

Non-Gaussian Galaxy Stochasticity and the Noise-Field Formulation

Henrique Rubira,^{a,b,c} Fabian Schmidt^d

^aUniversity Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 München, Germany

^bKavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, UK

^cCentre for Theoretical Cosmology, Department of Applied Mathematics and Theoretical Physics University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK

^dMax-Planck-Institut für Astrophysik,
Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

E-mail: fabians@mpa-garching.mpg.de, henrique.rubira@lmu.de

Abstract. We revisit the stochastic, or noise, contributions to the galaxy density field within the effective field theory (EFT) of large-scale structure. Starting from the general, all-order expression of the EFT partition function, we elucidate how the stochastic contributions can be described by local nonlinear couplings of a *single Gaussian noise field*. We introduce an alternative formulation of the partition function in terms of such a noise field, and derive the corresponding field-level likelihood for biased tracers. This noise-field formulation can capture the complete set of stochastic contributions to the galaxy density at the field level in a normalized, positive-definite probability density which is suitable for numerical sampling. We illustrate this by presenting the first results of EFT-based field-level inference with non-Gaussian and density-dependent stochasticity on dark matter halos using LEFTfield.

Keywords: Large-scale structure, galaxy clustering, field-level inference, bias, effective field theory

Contents

1	Introduction	1
1.1	Outline of the paper	3
1.2	Notation	3
2	From the partition function to the likelihood	6
2.1	The EFT partition function	6
2.2	EFT likelihood from the partition function	8
2.3	Limiting cases	9
2.3.1	Quadratic-in-current terms and all operators	9
2.3.2	All-orders-in-current terms but only unit operator $O = \mathbb{1}$	11
2.4	General expansion around the Gaussian likelihood	11
3	The noise-field formulation of stochasticity	13
3.1	Definition	14
3.2	Expansion around the Gaussian likelihood	15
3.3	Discussion	17
4	Numerical implementation and results	19
5	Conclusions	23
A	Comparison of standard (non-minimal) and minimal stochastic contributions for n-point functions	24
B	Issues with the real-space formulation of the likelihood	25
C	Derivation of coefficients $\mathcal{C}$ and $\tilde{\mathcal{C}}$	26
D	Expansion of likelihood in the noise-field formulation	29

1 Introduction

Surveys that probe tracers of matter are the backbone of modern cosmology. In this context, the general bias expansion (see [1] for a review) offers a symmetry-based framework to connect observed tracers to the underlying dark matter density field. While in general nonlocal in time, this bias expansion can be written as a local-in-time relation either in the initial conditions (Lagrangian approach) or the evolved density field (Eulerian approach), thanks to the factorization of time- and space-dependences in perturbation theory, provided linear growth is scale-independent.

Many investigations of the statistics of biased tracers (often simply referred to as “galaxies” in the following) in the EFT have focused on the deterministic part, i.e. the bias expansion,

$$\delta_{g,\text{det}}(\mathbf{x}, \tau) = \sum_O b_O(\tau) O(\mathbf{x}, \tau), \quad (1.1)$$

which describes the galaxy density in the mean-field sense. The operators $O(\mathbf{x}, \tau)$ encode the dependence of the galaxy density on large-scale perturbations via all local gravitational observables. The stochastic contributions, i.e. the scatter around Eq. (1.1), remain much less studied. In [1], motivated by expressions in [2], the stochastic part of the bias expansion was written as

$$\delta_g(\mathbf{x}, \tau) = \delta_{g,\text{det}} + \epsilon^{\text{n-min}}(\mathbf{x}, \tau) + \sum_O \epsilon_O^{\text{n-min}}(\mathbf{x}, \tau) O(\mathbf{x}, \tau). \quad (1.2)$$

with a set of noise fields $\{\epsilon^{\text{n-min}}, \epsilon_O^{\text{n-min}}\}$. Ref. [3] employs a similar expansion in terms of multiple stochastic fields. Each of those fields is completely characterized by their local cumulants $\langle (\epsilon_O^{\text{n-min}})^m(\mathbf{x}) \rangle$. The fields $\epsilon^{\text{n-min}}$ and $\epsilon_O^{\text{n-min}}$ are considered first-order in perturbations. We refer to this ansatz here as *non-minimal* noise theory.

Recently, Ref. [4] presented a general expression for galaxy statistics, at all orders in perturbation theory, at the partition function level. This partition function directly yields expressions for n -point correlation functions for any n and at any loop order, including all stochastic contributions generated in the EFT. However, the result is quite abstract, and its connection to field-level expressions such as Eq. (1.2) remained unclear. In this work, we show that the expression for the galaxy density *field* can in fact be reduced to

$$\delta_g(\mathbf{x}, \tau) = \sum_{m=0}^{\infty} \sum_{\mathbb{1}, O} b_O^{\{m\}}(\tau) [\epsilon_G(\mathbf{x})]^m O(\mathbf{x}, \tau), \quad (1.3)$$

