2\delta^2}(k) \equiv 2 \int_q P(q)P(|\mathbf{k} - \mathbf{q}|), \quad (101)$$

we can see that the large-scale limit of the effective noise is given by the variance of δ^2 . It is important to stress that this term can dominate the total noise even in a vanilla Λ CDM-like cosmology. Some well-known examples are small-mass dark matter halos [62,111] or neutral hydrogen [66,70,112]. In these examples the amplitude of the noise on large scales can be even 10 times larger than the naive Poisson expectation, and using the field-level methods one can show that this noise comes exactly from the quadratic bias nonlinearities we discuss here [62,66,70]. In a conventional analysis such a large noise contributes to the covariance matrix, and it can lead to larger errors. On the other hand, the field-level analysis remains optimal.

While it is possible to have a tracer such that $\mathcal{I}_{\delta^2\delta^2}(k) \gg P_e$ on all scales of interest, this does not automatically imply a dramatic difference in the errors between conventional analyses and forward modeling. The reason is that the errors can still be dominated by the cosmic variance. To see this explicitly, we can write

$$\begin{aligned} \mathcal{I}_{\delta^2\delta^2}(k) &= 2P(k) \int_{q<k} P(q) + 2 \int_{q>k} P^2(q) \\ &= 2P(k) \int_{q<k} P(q) + \frac{\sigma_{2,+}^2(k)}{k_{\text{NL}}^3}, \end{aligned} \quad (102)$$

where we have defined

$$\sigma_{2,+}^2(k) \equiv 2k_{\text{NL}}^3 \int_{q>k} P^2(q). \quad (103)$$

This new parameter is similar to $\sigma_{2,-}^2$ but with some very important differences. For instance, it is important to note that this parameter depends on scale. While the integral is still infrared dominated, the range of integration has a natural IR cutoff at the scale of interest k . This is a consequence of the fact that for each k we are considering the effects of shorter modes that combine to produce the effective noise on larger scales. The $+$ sign in the definition indicates that one has to calculate the contribution to the variance of δ^2 coming only from modes larger than k . In a simple power-law universe, we can estimate

$$\sigma_{2,+}^2(k) = 2k_{\text{NL}}^3 P(k) \frac{3-n}{2n-3} \left(\frac{k}{k_{\text{NL}}}\right)^{3-n} \approx 2k_{\text{NL}}^3 P(k) \Delta^2(k). \quad (104)$$

Note that this implies that

$$\mathcal{I}_{\delta^2\delta^2}(k) = 2P(k) \int_{q<k} P(q) + \frac{\sigma_{2,+}^2(k)}{k_{\text{NL}}^3} \approx 4P(k) \Delta^2(k), \quad (105)$$

While $\sigma_{2,+}^2(k)$ can be much bigger than one, its contribution to the variance of the estimator is controlled by the ratio of $\mathcal{I}_{\delta^2\delta^2}(k)$ and $P(k)$, which is small. More precisely

$$\text{var}(\tilde{\mathcal{E}}) \approx \frac{2V}{N_{\text{pix}}^2} \int_k (1 + b_2^2 \Delta^2(k))^2. \quad (106)$$

Depending on the value of b_2 , the linear theory variance can be increased by $\mathcal{O}(1)$ for $k_{\text{max}} \approx k_{\text{NL}}$. While these effects are not parametrically large, it would be worth exploring the possibility to reduce the effective noise using the field-level analysis in the future.

One might wonder whether higher-order terms in the bias expansion can contribute with new large parameters to the effective noise. For example, including $(b_3/6) \times \delta^3$ in the forward model leads to the following two-loop contribution to the shot noise:

$$\frac{b_3^2}{6} \int_{p,q} P(|\mathbf{k} - \mathbf{p} - \mathbf{q}|) P(p) P(q). \quad (107)$$

However, we see that in the low- k limit this contribution is suppressed with respect to $\mathcal{I}_{\delta^2\delta^2}(k)$ by the variance of the density field on large scales, which is a small parameter. This is generically true for higher-order contributions.

Let us finish this section by pointing out one interesting result that follows from this discussion. Given that at the field level we can predict the realization of the long-wavelength fluctuations that lead to the effective noise, we can also use these fluctuations in the data to infer cosmological and nuisance parameters. In particular, we

can measure b_2 from the amplitude of the effective noise. We show this explicitly in Appendix B. Note that this is very different from the conventional power spectrum analysis. There, following the standard prescription for renormalized bias, b_2 can be measured only from a shape of loop corrections which is different from the flat power spectrum of the noise. This suggest that there is a different, potentially more optimal renormalization scheme, in which information on bias parameters can be obtained from the amplitude of the effective noise even in the power spectrum analysis. We leave exploration of this interesting possibility for future work.

C. Simple nearly optimal estimators

We have seen in previous sections how in different situations standard power spectrum and bispectrum estimators can have large variances due to various large parameters which emerge from averaging over interactions of long-wavelength and short-wavelength modes. On the other hand, in forward modeling, all such interactions are explicitly taken into account and the analysis is always optimal. We have shown that for the simple perturbative model and in the limit of small noise, the posterior can be calculated analytically, and for any cosmological parameter of interest, one can find the optimal estimator in terms of simple operations on the data $\hat{\delta}_g$. However, our equations do not apply in all regimes of interest. One example where the variance of naive estimators can be large due to $\sigma_{2,-}^2$ and where one would benefit from doing the full field-level analysis is measurement of A_s from small scales beyond k_{NL} . Since our perturbative equations do not apply there, one would have to do the full forward modeling which is technically challenging. In this section we would like to show that it is possible to find a middle ground and keep the simplicity of the conventional analyses by using modified, forward-model-inspired estimators that are nearly optimal.

In order to see how to construct these nearly optimal estimators, we will first take a closer look at how the large parameters cancel in the field-level analysis. We will focus on $\sigma_{2,-}^2$ since it can have the largest impact in practice. First, in the case of the nonlinear dark matter, the variance of the optimal estimator of Eq. (72) scales as

$$\text{var}(\mathcal{E}) = \frac{2}{N_{\text{pix}}} [1 + \mathcal{O}(\Delta^4(k_{\text{max}}))], \quad (108)$$

i.e. without $\sigma_{2,-}^2$ appearing. This is manifest by the virtue of the inverse model which is correct at one-loop order. We emphasize that including Y_3 in the inverse model is crucial: $\sigma_{2,-}^2$ can appear in the one-loop trispectrum also from diagrams involving the X_3 kernel of the forward model.

A similar cancellation of $\sigma_{2,-}^2$ must happen in the case of a linearly biased tracer discussed in Sec. III B, even though it is a bit more difficult to show it explicitly. For simplicity,

we focus on the contribution of Eq. (89) to the variance of the simplified estimator $\text{var}(\tilde{\mathcal{E}})$ for the amplitude of the power spectrum. Let us recall that this contribution makes the variance of the estimator behave as (for a power-law universe with the spectral index $n \approx 2$)

$$\text{var}(\tilde{\mathcal{E}}) = \frac{2}{N_{\text{pix}}} (1 + \mathcal{O}(0.01) \times N_{\text{pix}}^{1/3}). \quad (109)$$

That is, while it is always a perturbative correction [$\text{var}(\tilde{\mathcal{E}})$ remains $\ll 1$ for a sufficiently large number of pixels], this contribution can make the variance suboptimal and scale as $1/N_{\text{pix}}^{2/3}$ if $N_{\text{pix}} \gtrsim 10^6$. To understand better this contribution and how it is canceled when using the optimal estimator, it is useful to expand the simplified estimator in terms of the realization of the initial conditions $\hat{\delta}$ from which the observed tracer field $\hat{\delta}_g$ is generated [93]:

$$\begin{aligned} \tilde{\mathcal{E}} \supset \tilde{\mathcal{E}}_{(1)} + \tilde{\mathcal{E}}_{(2)} &= \sigma_A^2 \int_{\mathbf{k}} \frac{1}{P(k)} \hat{\delta}(\mathbf{k}) \hat{\delta}_g^{(2)}(-\mathbf{k}) \\ &+ \sigma_A^2 \int_{\mathbf{k}} \frac{|\hat{\delta}_g^{(2)}(\mathbf{k})|^2}{P(k)}, \end{aligned} \quad (110)$$

where we have used the relation $Y_2 = -X_2$ and by the \supset symbol we mean that we work at leading and next-to-leading order in $\hat{\delta}$, but consider only terms that come from expanding $\hat{\delta}_g(\mathbf{k})$ in Eq. (87) at second order. The term that gives rise to $\sigma_{2,-}^2$ in the variance of the estimator is $\langle \tilde{\mathcal{E}}_{(2)}^2 \rangle$. More precisely, we have

$$\begin{aligned} \text{var} \left(\int_{\mathbf{k}} \frac{|\hat{\delta}_g^{(2)}(\mathbf{k})|^2}{P(k)} \right)_{\sigma_{2,-}^2} &= 48 \int_{\mathbf{k}, \mathbf{k}', \mathbf{p}} \frac{P^2(p) P(|\mathbf{k} + \mathbf{p}|) P(\mathbf{k}' - \mathbf{p})}{P(k) P(k')} \\ &\times X_2(\mathbf{k} + \mathbf{p}, \mathbf{p}) X_2(\mathbf{k}' - \mathbf{p}, \mathbf{p}) \\ &\times X_2(-\mathbf{k} - \mathbf{p}, \mathbf{p}) X_2(-\mathbf{k}' + \mathbf{p}, -\mathbf{p}) \\ &\sim N_{\text{pix}}^2 N_{\text{pix}}^{-\frac{2}{3}(3+n)}, \end{aligned} \quad (111)$$

where by the subscript we indicate that we are focusing only on the contribution that gives rise to $\sigma_{2,-}^2$. In the last line we have used the assumption of an exact power-law universe without the IR cutoff k_* to estimate the scaling of the variance with the number of pixels. On the other hand, the full field-level estimator of Eq. (48) is given by

$$\mathcal{E} = \frac{\tilde{\mathcal{E}}}{1 + \delta\mathcal{E}}, \quad (112)$$

where we have defined

$$\delta\mathcal{E} \equiv \frac{\text{quadr} - \langle \text{quadr} \rangle}{\langle \text{quadr} \rangle} \quad (113)$$

and used Eqs. (51) and (52). It is now straightforward to see that in the variance of the full estimator \mathcal{E} the contribution

from $\text{var}(\tilde{\mathcal{E}}_{(2)})$ is canceled; i.e. there is no appearance of $\sigma_{2,-}^2$. Indeed, $\delta\mathcal{E}$ has zero mean so its typical value is given by its variance. The contribution to its variance that contains the parameter $\sigma_{2,-}^2$ comes from the second line of Eq. (46), and is precisely given by

$$\text{var}(\delta\mathcal{E})_{\sigma_{2,-}^2} = (\sigma_A^2)^2 \text{var}\left(\int_k \frac{|\hat{\delta}_g^{(2)}(\mathbf{k})|^2}{P(k)}\right)_{\sigma_{2,-}^2} \sim N_{\text{pix}}^{-\frac{2}{3}(3+n)}. \quad (114)$$