with a *single unit Gaussian noise field* $\epsilon_G(\mathbf{x}) \sim \mathcal{N}(0, 1)$, and a set of free coefficients $\{b_{\mathbb{1}}^{\{m\}}, b_O^{\{m\}}\}$. Notice that the contributions with $m = 0$ correspond exactly to the deterministic part, $b_O^{\{0\}} = b_O$ in Eq. (1.1).¹ Conversely, the contributions with $m > 1$ and $O = \mathbb{1}$ generate the non-Gaussian stochasticity, i.e. the higher cumulants of the noise. The contributions with $m > 0$ and $O \neq \mathbb{1}$ generate a modulation of the noise by large-scale perturbations. Here, we consider a single tracer δ_g ; in case of multiple tracers, one would need to add an individual field ϵ_G for each tracer (see e.g. [5–7]). Further, we will drop the time argument from the fields and coefficients in most of the paper for clarity. Note that the field $\epsilon_G(\mathbf{x})$ is explicitly time-independent, with all time dependence of stochasticity absorbed in the coefficients $b_O^{\{m\}}$.

This result represents a substantial simplification: while Eq. (1.2) adds a *non-Gaussian stochastic field for every bias operator*, Eq. (1.3) states that the stochastic part of the galaxy density field can be written in terms of a *single Gaussian* stochastic field, and a set of generalized bias coefficients $b_O^{\{m\}}(\tau)$. The reason for this simplification is that statistics derived from Eq. (1.2) still contain a number of degenerate contributions. This is shown at the level of galaxy n -point functions in App. A, which also argues that Eq. (1.3) is sufficient to describe the non-degenerate stochastic contributions.

Our main focus in this paper however is the field-level likelihood for biased tracers. We show that Eq. (1.3) enables a practical implementation of a field-level likelihood that can be used to incorporate all stochastic contributions in the EFT for field-level inference applications [8–26]. We emphasize that our goal is to obtain a likelihood that can capture all contributions, order by order, in the EFT of LSS. Other non-Gaussian likelihoods, such as Poisson or log-normal that have been used in empirical models for field-level inference, are

¹The contribution $b_{\mathbb{1}}^{\{0\}}$ multiplying the constant operator $\mathbb{1}$ is usually dropped since it corresponds to a spatially constant contribution ($\mathbf{k} = 0$ mode).

not suited for this. This generalized likelihood in the “noise-field formulation” proceeds by jointly inferring the initial density field δ_{in} and the noise field ϵ_G , with independent Gaussian priors. Note however that the field ϵ_G is an *effective* (or “nuisance”) field, not a physical field; in other words, the specific realization of ϵ_G has no physical significance. This is in keeping with the fact that the initial conditions δ_{in} are characterized by a single adiabatic growing mode, which we constrain with a single tracer density field δ_g . Multiple tracers δ_g^a co-located in volume would require the introduction of multiple fields ϵ_G^a , as mentioned above.

We then present first results from an actual implementation of the noise-field formulation in the `LEFTfield` code, in the form of field-level inference results on dark matter halos, considering the same sample as studied in [26]. That is, we extend the forward model adopted in [26] to include non-Gaussian stochasticity.

1.1 Outline of the paper

We outline the main calculations of the paper here, providing a guide for the reader.

In Sec. 2, we adopt the partition-function formulation for the noise from [4], in which higher powers of the current J encode information about the non-Gaussian noise. We show how integrating out J yields the likelihood. We then examine two limiting cases of this likelihood in Sec. 2.3, and discuss the most general case, obtained by expanding terms of order J^3 and higher, in Sec. 2.4. We discuss issues with this formal likelihood expansion, including negative probability densities, that preclude it from being used for field-level inference in practice.

We then turn to the ansatz Eq. (1.3), which introduces a single Gaussian noise field with nonlinear couplings and generalized bias coefficients in Sec. 3. We demonstrate that this formulation leads to a field-level likelihood that is equivalent, order by order, to that obtained in Sec. 2.4 from the partition-function formulation. At the same time, this formulation avoids the issues of the formal likelihood expansion by essentially resumming higher-order terms to a positive-definite probability distribution.

In Sec. 4, we present the first numerical implementation of field-level non-Gaussian noise, by introducing the noise as an additional noise field to be sampled. We show that this noise formulation produces more stable results than the previously considered Gaussian noise model. We conclude in Sec. 5.

1.2 Notation

We introduce the notation of the paper in this section. For the momenta integrals, we use

$$\int_{\mathbf{p}_1, \dots, \mathbf{p}_n} = \int \frac{d^3 p_1}{(2\pi)^3} \cdots \int \frac{d^3 p_n}{(2\pi)^3}. \quad (1.4)$$

We consider W_Λ sharp-in- k filters at a scale Λ (and similarly k_{max}) and write the filtered fields as

$$f_\Lambda(\mathbf{k}) = W_\Lambda(\mathbf{k})f(\mathbf{k}). \quad (1.5)$$

The Gaussian linear density field $\delta_{\text{in},\Lambda}$ power spectra is given by

$$P_L^\Lambda(k) = \langle \delta_{\text{in},\Lambda}(\mathbf{k}) \delta_{\text{in},\Lambda}(\mathbf{k}') \rangle', \quad (1.6)$$

where we use the prime to encapsulate the momenta conservation in the n -point functions

$$\langle O(\mathbf{k}_1) \dots O(\mathbf{k}_n) \rangle = \hat{\delta}_D(\mathbf{k}_{1\dots n}) \langle O(\mathbf{k}_1) \dots O(\mathbf{k}_n) \rangle', \quad (1.7)$$

with $\hat{\delta}_D = (2\pi)^3 \delta_D$ the (3D) Dirac delta distribution. The variance of the linear density field on a scale Λ is

$$\sigma_\Lambda^2 = \int_{\mathbf{p}}^\Lambda P_L(p). \quad (1.8)$$

Throughout this work we assume initial conditions smoothed at a scale $\Lambda \gg k$, for all k values for which observables are evaluated. To avoid cluttering the text, we remove the Λ subscript from the fields in the following sections, but it should be assumed the initial conditions are always smoothed on the scale Λ .

The bias operators are constructed on top of the linear fields via the convolution with a K_O kernel

$$O(\mathbf{k}) = \sum_{n=n(O)}^\infty \int_{\mathbf{p}_1, \dots, \mathbf{p}_n} \hat{\delta}_D(\mathbf{k} - \mathbf{p}_{1\dots n}) K_O^{(n)}(\mathbf{p}_1, \dots, \mathbf{p}_n) \delta_{\text{in}}(\mathbf{p}_1) \cdots \delta_{\text{in}}(\mathbf{p}_n). \quad (1.9)$$

where $n(O)$ is the leading perturbative order at which the operator O contributes.

We always subtract the mean of nontrivial operators,

$$O \rightarrow O - \langle O \rangle. \quad (1.10)$$

We consider the general bias expansion consisting of a set of operators O , ordered in perturbation theory and spatial derivatives. Our results are valid at all orders in perturbations, so we do not consider a fixed maximum order here. Throughout, higher-derivative operators are also included, noting that they are controlled by k^2/k_{nl}^2 , where k_{nl} is the non-linear scale where perturbation theory fails, or $k^2 R_*^2$, where R_* is the spatial length scale associated with the formation of the galaxies considered. We also consider the (zeroth-order) unit operator

$$\mathbb{1}(\mathbf{k}) = \hat{\delta}_D(\mathbf{k}), \quad (1.11)$$

which only contributes to stochastic terms (higher-order-in-current operators) since the unit operator is removed from the bias basis by Eq. (1.10) (equivalent to demanding $\langle \delta_g \rangle = 0$).

In either the non-minimal or Gaussian-noise formulation presented above, the noise field ϵ is characterized by a two-point function given by

$$\langle \epsilon(\mathbf{x}) \epsilon(\mathbf{y}) \rangle = [P_{\epsilon, \mathbb{1}} + P_{\epsilon, \nabla^2 \mathbb{1}} \nabla_{\mathbf{x}}^2 + \dots] \delta_D(\mathbf{x} - \mathbf{y}), \quad (1.12)$$

corresponding to a local stochastic process. Here, we have also written the leading higher-derivative term to illustrate their structure. The field ϵ is by definition uncorrelated with the initial density field:

$$\langle \epsilon(\mathbf{k}) \delta_{\text{in}}(\mathbf{k}') \rangle = 0, \quad \text{and moreover} \quad \langle \epsilon(\mathbf{k}) \delta_{\text{in}}(\mathbf{k}_1) \cdots \delta_{\text{in}}(\mathbf{k}_n) \rangle = 0, \quad (1.13)$$

from which follows $\langle \epsilon(\mathbf{k}) O[\delta_{\text{in}}](\mathbf{k}') \rangle = 0$. See App. A for examples on how the noise field appears in n -point functions for both formulations Eqs. (1.2)–(1.3).

Index notation. We widely use the index notation for fields in Fourier space:

$$X(\mathbf{k}_i) \rightarrow X^i, \quad \text{and} \quad X(-\mathbf{k}_i) \rightarrow X^{-i}. \quad (1.14)$$

The integral over a field is written as

$$\int \mathcal{D}X \equiv \prod_i \int dX^i. \quad (1.15)$$

We define the corresponding Dirac delta distribution by

$$\hat{\delta}_D(\mathbf{k}_{j_1 \dots j_m} + \mathbf{k}_{i_1 \dots i_n}) \equiv \hat{\delta}_{D, i_1 \dots i_n}^{j_1 \dots j_m}, \quad (1.16)$$

which will be useful notationally. Contracted indices indicate an integral, e.g.

$$\begin{aligned} \hat{\delta}_{D, ij} X^i Y^j &= X^i Y^{-i} \equiv \int_{\mathbf{k}} X(\mathbf{k}) Y(-\mathbf{k}), \\ \hat{\delta}_{D, ijk} X^i Y^j Y^k &\equiv \int_{\mathbf{k}_1, \mathbf{k}_2} X(\mathbf{k}_1) Y(\mathbf{k}_2) Y(-\mathbf{k}_{12}). \end{aligned} \quad (1.17)$$

With this, Eq. (1.1) reads

$$\delta_{g, \det}^i[\{b_O\}, \delta_{\text{in}}] = \sum_O b_O O^i[\delta_{\text{in}}]. \quad (1.18)$$