Hence, we see that for large N_{pix} the typical value of $\delta\mathcal{E}$ is small, and we can approximate

$$\mathcal{E} \approx \frac{\tilde{\mathcal{E}}}{1 + \sqrt{\text{var}(\delta\mathcal{E})}} \approx \tilde{\mathcal{E}} - \tilde{\mathcal{E}} \sqrt{\text{var}(\delta\mathcal{E})}. \quad (115)$$

Let us stress again that the smallness of $\delta\mathcal{E}$ does not imply that the estimators $\tilde{\mathcal{E}}$ and \mathcal{E} are the same. As we saw above, the naive estimator can have the suboptimal variance, even though all contributions decay with the number of pixels N_{pix} . Computing the variance of \mathcal{E} in this approximation is easy. First, it is straightforward to see that at leading order

$$\text{var}(\tilde{\mathcal{E}}_{(1)}) = \sigma_A^2. \quad (116)$$

Then, using this leading order result $\text{var}(\tilde{\mathcal{E}}) = \text{var}(\tilde{\mathcal{E}}_{(1)})$, we see that the part of the variance of $\tilde{\mathcal{E}}_{(2)}$ that contains the large parameter $\sigma_{2,-}^2$ cancels with the contribution coming from the variance of $\delta\mathcal{E}$. We can treat the contribution from Eq. (90) in a similar way. This time it cancels with the part of “quadr” that contains the cubic interactions Y_3 . More precisely, the cancellation comes from the terms $Y_3 \sim X_2 X_2$. In conclusion, the approximation to the true estimator given by Eq. (115) is good enough to ensure that the estimator is nearly optimal. This is the analog of the estimator for local primordial non-Gaussianities derived in [93].

We can also construct a simple estimator for the case where A_s is measured from nonlinear modes where our formulas do not apply. This is particularly relevant for weak lensing surveys where most of the signal comes from $k > k_{\text{NL}}$. In order to do so, we can start from the observation that the large covariance matrix for the naive estimator in Eq. (73) was coming from two long modes modulating the short-scale power spectrum. Indeed, the leading contribution to the covariance matrix even in the nonlinear regime can be expressed through the so-called response of $P_g(k)$ to the two long modes [109].¹¹ This response is defined as

¹¹More generally, responses can be defined and measured for the nonlinear field rather than the power spectrum [113].

$$\mathcal{R}_2(\mathbf{k}, \mathbf{p}_1, \mathbf{p}_2) = \frac{1}{2} \frac{1}{P_g(k)} \frac{\partial^2 P_g(k)}{\partial \delta(\mathbf{p}_1) \partial \delta(\mathbf{p}_2)} \Big|_{\delta(\mathbf{p}_i)=0}, \quad (117)$$

where $\mathbf{p}_i \ll k$, and it can be in principle measured in simulations [109]. What we want to achieve is to “remove” this modulation of the power spectrum by the long modes. We can define the following modified estimator for the power spectrum:

$$\hat{P}_g^{\text{new}}(k) \equiv \hat{P}_g(k) - Y_1^2 \int_{\mathbf{p} \ll k} \mathcal{R}_2(\mathbf{k}, \mathbf{p}, -\mathbf{p}) \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(-\mathbf{p}) \hat{P}_g(k), \quad (118)$$

where we have used $\delta(\mathbf{p}) = Y_1 \delta_g(\mathbf{p})$ on very large scales. It is easy to show that large covariance of the standard power spectrum estimator $\hat{P}_g(k)$ is exactly canceled by the second term. For this reason we expect this new estimator to be nearly optimal. Some indirect evidence for this can be found for instance in Ref. [2] where it is shown that constraints on cosmological parameters for a joint power spectrum and bispectrum analysis improve once the cross covariance between the two is taken into account. The reason is that the bispectrum partially takes into account the effects of the long modes on the small-scale power spectrum that we discussed here. To include the whole information at leading order one would also have to include the trispectrum. It is also important to note that $\hat{P}_g^{\text{new}}(k)$ is not unbiased. However, the bias is proportional to the variance of the field on large scales, and therefore it can easily be computed in either perturbation theory or simulations. In its essence, the proposed new estimator is a version of reconstruction where the scatter in $\hat{P}_g(k)$ induced by two long modes is reduced by using the knowledge of realization of large-scale galaxy density field. This is in spirit very similar to the standard BAO reconstruction, but with the aim of undoing real gravitational nonlinearities rather than displacements. It would be interesting to explore this strategy in more detail and test it on simulations or real data. We leave this interesting investigation for future work.

Finally, let us briefly comment on the case of large noise from mildly nonlinear scales that we discussed in the previous section. There too we can define

$$\hat{P}_g^{\text{new}}(k) \equiv \hat{P}_g(k) - \frac{b_2 Y_1^2}{2} \int_{k < q < k_{\text{NL}}} \hat{\delta}_g(\mathbf{p}) \hat{\delta}_g(\mathbf{k} - \mathbf{p}), \quad (119)$$

and check again that the second term cancels the large contribution to the effective noise in $P_g(k)$. Even though we have argued that the noise contribution coming from a quadratic galaxy does not lead to parametrically large error bars, it would be interesting to explore this in more detail and check if even a moderate improvement of cosmological constraints is possible for very dense tracers such as neutral hydrogen. We leave this for future work.

In summary, we have shown in this section a few examples of new simple estimators for the power spectrum and bispectrum. Their form is inspired by computing the full perturbative posterior and identifying relevant long-short interactions that can make the standard analyses suboptimal. The variance of new estimators does not suffer from large parameters such as $\sigma_{2,-}^2$ or $\sigma_{2,+}^2(k)$. Therefore, these estimators are nearly optimal, but still much simpler to implement in practice than the full forward modeling. This is particularly true in regimes where the simple perturbative model does not hold, such as the nonlinear regime.

V. CONCLUSIONS

In this work we have studied how well perturbative forward modeling can constrain cosmological parameters compared to conventional analyses based on n -point functions. We have focused on the case where cosmic variance dominates the error budget. In this limit it is easy to derive the field-level posterior for cosmological parameters. We have shown that perturbative forward modeling is equivalent to conventional analyses: once the field is predicted at a given order in perturbation theory, an analysis with all the correlation functions one can correctly predict at that order will achieve the same errors on cosmological parameters.

As all theorems, our result relies on several assumptions, the main one being that the only relevant parameter for the nonlinear evolution is the nonlinear scale. While this is true in most situations of interest, even in Λ CDM-like cosmologies we know that there are other relevant parameters that change this simplified picture. In the conventional analyses these parameters can lead to either a depletion of the signal or an increase of errors, making them suboptimal. We discussed some examples, such as broadening of the BAO peak or large contributions to the covariance matrix due to the long modes. On the other hand, in all such cases the field-level analysis remains optimal. Given these counterexamples to our general claim, one may argue that the field-level inference is the only way to harvest all cosmological information. However, we have argued that in all relevant cases one can do a simple reconstruction and use simple modified estimators that are nearly optimal.

Do the results of this work mean that perturbative forward modeling should not be pursued further? We think it is still an interesting direction to follow. However, we argue that one should carry out forward modeling fully perturbatively; i.e. the marginalization over the initial conditions should also be carried out in perturbation theory. It is only at this level that: (1) the non-Gaussianity of the likelihood and the modulation of the noise by matter fluctuations can be consistently included; (2) the correct comparison to analyses based on correlation functions can be carried out; and (3) one can have a systematic understanding of how to implement the theoretical error [114,115] at the field level. In such a setup the usefulness of perturbative forward modeling over standard

analyses would not come from obtaining better constraints on cosmology, but from having an alternative way to include information from higher-order correlation functions. Importantly, we believe that if this direction is pursued further and perturbative forward modeling is to be successfully applied to data, a full understanding of renormalization at the field level and how it compares with renormalization of correlation functions must be achieved. While here we have not discussed this topic in detail, it being far from the scope of this work, an interesting application of the formalism we used in this work is presented in Appendix B, where we extend our formulas beyond the zero-noise, cosmic-variance-limited case discussed in the main text and apply them to show that the “effective noise” discussed in Sec. IV B carries information on the amplitude of the quadratic bias b_2 of the tracer under consideration, even if the amplitude of the shot noise P_ϵ is marginalized over. This goes against the usual lore, where it is assumed that the low- k limit of $\mathcal{I}_{\delta^2\delta^2}(k)$ is fully degenerate with P_ϵ , and hence, it is reabsorbed by it after renormalization. We leave a more detailed investigation to future work.