Gaussian action. The Gaussian free action or prior is given by the matter 2-point function

$$\mathcal{P}[\delta_{\text{in}}] = \left(\prod_{\mathbf{k}}^{\Lambda} 2\pi P_L(k) \right)^{-1/2} \exp \left[-\frac{1}{2} \int_{\mathbf{k}}^{\Lambda} \frac{|\delta_{\text{in}}|^2}{P_L(k)} \right]. \quad (1.19)$$

We drop any prefactors in the likelihood that are independent of the parameters of interest (in particular powers of 2π and such), since those are irrelevant for the inference of parameters in the field δ_{in} . Using the index notation, we can write Eq. (1.19) as

$$\mathcal{P}[\delta_{\text{in}}] = \left(\prod_{\mathbf{k}}^{\Lambda} 2\pi P_L(k) \right)^{-1/2} \exp \left[-\frac{1}{2} \delta_{\text{in}}^i (P_L^{-1})_{ij} \delta_{\text{in}}^j \right], \quad \text{with} \quad (P_L^{-1})_{ij} = [P_L(k_i)]^{-1} \hat{\delta}_D^{ij}. \quad (1.20)$$

Throughout, we will suppress time arguments, and often the dependence on cosmological parameters as well.

We will make much use of the well-known Gaussian integral identity

$$\int d^d X \exp \left[-\frac{1}{2} \mathbf{X}^\top \mathbf{A} \mathbf{X} + \mathbf{B}^\top \mathbf{X} \right] = \frac{(2\pi)^{d/2}}{|\mathbf{A}|^{1/2}} \exp \left[\frac{1}{2} \mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B} \right], \quad (1.21)$$

and use bold-face to denote field-space objects with one or two indices, in those cases where index contractions are obvious.

Kernel definition. We now introduce general kernels used in the paper, for reference. Since they make reference to concepts introduced later, they can be skipped at first reading. We define

$$\mathcal{K}^{\{m\}, j_1 \dots j_m}[\{C_O^{\{m\}}\}, \delta_{\text{in}}] \equiv \sum_{\mathbb{1}, O} C_O^{\{m\}} \sum_{n=n(O)}^{\infty} \hat{\delta}_{D, i_1 \dots i_n}^{j_1 \dots j_m} K_O^{(n)}(\mathbf{k}_{i_1}, \dots, \mathbf{k}_{i_n}) \delta_{\text{in}}^{i_1} \dots \delta_{\text{in}}^{i_n}. \quad (1.22)$$

Consistent with Eq. (1.3), we use braces to denote the m -th order in the current or noise field, and parentheses for the n -th order in perturbation theory. These kernels depend on a set of coefficients $\{C_O^{\{m\}}\}$ at fixed m which we will discuss in the following section. Note

that we always include the unit operator in the basis. We also define μ and Σ for the cases $m = 1, 2$ respectively:

$$\begin{aligned}
\mu^j[\{b_O\}, \delta_{\text{in}}] &\equiv \mathcal{K}^{\{1\},j}[\{b_O\}, \delta_{\text{in}}] = \sum_{\mathbb{1}, O} b_O \sum_{n=n(O)}^{\infty} \hat{\delta}_{\text{D}, i_1 \dots i_n}^j K_O^{(n)}(\mathbf{k}_{i_1}, \dots, \mathbf{k}_{i_n}) \delta_{\text{in}}^{i_1} \dots \delta_{\text{in}}^{i_n} \\
&= \delta_{g, \text{det}}^{-j}[\{b_O\}, \delta_{\text{in}}] = \sum_O b_O O^{-j}[\delta_{\text{in}}], \\
\Sigma^{jk}[\{C_O^{\{2\}}\}, \delta_{\text{in}}] &\equiv \mathcal{K}^{\{2\},jk}[\{C_O^{\{2\}}\}, \delta_{\text{in}}] = \sum_{\mathbb{1}, O} C_O^{\{2\}} \sum_{n=n(O)}^{\infty} \hat{\delta}_{\text{D}, i_1 \dots i_n}^{jk} K_O^{(n)}(\mathbf{k}_{i_1}, \dots, \mathbf{k}_{i_n}) \delta_{\text{in}}^{i_1} \dots \delta_{\text{in}}^{i_n} \\
&= C_{\mathbb{1}}^{\{2\}} \hat{\delta}_{\text{D}}^{jk} + \hat{\delta}_{\text{D}, l}^{jk} \sum_O C_O^{\{2\}} O^l[\delta_{\text{in}}].
\end{aligned} \tag{1.23}$$

2 From the partition function to the likelihood

In this section, the goal is to obtain the EFT likelihood starting from the EFT partition function.² We present the partition function in Sec. 2.1 and calculate the likelihood in Sec. 2.2. In Sec. 2.3 we discuss some simplified cases, and, in Sec. 2.4, perform a formal expansion of the likelihood in terms of, essentially, moments of the galaxy stochasticity.