There are many other aspects of our work that require further investigation. For example, a key ingredient in building the posterior was the inverse model. It would be very important to test the inverse model in simulations and check its range of validity. Related to this, it would also be interesting to evaluate our perturbative posterior given some data $\hat{\delta}_g$ and compare it to the full forward modeling. As we have explained, constraints on the BAO scale are the only ones we expect to differ significantly. Therefore, it would also be interesting to apply our methods to the reconstructed galaxy field, where the large displacements are largely removed and the agreement with the optimal analysis is expected to be much better. Along the same lines, one could try to implement our new nearly optimal estimators in practice and check if they lead to tightening of the error bars, particularly in the case of weak lensing. All our results can be straightforwardly generalized to redshift space. Finally, it would be interesting to do it explicitly, test the inverse model in redshift space and develop a pipeline for a realistic spectroscopy survey.

We conclude with some words on primordial non-Gaussianity. Reference [116] discussed constraints on primordial non-Gaussianity of the equilateral type at the field level. The results of this work can be straightforwardly applied to the case of non-Gaussian initial conditions¹²: the only difference in the formulas of Sec. II will be that the prior is not anymore a Gaussian but depends on the primordial bispectrum. The posterior at the field level will still be a combination of suitably weighted correlation functions of the data: hence, we expect that, for non-Gaussianity of the equilateral and orthogonal type, working

¹²For recent constraints on primordial non-Gaussianity from BOSS data see Refs. [117–119].

at a given loop order at the field level will give the same constraints on f_{NL} as the correlation functions that we correctly capture at that order. It is interesting that local-type primordial non-Gaussianities are enhanced in the infrared: however, unlike the case of CMB anisotropies discussed in Ref. [93], the presence of the transfer function never allows one to achieve scale invariance, and consequently the effects discussed in Sec. IV A will never be dramatic.

ACKNOWLEDGMENTS

It is a pleasure to thank Kazuyuki Akitsu, Alex Barreira, Stephen Chen, Misha Ivanov, Andrija Kostić, Matt Lewandowski, Chirag Modi, Minh Nguyen, Andrej Obuljen, Oliver Philcox, Fabian Schmidt, Blake Sherwin, and Zvonimir Vlah for useful discussions. G. C. acknowledges support from the Institute for Advanced Study. Part of the work of G. C. was carried out at the Aspen Center for Physics, which is supported by National Science Foundation Grant No. PHY-1607611. M. Z. is supported by NSF 2209991 and NSF-BSF 2207583.

APPENDIX A: FROM LIKELIHOOD TO POSTERIOR-NONPERTURBATIVE INVERSION

In this appendix we show how we can extend the results of Sec. II to all orders in perturbation theory. The difficulty of the calculation lies in how to treat the Dirac delta functional to which the likelihood reduces in the limit of small noise. For this purpose we introduce the following compact notation:

- (i) We use Greek indices to denote functionals of momentum \mathbf{k} . The initial conditions $\delta(\mathbf{k})$ become the ‘‘coordinates’’ δ^μ . Similarly, the fiducial initial conditions $\hat{\delta}(\mathbf{k})$ become $\hat{\delta}^\mu$. We will need to take functional derivatives with respect to $\hat{\delta}(\mathbf{k})$, and not with respect to $\delta(\mathbf{k})$: since there is no possibility of confusion we denote these simply by ∂_μ , without any superscript.
- (ii) Latin indices are used for the remaining parameters, i.e. θ . We use ∂_i to denote derivatives with respect to these parameters. These derivatives are always evaluated at the fiducial values $\hat{\theta}$ of these parameters: hence, we do not need to use any further symbol.
- (iii) We define the following quantities (we use the letter ‘‘F’’ as in ‘‘forward model’’):

$$F^\mu = \delta_g[\delta, \theta](\mathbf{k}), \quad (\text{A1a})$$

$$\tilde{F}^\mu = F[\hat{\delta}, \theta](\mathbf{k}), \quad (\text{A1b})$$

$$\hat{F}^\mu = F[\hat{\delta}, \hat{\theta}](\mathbf{k}), \quad (\text{A1c})$$

$$X^\mu = \hat{F}^\mu - \tilde{F}^\mu, \quad (\text{A1d})$$

$$\tilde{\mathcal{D}}^\mu{}_\nu = \partial_\nu \tilde{F}^\mu, \quad (\text{A1e})$$

$$\tilde{\mathcal{D}}^\mu{}_{\nu_1 \nu_2 \dots \nu_n} = \partial_{\nu_1} \partial_{\nu_2} \dots \partial_{\nu_n} \tilde{F}^\mu. \quad (\text{A1f})$$

Then, we define $\tilde{\mathcal{I}}^\mu{}_\nu$ as the matrix inverse of Eq. (A1e). That is,

$$\tilde{\mathcal{I}}^\mu{}_\rho \tilde{\mathcal{D}}^\rho{}_\nu = \delta^\mu{}_\nu = \delta_\nu{}^\mu = \tilde{\mathcal{D}}^\rho{}_\nu \tilde{\mathcal{I}}^\mu{}_\rho. \quad (\text{A2})$$

- (iv) Similar definitions hold for the derivatives of \hat{F}^μ . That is, we have

$$\hat{\mathcal{D}}^\mu{}_\nu = \partial_\nu \hat{F}^\mu, \text{ etc.} \quad (\text{A3})$$

- (v) The remaining definitions we need are those for the prior. We have

$$\hat{\mathcal{P}} = \mathcal{P}[\hat{\delta}]. \quad (\text{A4})$$

Moreover, we also define the symmetric matrices

$$P(k)(2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') = P^{\mu\nu}, \quad (\text{A5a})$$

$$P^{-1}(k)(2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') = P_{\mu\nu}, \quad (\text{A5b})$$

so that

$$\partial_\mu \ln \hat{\mathcal{P}} = -P_{\mu\nu} \hat{\delta}^\nu \quad \text{and} \quad \partial_\mu \partial_\nu \ln \hat{\mathcal{P}} = -P_{\mu\nu}. \quad (\text{A6})$$

After cleaning up we will arrive at expressions involving only $\hat{\mathcal{P}}$ together with its derivatives $\partial_\mu \ln \hat{\mathcal{P}}$ and $\partial_\mu \partial_\nu \ln \hat{\mathcal{P}}$, the derivatives of \hat{F}^μ with respect to the parameters θ at their fiducial values $\hat{\theta}$, the matrix $\hat{\mathcal{I}}^\mu{}_\nu$, and the matrices $\hat{\mathcal{D}}^\mu{}_{\nu_1 \nu_2 \dots \nu_n}$ with their derivatives $\partial_i \hat{\mathcal{D}}^\mu{}_{\nu_1 \nu_2 \dots \nu_n}$. At the end of Sec. A 3 we will drop out all ‘‘hat’’ superscripts to make the notation as compact as possible. We will also need to take averages of various expressions over $\hat{\delta}$: we denote these simply by $\langle \dots \rangle$.

1. Marginalization over initial conditions

As discussed in Sec. II A, in the cosmic-variance-limited case the conditional likelihood becomes a Dirac delta functional. We recall Eq. (7), i.e.

$$\mathcal{P}[\hat{\delta}_g | \theta] = \int \mathcal{D}\delta \delta_D^{(\infty)}(\hat{F} - F) \mathcal{P}[\delta]. \quad (\text{A7})$$

In Sec. II we carried out this integral by solving perturbatively the delta functional for δ in terms of δ_g . Now we will see how to carry out this integral at all orders in δ . Before proceeding notice that we have only considered a

dependence of the forward model, and not of the prior, on θ . As discussed in Sec. II there is no loss of generality in doing this.

Back to Eq. (A7). We do the change of variables

$$\delta^\mu = \hat{\delta}^\mu + \Delta^\mu. \quad (\text{A8})$$

The advantage is that we expect that the integrand peaks at $\Delta^\mu = 0$, so it pays to write

$$F^\mu - \hat{F}^\mu = \underbrace{\tilde{\mathcal{D}}^\mu{}_\nu \Delta^\nu + \frac{1}{2!} \tilde{\mathcal{D}}^\mu{}_{\nu\rho} \Delta^\nu \Delta^\rho + \dots - X^\mu}_{\equiv G[\Delta]^\mu}, \quad (\text{A9})$$

and after changing variables from Δ^μ to $Y^\mu = G[\Delta]^\mu$ we get

$$-\ln \mathcal{P}[\hat{\delta}_g | \theta] = -\ln \mathcal{P}[\hat{\delta} + G^{-1}[X]] - \ln \left| \frac{\partial G^{-1}[X]}{\partial X} \right|. \quad (\text{A10})$$

It is important to emphasize that here we consider only the solution of $Y^\mu = G[\Delta]^\mu$ connected to linear theory, as we have done throughout the rest of the paper. If the forward model is built from a filtered field, and is itself cut at a finite momentum (or if equivalently we are working with a coarse enough lattice in real space), this is a good assumption.

What we need to do now is to compute derivatives $\partial_{i_1} \dots \partial_{i_n}$: the key point that comes to our help is that X^μ vanishes if $\theta = \hat{\theta}$. Let us first check that the average of the first derivative vanishes once we average over the fiducial initial conditions (Sec. A 2), and then obtain expressions for the second derivatives (Sec. A 3).

2. Unbiasedness

Showing unbiasedness is now straightforward irrespective of what kind of parameter we are looking at. First, we need an expression for $G^{-1}[X]^\mu$. Luckily we only need this as a power series in X^μ . Given the definition of Eq. (A9), it is easy to see that

$$G^{-1}[X]^\mu = \tilde{\mathcal{I}}^\mu{}_\nu X^\nu - \frac{1}{2} \tilde{\mathcal{I}}^\nu{}_\beta \tilde{\mathcal{I}}^\mu{}_\rho \tilde{\mathcal{I}}^\sigma{}_\alpha \tilde{\mathcal{D}}^\rho{}_{\sigma\nu} X^\alpha X^\beta + \dots \quad (\text{A11})$$

To obtain this equation we have used the relation

$$\partial_\nu \tilde{\mathcal{I}}^\mu{}_\alpha = -\tilde{\mathcal{I}}^\mu{}_\rho \tilde{\mathcal{D}}^\rho{}_{\sigma\nu} \tilde{\mathcal{I}}^\sigma{}_\alpha, \quad (\text{A12})$$

and the tensor multiplying $X^\alpha X^\beta$ is symmetric in $\alpha \leftrightarrow \beta$ because $\tilde{\mathcal{D}}^\rho{}_{\sigma\nu} = \tilde{\mathcal{D}}^\rho{}_{\nu\sigma}$. Notice that we need the expansion up to second order because of the Jacobian in Eq. (A10). Being careful about the sign in the definition of X^μ in Eq. (A1d), we obtain

$$-\partial_i \ln \mathcal{P}[\hat{\delta}_g | \hat{\theta}] = (\partial_\nu \ln \hat{\mathcal{P}}) \hat{\mathcal{I}}^\nu{}_\mu \partial_i \hat{F}^\mu + \partial_\nu \{ \hat{\mathcal{I}}^\nu{}_\mu \partial_i \hat{F}^\mu \}. \quad (\text{A13})$$

It is easy to see that the average of Eq. (A13) over the fiducial initial conditions vanishes. We can—somewhat suggestively—rewrite it as

$$-\partial_i \ln \mathcal{P}[\hat{\delta}_g | \hat{\theta}] = \underbrace{(\partial_\nu \ln \hat{\mathcal{P}}) \hat{\mathcal{I}}^\nu{}_\mu}_{=-\Gamma^\rho{}_{\rho\nu}} \partial_i \hat{F}^\mu + \partial_\nu \underbrace{\{ \hat{\mathcal{I}}^\nu{}_\mu \partial_i \hat{F}^\mu \}}_{=V_i^\nu} = \nabla_\mu V_i^\mu, \quad (\text{A14})$$

where averaging over $\hat{\delta}$ is equivalent to integrating over $\hat{\delta}$ with a measure given by a diagonal metric with determinant equal to $\hat{\mathcal{P}}$. Using Stokes' theorem, and the fact that the measure vanishes exponentially fast on the “boundary,” we get

$$\langle -\partial_i \ln \mathcal{P}[\hat{\delta}_g | \hat{\theta}] \rangle = 0. \quad (\text{A15})$$

3. Second derivatives of log-posterior

When looking at second derivatives the algebra becomes more complicated. It is helpful to define the matrix $Z^\mu{}_\nu$ as

$$Z^\mu{}_\nu = \frac{\partial G^{-1}[X]^\mu}{\partial X^\nu}, \quad (\text{A16})$$

with $(Z^{-1})^\mu{}_\nu$ as its matrix inverse. Then, we have that $-\partial_i \partial_j \ln \mathcal{P}[\hat{\delta}_g | \hat{\theta}]$ contains four terms: two come from the prior and two come from the Jacobian. More precisely, we have

$$\begin{aligned} & -\partial_i \partial_j \ln \mathcal{P}[\hat{\delta}_g | \hat{\theta}] \\ &= -(\partial_\mu \partial_\nu \ln \hat{\mathcal{P}}) \partial_i G^{-1}[X]^\mu \partial_j G^{-1}[X]^\nu - (\partial_\mu \ln \hat{\mathcal{P}}) \partial_i \partial_j G^{-1}[X]^\mu \\ & \quad + \underbrace{(Z^{-1})^\nu{}_\rho (Z^{-1})^\sigma{}_\mu \partial_i Z^\mu{}_\nu \partial_j Z^\rho{}_\sigma - (Z^{-1})^\nu{}_\mu \partial_i \partial_j Z^\mu{}_\nu}_{\equiv J_{ij}^{11} + J_{ij}^{02}}, \end{aligned} \quad (\text{A17})$$

where we have used the relation

$$\partial_i (Z^{-1})^\mu{}_\alpha = -(Z^{-1})^\mu{}_\rho (\partial_i Z^\rho{}_\sigma) (Z^{-1})^\sigma{}_\alpha. \quad (\text{A18})$$

In order to compute the last term in Eq. (A17) we need to extend Eq. (A11) to third order. We get

$$\begin{aligned} G^{-1}[X]^\mu &= \tilde{\mathcal{I}}^\mu{}_\nu X^\nu - \frac{1}{2} \tilde{\mathcal{I}}^\nu{}_\beta \tilde{\mathcal{I}}^\mu{}_\rho \tilde{\mathcal{I}}^\sigma{}_\alpha \tilde{\mathcal{D}}^\rho{}_{\sigma\nu} X^\alpha X^\beta \\ & \quad + \tilde{\mathcal{I}}^\mu{}_\nu M^\nu{}_{\alpha\beta\gamma} X^\alpha X^\beta X^\gamma + \dots, \end{aligned} \quad (\text{A19})$$

where

$$M^\nu{}_{\alpha\beta\gamma} = -\frac{1}{6} \tilde{\mathcal{D}}^\nu{}_{\rho\sigma\lambda} \tilde{\mathcal{I}}^\rho{}_\alpha \tilde{\mathcal{I}}^\sigma{}_\beta \tilde{\mathcal{I}}^\lambda{}_\gamma + \frac{1}{2} \tilde{\mathcal{D}}^\nu{}_{\rho\sigma} \tilde{\mathcal{I}}^\rho{}_\alpha \tilde{\mathcal{I}}^\sigma{}_\lambda \tilde{\mathcal{D}}^\lambda{}_{\kappa\eta} \tilde{\mathcal{I}}^\kappa{}_\beta \tilde{\mathcal{I}}^\eta{}_\gamma. \quad (\text{A20})$$

We then have all the ingredients to compute all four terms of Eq. (A17). From now on we drop all “hat” superscripts: all quantities are intended as evaluated at $\hat{\theta}$.

- (i) The first prior term, involving the first derivative squared of the forward model, is given by

$$\partial_i G^{-1}[X]^\mu \partial_j G^{-1}[X]^\nu = \mathcal{I}^\mu_\rho \mathcal{I}^\nu_\sigma \partial_i F^\rho \partial_j F^\sigma. \quad (\text{A21})$$

- (ii) The second prior term, which instead involves the second derivative of the forward model, is given by

$$\begin{aligned} \partial_i \partial_j G^{-1}[X]^\mu = & -\mathcal{I}^\mu_\nu \partial_i \partial_j F^\nu + 2\mathcal{I}^\mu_\alpha \mathcal{I}^\beta_\nu \partial_{(i} F^\nu \partial_{j)} \mathcal{D}^\alpha_\beta \\ & - \mathcal{I}^\nu_\beta \mathcal{I}^\mu_\rho \mathcal{I}^\sigma_\alpha \mathcal{D}^\rho_{\sigma\nu} \partial_{(i} F^\alpha \partial_{j)} F^\beta. \end{aligned} \quad (\text{A22})$$

- (iii) The first of the two Jacobian terms is

$$\begin{aligned} J_{ij}^{11} = & \mathcal{I}^\nu_\rho \mathcal{I}^\sigma_\mu \partial_i \mathcal{D}^\rho_\sigma \partial_j \mathcal{D}^\mu_\nu \\ & - 2\mathcal{I}^\mu_\nu \mathcal{D}^\rho_{\sigma\mu} \mathcal{I}^\sigma_\lambda \mathcal{I}^\gamma_\rho \partial_{(i} F^\nu \partial_{j)} \mathcal{D}^\lambda_\gamma \\ & + \mathcal{I}^\gamma_\lambda \mathcal{D}^\sigma_{\rho\gamma} \mathcal{D}^\mu_{\nu\alpha} \mathcal{I}^\rho_\mu \mathcal{I}^\alpha_\beta \mathcal{I}^\nu_\sigma \partial_i F^\lambda \partial_j F^\beta. \end{aligned} \quad (\text{A23})$$

- (iv) Unsurprisingly, $-(Z^{-1})^\nu_\mu \partial_i \partial_j Z^\mu_\nu$ is by far the most complicated term. We have

$$\begin{aligned} J_{ij}^{02} = & \mathcal{I}^\beta_\alpha \partial_i \partial_j \mathcal{D}^\alpha_\beta - 2\mathcal{I}^\sigma_\alpha \mathcal{I}^\beta_\rho \partial_i \mathcal{D}^\rho_\sigma \partial_j \mathcal{D}^\alpha_\beta + 4\mathcal{I}^\rho_\sigma \mathcal{I}^\gamma_\alpha \mathcal{D}^\alpha_{\beta\rho} \mathcal{I}^\beta_\lambda \partial_{(i} F^\sigma \partial_{j)} \mathcal{D}^\lambda_\gamma \\ & + \mathcal{I}^\rho_\lambda \mathcal{I}^\gamma_\sigma \mathcal{D}^\alpha_{\beta\rho} \mathcal{I}^\beta_\alpha \partial_{(i} F^\sigma \partial_{j)} \mathcal{D}^\lambda_\gamma + \mathcal{I}^\rho_\alpha \mathcal{D}^\alpha_{\beta\rho} \mathcal{I}^\beta_\lambda \mathcal{I}^\gamma_\sigma \partial_{(i} F^\sigma \partial_{j)} \mathcal{D}^\lambda_\gamma \\ & - 2\mathcal{I}^\rho_\alpha \mathcal{I}^\beta_\rho \partial_{(i} F^\sigma \partial_{j)} \mathcal{D}^\alpha_{\beta\rho} - \mathcal{I}^\rho_\sigma \mathcal{I}^\beta_\alpha \mathcal{D}^\alpha_{\beta\rho} \partial_i \partial_j F^\sigma + \mathcal{D}^\mu_{\nu\rho\sigma} \mathcal{I}^\nu_\mu \mathcal{I}^\rho_\alpha \mathcal{I}^\sigma_\beta \partial_i F^\beta \partial_j F^\alpha \\ & - \mathcal{D}^\rho_{\gamma\lambda} \mathcal{D}^\mu_{\eta\sigma} \mathcal{I}^\lambda_\mu \mathcal{I}^\gamma_\rho \mathcal{I}^\eta_\alpha \mathcal{I}^\sigma_\beta \partial_i F^\beta \partial_j F^\alpha - 2\mathcal{D}^\rho_{\gamma\lambda} \mathcal{D}^\mu_{\eta\sigma} \mathcal{I}^\lambda_\mu \mathcal{I}^\gamma_\alpha \mathcal{I}^\eta_\beta \mathcal{I}^\sigma_\rho \partial_i F^\beta \partial_j F^\alpha. \end{aligned} \quad (\text{A24})$$

In the next sections we use these expressions to show that, if one knows the full matrix \mathcal{I}^μ_ν , the field level reproduces the linear-theory errors for parameters that appear only in the linear power spectrum.