2.1 The EFT partition function

In this section we review the EFTofLSS partition function including noise terms. In [4] (see also [27, 28]), it was shown that the partition function can be written as

$$\begin{aligned}
Z[\mathbf{J}] &= \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp(S_{\text{eff}}[\delta_{\text{in}}, \mathbf{J}]), \\
\text{with } S_{\text{eff}}[\delta_{\text{in}}, \mathbf{J}] &\equiv \sum_{m=1}^{\infty} \sum_{\mathbb{1}, O} \frac{C_O^{\{m\}}}{m!} \int_{\mathbf{x}} [J(\mathbf{x})]^m O[\delta_{\text{in}}](\mathbf{x}).
\end{aligned} \tag{2.1}$$

Provided that the set $\{O\}$ is a complete set of linearly independent bias operators (see [1]), this partition function can be shown to be closed under renormalization [4] (see also [29–31]). In the partition function, the terms with $m = 1$ correspond to the usual bias parameters, $b_O = C_O^{\{1\}}$, while the terms with $m \geq 2$ correspond to stochastic contributions. Notice that, for $m = 1$, the zeroth-order $O = \mathbb{1}$ is removed by Eq. (1.10). Higher-derivative contributions for both the deterministic and stochastic fields are included in the partition function via current-derivative terms such as $\nabla^2 J$ and $\partial_i J \partial^i J$. We focus in this work on leading-in-derivative operators, but the results are straightforwardly generalized to include higher-derivative terms, including then derivative expansions of the noise field such as those in Eq. (1.12).

The relatively abstract expression Eq. (2.1) becomes more concrete when calculating n -point correlation functions, obtained by taking n derivatives of the partition function with

²Strictly speaking, we will be deriving the posterior for cosmological and bias parameters given the galaxy density field, while the likelihood is part of the integrand, see Sec. 2.2. We continue to use the term “likelihood” here loosely, following the literature (e.g. [27]).

respect to the current $\mathbf{J}$. Since

$$\langle \delta_g(\mathbf{k}_1) \dots \delta_g(\mathbf{k}_n) \rangle = \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \delta_g(\mathbf{k}_1) \dots \delta_g(\mathbf{k}_n) e^{S_{\text{eff}}}, \quad (2.2)$$

the n -point function is given by

$$\langle \delta_g(\mathbf{k}_1) \dots \delta_g(\mathbf{k}_n) \rangle = \frac{1}{Z[\mathbf{J} = \mathbf{0}]} \frac{\delta^n Z}{\delta J(\mathbf{k}_1) \dots \delta J(\mathbf{k}_n)} \Big|_{\mathbf{J}=\mathbf{0}}. \quad (2.3)$$

Therefore, as derived in [4]³

$$\begin{aligned} \langle \delta_g(\mathbf{k}) \rangle &= C_{\mathbb{1}}^{\{1\}} \hat{\delta}_{\text{D}}(\mathbf{k}) = 0, \\ \langle \delta_g(\mathbf{k}_1) \delta_g(\mathbf{k}_2) \rangle &= \sum_O \sum_{O'} C_O^{\{1\}} C_{O'}^{\{1\}} \langle O(\mathbf{k}_1) O'(\mathbf{k}_2) \rangle + C_{\mathbb{1}}^{\{2\}} \hat{\delta}_{\text{D}}(\mathbf{k}_{12}), \\ \langle \delta_g(\mathbf{k}_1) \delta_g(\mathbf{k}_2) \delta_g(\mathbf{k}_3) \rangle &= \sum_O \sum_{O'} \sum_{O''} C_O^{\{1\}} C_{O'}^{\{1\}} C_{O''}^{\{1\}} \langle O(\mathbf{k}_1) O'(\mathbf{k}_2) O''(\mathbf{k}_3) \rangle \\ &\quad + \left(\sum_O \sum_{O'} C_O^{\{1\}} C_{O'}^{\{2\}} \langle O(\mathbf{k}_1) O'(\mathbf{k}_{23}) \rangle + 2 \text{ perm.} \right) + C_{\mathbb{1}}^{\{3\}} \hat{\delta}_{\text{D}}(\mathbf{k}_{123}). \end{aligned} \quad (2.4)$$

By simple dimensional analysis we have

$$[O(\mathbf{x})] = d_O \quad \Rightarrow \quad [O(\mathbf{k})] = d_O - 3, \quad (2.5)$$

$$[J(\mathbf{x})] = 3 \quad \Rightarrow \quad [J(\mathbf{k})] = 0, \quad (2.6)$$

$$[C_O^{\{m\}}] = - \left[\int_{\mathbf{x}} \mathbf{J}^m O \right] = 3 - 3m - d_O. \quad (2.7)$$

where $d_O = n_{\text{deriv}}(O)$ is the number of derivatives in the operator O . We emphasize that the coefficients $C_O^{\{m\}}$, defined in Eq. (2.1) are not the same as $b_O^{\{m\}}$ introduced in Eq. (1.3), the latter having dimension $-d_O$ while, for $d_O = 0$, $C_O^{\{m\}}$ have dimensions of an m -point function in Fourier space. We will return to the precise relation between the $C_O^{\{m\}}$ and the $b_O^{\{m\}}$ in the following section.