Before proceeding, we emphasize that there is actually a great simplification in Eq. (A17). By direct calculation one can show that $-(\partial_\mu \ln \hat{\mathcal{P}}) \partial_i \partial_j G^{-1}[X]^\mu$ and $-(Z^{-1})^\nu_\mu \partial_i \partial_j Z^\mu_\nu$ combine to give

$$\begin{aligned} & -(\partial_\mu \ln \hat{\mathcal{P}}) \partial_i \partial_j G^{-1}[X]^\mu - (Z^{-1})^\nu_\mu \partial_i \partial_j Z^\mu_\nu \\ & = \nabla_\mu \{ \partial_i \partial_j G^{-1}[X]^\mu \}, \end{aligned} \quad (\text{A25})$$

where the ‘‘covariant derivative’’ ∇_μ is defined in the same way as in Eq. (A14).

Also, notice that a relation of the sort of Eq. (A25) should have been expected. It essentially shows that, on average, the error is controlled by ‘‘first derivatives squared’’ of the forward model, and not second derivatives. This is in line with the ‘‘standard’’ Fisher matrix expressions for marginalization over parameters; see e.g. Eq. (62) of Ref. [120] for a review.

4. Reproducing linear theory

Let us now show that, for parameters appearing in the linear matter power spectrum, the forward model reproduces linear theory if the inverse matrix \mathcal{I}^μ_ν is known exactly. Recall the discussion in Sec. II A: whatever the parameter we are looking at, we can make the change of variables

$$\delta(\mathbf{k}) \rightarrow \tau(k, \boldsymbol{\theta}) \delta(\mathbf{k}), \quad \text{with} \quad \tau(k, \boldsymbol{\theta}) = \frac{\mathcal{M}(k, \boldsymbol{\theta})}{\mathcal{M}(k, \hat{\boldsymbol{\theta}})}. \quad (\text{A26})$$

The prior for the new δ is a Gaussian with a power spectrum equal to the fiducial linear power spectrum at that redshift, and all the dependence on the parameters $\boldsymbol{\theta}$ is now in the forward model. We can then use the results of Sec. A 3.

For simplicity of notation we denote derivatives with respect to the single cosmological parameter we are focusing on (evaluated at their fiducial) via $'$ and ∂_{i_0} . It then proves useful to define the field $\gamma(\mathbf{k})$ and the matrix $\Gamma(\mathbf{k}, \mathbf{k}')$ as

$$\begin{aligned} \gamma(\mathbf{k}) &= \frac{\tau'(k, \hat{\boldsymbol{\theta}})}{\tau(k, \hat{\boldsymbol{\theta}})} \delta(\mathbf{k}), \\ \Gamma(\mathbf{k}, \mathbf{k}') &= \frac{\partial \gamma(\mathbf{k})}{\partial \delta(\mathbf{k}')} = \frac{\tau'(k, \hat{\boldsymbol{\theta}})}{\tau(k, \hat{\boldsymbol{\theta}})} (2\pi)^3 \delta_D^{(3)}(\mathbf{k} - \mathbf{k}'). \end{aligned} \quad (\text{A27})$$

Then, the key result is that derivatives with respect to any parameter can be rewritten in terms of derivatives with respect to the initial conditions. For example, it is easy to see that

$$\partial_{i_0} F^\mu = \gamma^\nu \mathcal{D}^\mu_\nu, \quad \text{i.e.} \quad F'(\mathbf{k}, \hat{\boldsymbol{\theta}}) = \int_{\mathbf{k}'} \frac{\tau'(k', \hat{\boldsymbol{\theta}})}{\tau(k', \hat{\boldsymbol{\theta}})} \delta(\mathbf{k}') \frac{\partial F(\mathbf{k}, \hat{\boldsymbol{\theta}})}{\partial \delta(\mathbf{k}')}. \quad (\text{A28})$$

One can see why this formula works by expanding the forward model in a power series in $\tau(k, \boldsymbol{\theta}) \delta(\mathbf{k})$: Eq. (A28) is a consequence of the fact that $\tau(k, \boldsymbol{\theta})$ and $\delta(\mathbf{k})$ enter always in this particular combination in this power series.

With the chain rule, and the fact that $\partial_{i_0} \partial_\mu = \partial_\mu \partial_{i_0}$ we can obtain similar formulas for other derivatives appearing in Eq. (A23). More precisely, we have

$$\partial_{i_0} \mathcal{D}^\mu{}_\nu = \partial_\nu \partial_{i_0} F^\mu = \partial_\nu (\gamma^\rho \mathcal{D}^\mu{}_\rho) = \Gamma^\rho{}_\nu \mathcal{D}^\mu{}_\rho + \gamma^\rho \mathcal{D}^\mu{}_{\rho\nu}. \quad (\text{A29})$$

We can then use Eqs. (A28), (A29) in Eq. (A17). Recalling Eq. (A6), all terms combine to give

$$-\partial_{i_0}^2 \ln \mathcal{P}[\hat{\delta}_g | \hat{\theta}] = P_{\mu\nu} \gamma^\mu \gamma^\nu + \Gamma^\alpha{}_\beta \Gamma^\beta{}_\alpha + \text{total divergence}. \quad (\text{A30})$$

Using the relations

$$\gamma^\mu = \Gamma^\mu{}_\nu \delta^\nu \quad \text{and} \quad \langle \delta^\mu \delta^\nu \rangle = P^{\mu\nu}, \quad (\text{A31})$$

we see that the average error per $d^3 k / (2\pi)^3$ and per unit volume is

$$2 \left(\frac{\tau'(k, \hat{\theta})}{\tau(k, \hat{\theta})} \right)^2, \quad (\text{A32})$$

i.e. the same as in linear theory.

We also emphasize that, while to arrive at Eq. (A30) we have dropped the total divergence, we have explicitly checked that if we consider also all the terms coming from Eqs. (A22), (A24) we obtain¹³

$$-\partial_{i_0}^2 \ln \mathcal{P}[\hat{\delta}_g | \hat{\theta}] = 3P_{\mu\nu} \gamma^\mu \gamma^\nu - \Gamma^\alpha{}_\beta \Gamma^\beta{}_\alpha, \quad (\text{A33})$$

whose average is clearly the same as Eq. (A30).

APPENDIX B: INCLUDING FINITE SHOT NOISE (AND SHOT NOISE RENORMALIZATION)

In this appendix we want to address the following question. In Sec. IV B we have seen how it is the ‘‘effective noise’’

$$P_\epsilon + \frac{b_2^2}{4} \lim_{k \rightarrow 0} \mathcal{I}_{\delta^2 \delta^2}(k) \quad (\text{B1})$$

and not only the shot noise P_ϵ that controls the variance of the naive estimator for the amplitude of the linear power spectrum. There we were focusing on the scenario where $P_\epsilon \ll P(k)$, the quadratic bias b_2 is fixed, and we want to measure A_s . But what if we instead want to measure b_2 ? And what if we do not know what P_ϵ is, and want to marginalize over it? Is there some information on b_2 contained in the effective noise?

To answer this question we need to include a finite P_ϵ in our calculation of the posterior. It is sufficient, however, to assume the following: (1) the limit $P_\epsilon \ll P(k)$; (2) that the

¹³Notice that, in order to derive this result, one needs also quantities such as $\partial_{i_0}^2 F^\mu$ and $\partial_{i_0} \mathcal{D}^\mu{}_{\nu\rho}$. These are straightforwardly obtained from Eqs. (A28), (A29).

fiducial noise is zero. In this case, we can derive a perturbative expansion of the posterior via¹⁴

$$\begin{aligned} \mathcal{L}[\hat{\delta}_g | \delta, \theta] &= \left(1 + \frac{P_\epsilon}{2} \int_{\mathbf{k}} \frac{\partial^2}{\partial \hat{\delta}_g(\mathbf{k}) \partial \hat{\delta}_g(-\mathbf{k})} \right. \\ &\quad \left. + \frac{P_\epsilon^2}{8} \int_{\mathbf{k}, \mathbf{p}} \frac{\partial^4}{\partial \hat{\delta}_g(\mathbf{k}) \partial \hat{\delta}_g(-\mathbf{k}) \partial \hat{\delta}_g(\mathbf{p}) \partial \hat{\delta}_g(-\mathbf{p})} + \dots \right) \\ &\quad \times \delta_{\text{D}}^{(\infty)}(\hat{\delta}_g - \delta_g[\delta, \theta]). \end{aligned} \quad (\text{B3})$$

At this point it is straightforward to compute the posterior at each order in $P_\epsilon/P(k)$ by bringing the series of derivatives out of the functional integral of Eq. (5). We obtain

$$\begin{aligned} -\ln \mathcal{P}[\theta | \hat{\delta}_g] &= \underbrace{\frac{1}{2} \chi_{\text{prior}}^2[\hat{\delta}_g, \theta] - \text{Tr} \ln J[\hat{\delta}_g, \theta]}_{\equiv -\ln \mathcal{P}[\theta | \hat{\delta}_g]_0} \\ &\quad + \frac{P_\epsilon}{2} \int_{\mathbf{k}} \left(\frac{\partial^2 \langle -\ln \mathcal{P}[\theta | \hat{\delta}_g]_0}{\partial \hat{\delta}_g(\mathbf{k}) \partial \hat{\delta}_g(-\mathbf{k})} \right. \\ &\quad \left. + \frac{\partial \langle -\ln \mathcal{P}[\theta | \hat{\delta}_g]_0}{\partial \hat{\delta}_g(\mathbf{k})} \frac{\partial \langle -\ln \mathcal{P}[\theta | \hat{\delta}_g]_0}{\partial \hat{\delta}_g(-\mathbf{k})} \right) \\ &\quad + \dots, \end{aligned} \quad (\text{B4})$$

where the higher orders in P_ϵ are straightforwardly obtained from Eq. (B3) using

$$\frac{\partial \mathcal{P}[\theta | \hat{\delta}_g]_0}{\partial \hat{\delta}_g(\mathbf{k})} = -\mathcal{P}[\theta | \hat{\delta}_g]_0 \frac{\partial (-\ln \mathcal{P}[\theta | \hat{\delta}_g]_0)}{\partial \hat{\delta}_g(\mathbf{k})}. \quad (\text{B5})$$

After averaging over the fiducial initial conditions, we see that the leading order in this expansion is still given by the formulas of Sec. II. More precisely, we find

¹⁴There are many ways to derive this expansion, e.g. via the Fourier transform. For one-dimensional integrals in dx one can write

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[d-f(x)]^2}{2\sigma^2}} &= \int_{-\infty}^{+\infty} dJ e^{-\frac{\sigma^2 J^2}{2}} e^{iJ[d-f(x)]} \\ &= \int_{-\infty}^{+\infty} dJ \left(\sum_{n=0}^{+\infty} \frac{(-1)^n \sigma^{2n} J^{2n}}{2^n n!} \right) e^{iJ[d-f(x)]} \\ &= \int_{-\infty}^{+\infty} dJ \left(\sum_{n=0}^{+\infty} \frac{\sigma^{2n}}{2^n n!} \right) \frac{\partial^{2n}}{\partial d^{2n}} e^{iJ[d-f(x)]} \\ &= \sum_{n=0}^{+\infty} \frac{\sigma^{2n}}{2^n n!} \frac{\partial^{2n}}{\partial d^{2n}} \delta_{\text{D}}^{(1)}(d-f(x)), \end{aligned} \quad (\text{B2})$$

where the Dirac delta is normalized such that $\int_{-\infty}^{+\infty} dx \delta_{\text{D}}^{(1)}(x) = 1$.

$$\langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g]_0 \rangle = \frac{\delta b_2^2}{4} V \int_{k,p} \frac{P(p)P(|\mathbf{k}-\mathbf{p}|)}{P(k)}, \quad (\text{B6})$$

which is a one-loop term. Notice that here we have already expanded up to second order in $\delta\boldsymbol{\theta} = \delta b_2$. The first order in δb_2 vanishes, consistently with unbiasedness. For the same reason, also the linear order in P_e vanishes (one can use the formalism of Appendix A to provide a nonperturbative proof of this). One can then derive, in a straightforward albeit tedious way, the orders $\delta b_2 \times P_e$ and P_e^2 , both at tree level and at one-loop order. Let us, however, discuss the expression for the error on b_2 after marginalization over P_e . By inverting the matrix of second derivatives of minus the log-posterior with respect to b_2 and P_e , we find

$$\frac{1}{\sigma_{b_2}^2} = \frac{\partial^2 \langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \rangle}{\partial b_2^2} - \left(\frac{\partial^2 \langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \rangle}{\partial b_2 \partial P_e} \right)^2 \times \left(\frac{\partial^2 \langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \rangle}{\partial P_e^2} \right)^{-1}. \quad (\text{B7})$$

This error can be expanded in loops, i.e. in powers of $\Delta^2(k)$. The leading-order (LO) contribution will be a one-loop term, the next-to-leading-order contribution (NLO) a two-loop term, and so on. From this expression we can now see that the explicit form of minus the log-posterior at orders $\delta b_2 \times P_e$ and P_e^2 is not important. What is crucial is

that at tree level only the order P_e^2 is not vanishing, being equal to

$$\frac{\partial^2 \langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \rangle}{\partial P_e^2} = \frac{V}{2} \int_k \frac{1}{P^2(k)} \text{ at tree level.} \quad (\text{B8})$$

This leads to

$$-\left(\frac{\partial^2 \langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \rangle}{\partial b_2 \partial P_e} \right)^2 \left(\frac{\partial^2 \langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \rangle}{\partial P_e^2} \right)^{-1} = \text{NLO}, \quad (\text{B9})$$

and consequently

$$\frac{1}{\sigma_{b_2}^2} \Big|_{\text{LO}} = \frac{\partial^2 \langle -\ln \mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g]_0 \rangle}{\partial b_2^2} = \frac{V}{2} \int_{k,p} \frac{P(p)P(|\mathbf{k}-\mathbf{p}|)}{P(k)}. \quad (\text{B10})$$

This tells us that at the field level not only can we still constrain b_2 after marginalizing over P_e and that there is information on b_2 in the amplitude of the effective noise, but also that the error on b_2 becomes smaller the larger the low- k limit of $\mathcal{I}_{\delta^2\delta^2}(k)$ is. This suggests that one should not let P_e absorb after renormalization the full $\lim_{k \rightarrow 0} \mathcal{I}_{\delta^2\delta^2}(k)$, but only the part of the loop integral in $\lim_{k \rightarrow 0} \mathcal{I}_{\delta^2\delta^2}(k)$ from $\sim k_{\text{NL}}$ to very UV modes. This would be similar to what happens with the speed of sound and the low- k limit of P_{13} for the nonlinear dark matter. A more detailed investigation of this is left to future work.

-
- [1] R. Scoccimarro, The bispectrum: From theory to observations, *Astrophys. J.* **544**, 597 (2000).
 - [2] E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro, Cosmology and the bispectrum, *Phys. Rev. D* **74**, 023522 (2006).
 - [3] T. Baldauf, L. Mercolli, M. Mirbabayi, and E. Pajer, The bispectrum in the effective field theory of large scale structure, *J. Cosmol. Astropart. Phys.* **05** (2015) 007.
 - [4] R. E. Angulo, S. Foreman, M. Schmittfull, and L. Senatore, The one-loop matter bispectrum in the effective field theory of large scale structures, *J. Cosmol. Astropart. Phys.* **10** (2015) 039.
 - [5] H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, The power spectrum and bispectrum of SDSS DR11 BOSS galaxies—I. Bias and gravity, *Mon. Not. R. Astron. Soc.* **451**, 539 (2015).
 - [6] Z. Slepian *et al.*, The large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies, *Mon. Not. R. Astron. Soc.* **468**, 1070 (2017).
 - [7] H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres, and M. D. Olmstead, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies, *Mon. Not. R. Astron. Soc.* **465**, 1757 (2017).
 - [8] Z. Slepian *et al.*, Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies, *Mon. Not. R. Astron. Soc.* **469**, 1738 (2017).
 - [9] D. W. Pearson and L. Samushia, A detection of the baryon acoustic oscillation features in the SDSS BOSS DR12 galaxy bispectrum, *Mon. Not. R. Astron. Soc.* **478**, 4500 (2018).
 - [10] A. Eggemeier, R. Scoccimarro, and R. E. Smith, Bias loop corrections to the galaxy bispectrum, *Phys. Rev. D* **99**, 123514 (2019).
 - [11] A. Oddo, E. Sefusatti, C. Porciani, P. Monaco, and A. G. Sánchez, Toward a robust inference method for the galaxy bispectrum: Likelihood function and model selection, *J. Cosmol. Astropart. Phys.* **03** (2020) 056.
 - [12] A. Moradinezhad Dizgah, M. Biagetti, E. Sefusatti, V. Desjacques, and J. Noreña, Primordial non-Gaussianity from biased tracers: Likelihood analysis of real-space power spectrum and bispectrum, *J. Cosmol. Astropart. Phys.* **05** (2021) 015.