Using the index notation, Eq. (1.20), and the kernels $\mathcal{K}$ introduced in Eq. (1.22), the partition function can also be written as

$$\begin{aligned} Z[\mathbf{J} | \{C_O^{\{m\}}\}] &= \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[J_i \mu^i [\{b_O\}, \delta_{\text{in}}] + \frac{1}{2} J_i J_j \Sigma^{ij} [\{C_O^{\{2\}}\}, \delta_{\text{in}}] \right. \\ &\quad \left. + \frac{1}{6} J_i J_j J_k \mathcal{K}^{\{3\}, ij k} [\{C_O^{\{3\}}\}, \delta_{\text{in}}] + \dots \right]. \end{aligned} \quad (2.8)$$

For more examples of the momentum structure of these terms, see [4]. Note once more that $\mu^i = \mathcal{K}^{\{1\}, i} = \delta_{g, \text{det}}^{-i}$ and that $\Sigma^{ij} = \mathcal{K}^{\{2\}, ij}$.

³Ref. [4] was lacking the Dirac factors with the $C_{\mathbb{1}}^{\{m\}}$ terms in their Eqs. (2.8–2.10).

2.2 EFT likelihood from the partition function

Rather than n -point functions, we are interested here in the “likelihood” for the galaxy density field. First, this is the crucial ingredient in field-level inference approaches and for EFT-based generative models, which are required for simulation-based inference, for example [26, 32]. Second, the field-level likelihood allows us to formulate field-level expressions such as Eqs. (1.2)–(1.3) with more rigor, as it is not immediately obvious how such relations follow from the partition function.

The *posterior* for a set of cosmological parameters θ and EFT coefficients $C_O^{\{m\}}$ given some observed data $\hat{\delta}_g$ is given by

$$\mathcal{P} \left[\{\theta\}, \{C_O^{\{m\}}\} \mid \hat{\delta}_{g,k_{\max}} \right] \propto \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}} \mid \{\theta\}] \mathcal{L}_{k_{\max}}[\hat{\delta}_{g,k_{\max}} \mid \delta_{\text{in}}, \{\theta\}, \{C_O^{\{m\}}\}] \mathcal{P}(\{\theta\}, \{C_O^{\{m\}}\}), \quad (2.9)$$

where $\mathcal{L}_{k_{\max}}$ is the *likelihood* proper, which depends on δ_{in} (including modes up to Λ) through the bias operators $O[\delta_{\text{in}}]$, and the subscript $k_{\max}$ indicates that only modes in the data up to $k_{\max}$ are included. Following [33], this corresponds to cutting all external momenta of n -point functions of the data at $k_{\max}$. Finally, $\mathcal{P}(\{\theta\}, \{C_O^{\{m\}}\})$ denotes the prior on EFT and cosmological parameters.

Our main goal in this section is to connect the likelihood $\mathcal{L}$ to the partition function. In order to keep the equations simple, we will drop the prior $\mathcal{P}(\{\theta\}, \{C_O^{\{m\}}\})$ in the following, and also keep only the dependence of the data in the posterior explicit, defining

$$\mathcal{P}'[\hat{\delta}_{g,k_{\max}}] \equiv \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}} \mid \{\theta\}] \mathcal{L}_{k_{\max}}[\hat{\delta}_{g,k_{\max}} \mid \delta_{\text{in}}, \{\theta\}, \{C_O^{\{m\}}\}], \quad (2.10)$$

using a prime to differentiate from Eq. (2.9). Once an explicit expression for $\mathcal{P}'$ is obtained, we can then read off the likelihood proper, $\mathcal{L}_{k_{\max}}$, from the integrand.

The quantity $\mathcal{P}'[\hat{\delta}_{g,k_{\max}}]$ represents the probability of finding the observed tracer field $\hat{\delta}_{g,k_{\max}}$, given (implicit) fixed values for the cosmological and EFT parameters θ , $C_O^{\{m\}}$ and including modes up to $k_{\max}$, marginalized over the entire initial conditions δ_{in} (up to Λ). Following [27], this probability can be written as a field-level Dirac delta functional, which we in turn represent in Fourier space:

$$\begin{aligned} \mathcal{P}'[\hat{\delta}_{g,k_{\max}}] &= \int \mathcal{D}\delta_{g,k_{\max}} \mathcal{P}[\delta_{g,k_{\max}}] \delta_{\text{D}}^{[0,k_{\max}]}(\hat{\delta}_{g,k_{\max}} - \delta_{g,k_{\max}}) \\ &= \int \mathcal{D}(\text{i}^{-1} \mathbf{J}_{k_{\max}}) \left\langle \exp \left[J_{k_{\max}}^i (\hat{\delta}_g - \delta_g)_{-i} \right] \right\rangle \\ &= \int \mathcal{D}(\text{i}^{-1} \mathbf{J}_{k_{\max}}) \exp \left[J_{k_{\max}}^i (\hat{\delta}_g)_{-i} \right] \left\langle \exp \left[-J_{k_{\max}}^i (\delta_g)_{-i} \right] \right\rangle \\ &= (Z[\mathbf{J} = \mathbf{0}])^{-1} \int \mathcal{D}(\text{i} \mathbf{J}_{k_{\max}}) \exp \left[-J_{k_{\max}}^i (\hat{\delta}_g)_{-i} \right] Z[\mathbf{J}_{k_{\max}}]. \end{aligned} \quad (2.11)$$