- [13] A. Eggemeier, R. Scoccimarro, R. E. Smith, M. Crocce, A. Pezzotta, and A. G. Sánchez, Testing one-loop galaxy bias: Joint analysis of power spectrum and bispectrum, *Phys. Rev. D* **103**, 123550 (2021).
- [14] D. Alkhanishvili, C. Porciani, E. Sefusatti, M. Biagetti, A. Lazanu, A. Oddo, and V. Yankelevich, The reach of next-to-leading-order perturbation theory for the matter bispectrum, *Mon. Not. R. Astron. Soc.* **512**, 4961 (2022).
- [15] A. Oddo, F. Rizzo, E. Sefusatti, C. Porciani, and P. Monaco, Cosmological parameters from the likelihood analysis of the galaxy power spectrum and bispectrum in real space, *J. Cosmol. Astropart. Phys.* **11** (2021) 038.
- [16] T. Baldauf, M. Garny, P. Taule, and T. Steele, Two-loop bispectrum of large-scale structure, *Phys. Rev. D* **104**, 123551 (2021).
- [17] M. M. Ivanov, O. H. E. Philcox, T. Nishimichi, M. Simonović, M. Takada, and M. Zaldarriaga, Precision analysis of the redshift-space galaxy bispectrum, *Phys. Rev. D* **105**, 063512 (2022).
- [18] O. H. E. Philcox and M. M. Ivanov, BOSS DR12 full-shape cosmology: Λ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole, *Phys. Rev. D* **105**, 043517 (2022).
- [19] O. H. E. Philcox, M. M. Ivanov, M. Zaldarriaga, M. Simonovic, and M. Schmittfull, Fewer mocks and less noise: Reducing the dimensionality of cosmological observables with subspace projections, *Phys. Rev. D* **103**, 043508 (2021).
- [20] O. H. E. Philcox, Cosmology without window functions: Quadratic estimators for the galaxy power spectrum, *Phys. Rev. D* **103**, 103504 (2021).
- [21] O. H. E. Philcox, Cosmology without window functions. II. Cubic estimators for the galaxy bispectrum, *Phys. Rev. D* **104**, 123529 (2021).
- [22] K. Pardede, F. Rizzo, M. Biagetti, E. Castorina, E. Sefusatti, and P. Monaco, Bispectrum-window convolution via Hankel transform, *J. Cosmol. Astropart. Phys.* **10** (2022) 066.
- [23] F. Rizzo, C. Moretti, K. Pardede, A. Eggemeier, A. Oddo, E. Sefusatti, C. Porciani, and P. Monaco, The halo bispectrum multipoles in redshift space, *J. Cosmol. Astropart. Phys.* **01** (2023) 031.
- [24] M. M. Ivanov, O. H. E. Philcox, G. Cabass, T. Nishimichi, M. Simonović, and M. Zaldarriaga, Cosmology with the galaxy bispectrum multipoles: Optimal estimation and application to BOSS data, *Phys. Rev. D* **107**, 083515 (2023).
- [25] G. D'Amico, Y. Donath, M. Lewandowski, L. Senatore, and P. Zhang, The one-loop bispectrum of galaxies in redshift space from the effective field theory of large-scale structure, [arXiv:2211.17130](https://arxiv.org/abs/2211.17130).
- [26] G. D'Amico, Y. Donath, M. Lewandowski, L. Senatore, and P. Zhang, The BOSS bispectrum analysis at one loop from the effective field theory of large-scale structure, [arXiv:2206.08327](https://arxiv.org/abs/2206.08327).
- [27] K. Pardede, E. Di Dio, and E. Castorina, Wide-angle effects in the galaxy bispectrum, *J. Cosmol. Astropart. Phys.* **09** (2023) 030.
- [28] F. S. Kitaura and T. A. Ensslin, Bayesian reconstruction of the cosmological large-scale structure: Methodology, inverse algorithms and numerical optimization, *Mon. Not. R. Astron. Soc.* **389**, 497 (2008).
- [29] J. Jasche and B. D. Wandelt, Bayesian physical reconstruction of initial conditions from large scale structure surveys, *Mon. Not. R. Astron. Soc.* **432**, 894 (2013).
- [30] H. Wang, H. J. Mo, X. Yang, and F. C. v. d. Bosch, Reconstructing the initial density field of the local universe: Methods and tests with mock catalogs, *Astrophys. J.* **772**, 63 (2013).
- [31] M. Ata, F.-S. Kitaura, and V. Müller, Bayesian inference of cosmic density fields from non-linear, scale-dependent, and stochastic biased tracers, *Mon. Not. R. Astron. Soc.* **446**, 4250 (2015).
- [32] U. Seljak, G. Aslanyan, Y. Feng, and C. Modi, Towards optimal extraction of cosmological information from non-linear data, *J. Cosmol. Astropart. Phys.* **12** (2017) 009.
- [33] C. Modi, Y. Feng, and U. Seljak, Cosmological reconstruction from galaxy light: Neural network based light-matter connection, *J. Cosmol. Astropart. Phys.* **10** (2018) 028.
- [34] J. Alsing, B. Wandelt, and S. Feeney, Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology, *Mon. Not. R. Astron. Soc.* **477**, 2874 (2018).
- [35] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, Fast likelihood-free cosmology with neural density estimators and active learning, *Mon. Not. R. Astron. Soc.* **488**, 4440 (2019).
- [36] N. Jeffrey, J. Alsing, and F. Lanusse, Likelihood-free inference with neural compression of DES SV weak lensing map statistics, *Mon. Not. R. Astron. Soc.* **501**, 954 (2021).
- [37] J. Jasche and F. S. Kitaura, Fast Hamiltonian sampling for large scale structure inference, *Mon. Not. R. Astron. Soc.* **407**, 29 (2010).
- [38] J. Jasche and G. Lavaux, Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe, *Astron. Astrophys.* **625**, A64 (2019).
- [39] B. Dai and U. Seljak, Translation and rotation equivariant normalizing flow (TRENF) for optimal cosmological analysis, *Mon. Not. R. Astron. Soc.* **516**, 2363 (2022).
- [40] J. Robnik, G. B. De Luca, E. Silverstein, and U. Seljak, Microcanonical Hamiltonian Monte Carlo, [arXiv:2212.08549](https://arxiv.org/abs/2212.08549).
- [41] C. Modi, Y. Li, and D. Blei, Reconstructing the universe with variational self-boosted sampling, *J. Cosmol. Astropart. Phys.* **03** (2023) 059.
- [42] P. McDonald and A. Roy, Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS, *J. Cosmol. Astropart. Phys.* **08** (2009) 020.
- [43] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, Cosmological non-linearities as an effective fluid, *J. Cosmol. Astropart. Phys.* **07** (2012) 051.
- [44] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, The effective field theory of cosmological large scale structures, *J. High Energy Phys.* **09** (2012) 082.
- [45] R. A. Porto, L. Senatore, and M. Zaldarriaga, The Lagrangian-space effective field theory of large scale structures, *J. Cosmol. Astropart. Phys.* **05** (2014) 022.

- [46] L. Senatore and M. Zaldarriaga, Redshift space distortions in the effective field theory of large scale structures, [arXiv:1409.1225](https://arxiv.org/abs/1409.1225).
- [47] L. Senatore, Bias in the effective field theory of large scale structures, *J. Cosmol. Astropart. Phys.* **11** (2015) 007.
- [48] M. Lewandowski, A. Perko, and L. Senatore, Analytic prediction of baryonic effects from the EFT of large scale structures, *J. Cosmol. Astropart. Phys.* **05** (2015) 019.
- [49] M. Mirbabayi, F. Schmidt, and M. Zaldarriaga, Biased tracers and time evolution, *J. Cosmol. Astropart. Phys.* **07** (2015) 030.
- [50] R. Angulo, M. Fasiello, L. Senatore, and Z. Vlah, On the statistics of biased tracers in the effective field theory of large scale structures, *J. Cosmol. Astropart. Phys.* **09** (2015) 029.
- [51] M. Lewandowski, L. Senatore, F. Prada, C. Zhao, and C.-H. Chuang, EFT of large scale structures in redshift space, *Phys. Rev. D* **97**, 063526 (2018).
- [52] A. A. Abolhasani, M. Mirbabayi, and E. Pajer, Systematic renormalization of the effective theory of large scale structure, *J. Cosmol. Astropart. Phys.* **05** (2016) 063.
- [53] Z. Vlah, M. White, and A. Aviles, A Lagrangian effective field theory, *J. Cosmol. Astropart. Phys.* **09** (2015) 014.
- [54] D. Blas, M. Garny, M. M. Ivanov, and S. Sibiryakov, Time-sliced perturbation theory for large scale structure I: General formalism, *J. Cosmol. Astropart. Phys.* **07** (2016) 052.
- [55] Z. Vlah, E. Castorina, and M. White, The Gaussian streaming model and convolution Lagrangian effective field theory, *J. Cosmol. Astropart. Phys.* **12** (2016) 007.
- [56] V. Desjacques, D. Jeong, and F. Schmidt, Large-scale galaxy bias, *Phys. Rep.* **733**, 1 (2018).
- [57] A. Perko, L. Senatore, E. Jennings, and R. H. Wechsler, Biased tracers in redshift space in the EFT of large-scale structure, [arXiv:1610.09321](https://arxiv.org/abs/1610.09321).
- [58] S.-F. Chen, Z. Vlah, E. Castorina, and M. White, Redshift-space distortions in Lagrangian perturbation theory, *J. Cosmol. Astropart. Phys.* **03** (2021) 100.
- [59] G. Cabass, M. M. Ivanov, M. Lewandowski, M. Mirbabayi, and M. Simonović, Snowmass white paper: Effective field theories in cosmology, *Phys. Dark Universe* **40**, 101193 (2023).
- [60] T. Baldauf, E. Schaan, and M. Zaldarriaga, On the reach of perturbative descriptions for dark matter displacement fields, *J. Cosmol. Astropart. Phys.* **03** (2016) 017.
- [61] T. Baldauf, E. Schaan, and M. Zaldarriaga, On the reach of perturbative methods for dark matter density fields, *J. Cosmol. Astropart. Phys.* **03** (2016) 007.
- [62] M. Schmittfull, M. Simonović, V. Assassi, and M. Zaldarriaga, Modeling biased tracers at the field level, *Phys. Rev. D* **100**, 043514 (2019).
- [63] A. Taruya, T. Nishimichi, and D. Jeong, Grid-based calculation for perturbation theory of large-scale structure, *Phys. Rev. D* **98**, 103532 (2018).
- [64] M. McQuinn and A. D’Aloisio, The observable 21 cm signal from reionization may be perturbative, *J. Cosmol. Astropart. Phys.* **10** (2018) 016.
- [65] M. Schmittfull, M. Simonović, M. M. Ivanov, O. H. E. Philcox, and M. Zaldarriaga, Modeling galaxies in redshift space at the field level, *J. Cosmol. Astropart. Phys.* **05** (2021) 059.
- [66] C. Modi, M. White, A. Slosar, and E. Castorina, Reconstructing large-scale structure with neutral hydrogen surveys, *J. Cosmol. Astropart. Phys.* **11** (2019) 023.
- [67] F. Schmidt, An n -th order Lagrangian forward model for large-scale structure, *J. Cosmol. Astropart. Phys.* **04** (2021) 033.
- [68] N. Kokron, J. DeRose, S.-F. Chen, M. White, and R. H. Wechsler, The cosmology dependence of galaxy clustering and lensing from a hybrid N-body-perturbation theory model, *Mon. Not. R. Astron. Soc.* **505**, 1422 (2021).
- [69] A. Taruya, T. Nishimichi, and D. Jeong, Grid-based calculations of redshift-space matter fluctuations from perturbation theory: UV sensitivity and convergence at the field level, *Phys. Rev. D* **105**, 103507 (2022).
- [70] A. Obuljen, M. Simonović, A. Schneider, and R. Feldmann, Modeling HI at the field level, *Phys. Rev. D* **108**, 083528 (2023).
- [71] F. Schmidt, F. Elsner, J. Jasche, N. M. Nguyen, and G. Lavaux, A rigorous EFT-based forward model for large-scale structure, *J. Cosmol. Astropart. Phys.* **01** (2019) 042.
- [72] F. Elsner, F. Schmidt, J. Jasche, G. Lavaux, and N.-M. Nguyen, Cosmology inference from a biased density field using the EFT-based likelihood, *J. Cosmol. Astropart. Phys.* **01** (2020) 029.
- [73] G. Cabass and F. Schmidt, The EFT likelihood for large-scale structure, *J. Cosmol. Astropart. Phys.* **04** (2020) 042.
- [74] F. Schmidt, Sigma-eight at the percent level: The EFT likelihood in real space, *J. Cosmol. Astropart. Phys.* **04** (2021) 032.
- [75] G. Cabass and F. Schmidt, The likelihood for LSS: Stochasticity of bias coefficients at all orders, *J. Cosmol. Astropart. Phys.* **07** (2020) 051.
- [76] F. Schmidt, G. Cabass, J. Jasche, and G. Lavaux, Unbiased cosmology inference from biased tracers using the EFT likelihood, *J. Cosmol. Astropart. Phys.* **11** (2020) 008.
- [77] G. Cabass, The EFT likelihood for large-scale structure in redshift space, *J. Cosmol. Astropart. Phys.* **01** (2021) 067.
- [78] A. Barreira, T. Lazeyras, and F. Schmidt, Galaxy bias from forward models: Linear and second-order bias of IllustrisTNG galaxies, *J. Cosmol. Astropart. Phys.* **08** (2021) 029.
- [79] T. Lazeyras, A. Barreira, and F. Schmidt, Assembly bias in quadratic bias parameters of dark matter halos from forward modeling, *J. Cosmol. Astropart. Phys.* **10** (2021) 063.
- [80] I. Babić, F. Schmidt, and B. Tucci, BAO scale inference from biased tracers using the EFT likelihood, *J. Cosmol. Astropart. Phys.* **08** (2022) 007.
- [81] A. Andrews, J. Jasche, G. Lavaux, and F. Schmidt, Bayesian field-level inference of primordial non-Gaussianity using next-generation galaxy surveys, *Mon. Not. R. Astron. Soc.* **520**, 5746 (2023).
- [82] A. Kostić, N.-M. Nguyen, F. Schmidt, and M. Reinecke, Consistency tests of field level inference with the EFT likelihood, *J. Cosmol. Astropart. Phys.* **07** (2023) 063.
- [83] J. Stadler, F. Schmidt, and M. Reinecke, Cosmology inference at the field level from biased tracers in redshift-space, *J. Cosmol. Astropart. Phys.* **10** (2023) 060.