Here, $\hat{\delta}_g$ represents the point in field space at which we evaluate the probability $\mathcal{P}'$ (e.g., the observed data), while δ_g represents the random field itself. We denote the field-level Dirac delta that includes all modes in the field up to $k_{\max}$ (except for the zero mode) with $\delta_{\text{D}}^{[0,k_{\max}]}$. In the second line, we have used the Fourier representation of the Dirac delta in field space,

and expressed the integral over $\delta_{g,k_{\max}}$ as expectation value. That is, the expectation values in the second and third lines are taken with respect to the measure $\mathcal{P}[\delta_{\text{in}}|\{\theta\}]$, just as for n -point correlators. Note that we have chosen an imaginary current $\mathbf{J}$ here for convenience, so that the integral measure is accompanied by a factor i^{-1} , and changed the sign of $\mathbf{J}$ in the last line. We have emphasized that the current $\mathbf{J}$ employed here likewise only has support up to $k_{\max}$, and that $Z[\mathbf{J} = \mathbf{0}]$ is the partition function evaluated at zero current, which still has a nontrivial dependence on the parameters $\{C_O^{\{m\}}\}$, implicit here but important to obtain a normalized probability distribution. The last equality can be shown via a formal Taylor expansion

$$\begin{aligned} \langle \exp [J_{k_{\max}}^i (\delta_g)_{-i}] \rangle &= \sum_{l=0}^{\infty} \frac{1}{l!} J_{k_{\max}}^{i_1} \cdots J_{k_{\max}}^{i_l} \langle (\delta_g)_{-i_1} \cdots (\delta_g)_{-i_l} \rangle \\ &= (Z[\mathbf{J} = \mathbf{0}])^{-1} \sum_{l=0}^{\infty} \frac{1}{l!} J_{k_{\max}}^{i_1} \cdots J_{k_{\max}}^{i_l} \frac{\mathcal{D}}{\mathcal{D}J_{k_{\max}}^{i_1}} \cdots \frac{\mathcal{D}}{\mathcal{D}J_{k_{\max}}^{i_l}} Z[\mathbf{J}_{k_{\max}}] \Big|_{\mathbf{J}_{k_{\max}}=\mathbf{0}} \\ &= (Z[\mathbf{J} = \mathbf{0}])^{-1} Z[\mathbf{J}_{k_{\max}}]. \end{aligned} \quad (2.12)$$

Reinstating the explicit dependencies [but dropping the prior $\mathcal{P}(\{\theta\}, \{C_O^{\{m\}}\})$], this becomes

$$\begin{aligned} \mathcal{P}[\{\theta\}, \{C_O^{\{m\}}\} | \hat{\delta}_{g,k_{\max}}] &\propto \\ \frac{1}{Z[\mathbf{0}|\{\theta\}, \{C_O^{\{m\}}\}]} &\int \mathcal{D}(i\mathbf{J}_{k_{\max}}) \exp \left[-J_{k_{\max}}^{-i} (\hat{\delta}_{g,k_{\max}})_i \right] Z[\mathbf{J}_{k_{\max}}|\{\theta\}, \{C_O^{\{m\}}\}]. \end{aligned} \quad (2.13)$$

Eq. (2.13) states that the EFT posterior is obtained as the functional Fourier transform of the partition function, in agreement with [27], but making the cutoff $k_{\max}$ explicit. Notice that, since the current is cut at $k_{\max}$, no modes above this value are excited in the partition function. Thus, the likelihood is related to the partition function in precisely the same regime as n -point functions up to the same $k_{\max}$. Finally, note that we have kept the normalization $Z[\mathbf{0}|\{\theta\}, \{C_O^{\{m\}}\}]$, as it is parameter-dependent and hence contributes nontrivially to the posterior in $\{\theta\}, \{C_O^{\{m\}}\}$ (any parameter-independent constants on the other hand can safely be ignored, as we are not attempting to compute the normalizing evidence). In the following, we will neglect the dependence on the cosmological parameters $\{\theta\}$, since it no longer plays any role in the derivation.

2.3 Limiting cases

While we have formally obtained the likelihood as Eq. (2.13), the integration over $\mathbf{J}$ in Eq. (2.13) is not tractable for the general $Z[\mathbf{J}]$ in Eq. (2.1). In the following, we will discuss simpler limiting cases for which the integration can be done analytically. While in Sec. 2.3.1 we truncate the $\mathbf{J}$ series keeping only terms up to $\mathbf{J}^2$, in Sec. 2.3.2 we keep all powers in $\mathbf{J}$ but neglect the coupling with δ_{in} . Finally, in Sec. 2.4 we consider an expansion in the current, which is equivalent to expanding around the case considered in Sec. 2.3.1.