- [84] M. Crocce and R. Scoccimarro, Nonlinear evolution of baryon acoustic oscillations, *Phys. Rev. D* **77**, 023533 (2008).
- [85] N. Padmanabhan, M. White, and J. D. Cohn, Reconstructing baryon oscillations: A Lagrangian theory perspective, *Phys. Rev. D* **79**, 063523 (2009).
- [86] N. S. Sugiyama and D. N. Spergel, How does non-linear dynamics affect the baryon acoustic oscillation?, *J. Cosmol. Astropart. Phys.* **02** (2014) 042.
- [87] D. J. Eisenstein, H.-j. Seo, E. Sirko, and D. Spergel, Improving cosmological distance measurements by reconstruction of the baryon acoustic peak, *Astrophys. J.* **664**, 675 (2007).
- [88] M. Schmittfull, T. Baldauf, and M. Zaldarriaga, Iterative initial condition reconstruction, *Phys. Rev. D* **96**, 023505 (2017).
- [89] C. D. Rimes and A. J. S. Hamilton, Information content of the non-linear matter power spectrum, *Mon. Not. R. Astron. Soc.* **360**, L82 (2005).
- [90] A. J. S. Hamilton, C. D. Rimes, and R. Scoccimarro, On measuring the covariance matrix of the nonlinear power spectrum from simulations, *Mon. Not. R. Astron. Soc.* **371**, 1188 (2006).
- [91] I. Mohammed, U. Seljak, and Z. Vlah, Perturbative approach to covariance matrix of the matter power spectrum, *Mon. Not. R. Astron. Soc.* **466**, 780 (2017).
- [92] A. Barreira and F. Schmidt, Response approach to the matter power spectrum covariance, *J. Cosmol. Astropart. Phys.* **11** (2017) 051.
- [93] P. Creminelli, L. Senatore, and M. Zaldarriaga, Estimators for local non-Gaussianities, *J. Cosmol. Astropart. Phys.* **03** (2007) 019.
- [94] M. McQuinn, On the primordial information available to galaxy redshift surveys, *J. Cosmol. Astropart. Phys.* **06** (2021) 024.
- [95] Y. Feng, U. Seljak, and M. Zaldarriaga, Exploring the posterior surface of the large scale structure reconstruction, *J. Cosmol. Astropart. Phys.* **07** (2018) 043.
- [96] M. Peloso and M. Pietroni, Galilean invariance and the consistency relation for the nonlinear squeezed bispectrum of large scale structure, *J. Cosmol. Astropart. Phys.* **05** (2013) 031.
- [97] A. Kehagias and A. Riotto, Symmetries and consistency relations in the large scale structure of the universe, *Nucl. Phys.* **B873**, 514 (2013).
- [98] P. Creminelli, J. Noreña, M. Simonović, and F. Vernizzi, Single-field consistency relations of large scale structure, *J. Cosmol. Astropart. Phys.* **12** (2013) 025.
- [99] P. Creminelli, J. Gleyzes, M. Simonović, and F. Vernizzi, Single-field consistency relations of large scale structure. Part II: Resummation and redshift space, *J. Cosmol. Astropart. Phys.* **02** (2014) 051.
- [100] P. Creminelli, J. Gleyzes, L. Hui, M. Simonović, and F. Vernizzi, Single-field consistency relations of large scale structure. Part III: Test of the equivalence principle, *J. Cosmol. Astropart. Phys.* **06** (2014) 009.
- [101] M. Mirbabayi, M. Simonović, and M. Zaldarriaga, Baryon acoustic peak and the squeezed limit bispectrum, [arXiv:1412.3796](https://arxiv.org/abs/1412.3796).
- [102] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga, Equivalence principle and the baryon acoustic peak, *Phys. Rev. D* **92**, 043514 (2015).
- [103] D. Blas, M. Garny, M. M. Ivanov, and S. Sibiryakov, Time-sliced perturbation theory II: Baryon acoustic oscillations and infrared resummation, *J. Cosmol. Astropart. Phys.* **07** (2016) 028.
- [104] J. Carlson, B. Reid, and M. White, Convolution Lagrangian perturbation theory for biased tracers, *Mon. Not. R. Astron. Soc.* **429**, 1674 (2013).
- [105] L. Senatore and M. Zaldarriaga, The IR-resummed effective field theory of large scale structures, *J. Cosmol. Astropart. Phys.* **02** (2015) 013.
- [106] Z. Vlah, U. Seljak, M. Y. Chu, and Y. Feng, Perturbation theory, effective field theory, and oscillations in the power spectrum, *J. Cosmol. Astropart. Phys.* **03** (2016) 057.
- [107] L. Senatore and G. Trevisan, On the IR-resummation in the EFTofLSS, *J. Cosmol. Astropart. Phys.* **05** (2018) 019.
- [108] M. M. Ivanov and S. Sibiryakov, Infrared resummation for biased tracers in redshift space, *J. Cosmol. Astropart. Phys.* **07** (2018) 053.
- [109] A. Barreira and F. Schmidt, Responses in large-scale structure, *J. Cosmol. Astropart. Phys.* **06** (2017) 053.
- [110] D. Wadekar, M. M. Ivanov, and R. Scoccimarro, Cosmological constraints from BOSS with analytic covariance matrices, *Phys. Rev. D* **102**, 123521 (2020).
- [111] C. Modi, E. Castorina, and U. Seljak, Halo bias in Lagrangian space: Estimators and theoretical predictions, *Mon. Not. R. Astron. Soc.* **472**, 3959 (2017).
- [112] F. Villaescusa-Navarro *et al.*, Ingredients for 21 cm intensity mapping, *Astrophys. J.* **866**, 135 (2018).
- [113] A. Taruya and K. Akitsu, Lagrangian approach to super-sample effects on biased tracers at field level: Galaxy density fields and intrinsic alignments, *J. Cosmol. Astropart. Phys.* **11** (2021) 061.
- [114] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga, Large-scale structure constraints with controlled theoretical uncertainties, [arXiv:1602.00674](https://arxiv.org/abs/1602.00674).
- [115] A. Chudaykin, M. M. Ivanov, and M. Simonović, Optimizing large-scale structure data analysis with the theoretical error likelihood, *Phys. Rev. D* **103**, 043525 (2021).
- [116] D. Baumann and D. Green, The power of locality: Primordial non-Gaussianity at the map level, *J. Cosmol. Astropart. Phys.* **08** (2022) 061.
- [117] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonović, and M. Zaldarriaga, Constraints on single-field inflation from the BOSS galaxy survey, *Phys. Rev. Lett.* **129**, 021301 (2022).
- [118] G. D'Amico, M. Lewandowski, L. Senatore, and P. Zhang, Limits on primordial non-Gaussianities from BOSS galaxy-clustering data, [arXiv:2201.11518](https://arxiv.org/abs/2201.11518).
- [119] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonović, and M. Zaldarriaga, Constraints on multifield inflation from the BOSS galaxy survey, *Phys. Rev. D* **106**, 043506 (2022).
- [120] A. Heavens, Statistical techniques in cosmology, [arXiv:0906.0664](https://arxiv.org/abs/0906.0664).