2.3.1 Quadratic-in-current terms and all operators

We start by restricting the partition function Eq. (2.1) to terms up to second-order in the current $\mathbf{J}_{k_{\max}}$ and including the most general set of bias operators $O[\delta_{\text{in}}]$. Inserting that into

Eq. (2.13) yields

$$\begin{aligned} \mathcal{P}'[\hat{\delta}_{g,k_{\max}}] &= (Z[\mathbf{0}|\{b_O, C_O^{\{2\}}\}])^{-1} \int \mathcal{D}(\mathbf{i} \mathbf{J}_{k_{\max}}) \exp \left[-J_{k_{\max}}^i (\hat{\delta}_g)_{-i} \right] \\ &\times \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[J_{k_{\max},i} \mu^i [\{b_O\}, \delta_{\text{in}}] + \frac{1}{2} J_{k_{\max},i} J_{k_{\max},j} \Sigma^{ij} [\{C_O^{\{2\}}\}, \delta_{\text{in}}] \right], \end{aligned} \quad (2.14)$$

where we have used μ defined in Eq. (1.23), and

$$\Sigma^{ij}[\{C_O^{\{2\}}\}, \delta_{\text{in}}] = C_{\mathbb{1}}^{\{2\}} \hat{\delta}_{\text{D}}^{ij} + \hat{\delta}_{\text{D},k}^{ij} \sum_O C_O^{\{2\}} O^k [\delta_{\text{in}}]. \quad (2.15)$$

We omit the higher-derivative terms in Eq. (1.12), but one can straightforwardly generalize the conclusions of this section to include them.

In this case we can perform the Gaussian integral over $\mathbf{J}_{k_{\max}}$ in Eq. (2.14) using Eq. (1.21), yielding

$$\mathcal{P}'[\hat{\delta}_{g,k_{\max}}] = \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \mathcal{L}_{k_{\max}}[\hat{\delta}_{g,k_{\max}}|\delta_{\text{in}}, \{\theta\}, \{C_O^{\{1,2\}}\}]$$

with

$$\begin{aligned} \mathcal{L}_{k_{\max}}[\hat{\delta}_{g,k_{\max}}|\delta_{\text{in}}, \{\theta\}, \{C_O^{\{1,2\}}\}] &= \mathcal{N}_{\mathcal{L}}[\{C_O^{\{2\}}\}, \delta_{\text{in}}] \\ &\times \exp \left[Y^i [\hat{\delta}_{g,k_{\max}}, \{b_O\}, \delta_{\text{in}}] \left(\Sigma^{-1}[\delta_{\text{in}}, \{C_O^{\{2\}}\}] \right)_{ij} Y^j [\hat{\delta}_{g,k_{\max}}, \{b_O\}, \delta_{\text{in}}] \right], \end{aligned} \quad (2.16)$$

and

$$Y^i [\hat{\delta}_{g,k_{\max}}, \{b_O\}, \delta_{\text{in}}] \equiv \hat{\delta}_{g,k_{\max}}^{-i} - \delta_{g,\text{det}}^{-i} [\{b_O\}, \delta_{\text{in}}] = \hat{\delta}_{g,k_{\max}}^{-i} - \mu^i [\{b_O\}, \delta_{\text{in}}], \quad (2.17)$$

which we will widely use below, as the likelihood becomes centered in terms of Y . Note that the coupling with $\mathbf{J}_{k_{\max}}$ in Eq. (2.14) ensures that both $\hat{\delta}_g$ and μ are cut at $k_{\max}$, hence we will always only encounter the filtered $\mathbf{Y} = \mathbf{Y}_{k_{\max}}$, and throughout we omit $k_{\max}$. Further, we defined the likelihood normalization

$$\mathcal{N}_{\mathcal{L}}[\{C_O^{\{2\}}\}, \delta_{\text{in}}] \propto (Z[\mathbf{J} = \mathbf{0}])^{-1} |\Sigma[\delta_{\text{in}}]|^{-1/2}. \quad (2.18)$$

In the following, we will drop the explicit parameter dependence in μ , Σ , and Y for clarity.

It is clear that the likelihood (proper) in Eq. (2.16) can be interpreted as being due to a single, Gaussian stochastic degree of freedom ϵ^i with covariance $\Sigma^{ij}[\delta_{\text{in}}]$. The evaluation of this likelihood however requires inverting the matrix Σ and evaluating its determinant. If we keep only the first, diagonal term in Eq. (2.15), Σ^{ij} is easily inverted; in fact, Eq. (2.16) then becomes the Fourier-space likelihood first derived in [34], where the data $\hat{\delta}_{g,k_{\max}}$ and model prediction Eq. (1.18) are compared with a diagonal, Fourier-space covariance up to $k_{\max}$. Going beyond the leading term in Eq. (2.15) however, i.e. when attempting to include the coupling between δ_{in} and noise, this matrix can no longer be inverted analytically in general. Moreover, it is a dense matrix, which renders a numerical evaluation of the likelihood Eq. (2.16) essentially intractable.⁴ An alternative route was proposed in [28], by instead formulating the likelihood in real space. This approach is unfortunately hampered by the need to enforce the Fourier-space cut at $k_{\max}$. We discuss this in App. B.

These issues aside, Eq. (2.16) is not complete in any case, as we have truncated the partition function at order $\mathbf{J}^2$. We turn to the higher-order terms in $\mathbf{J}$ next.

⁴In a field-level analysis, the likelihood has to be evaluated at every step in a sampling process. Since the matrix Σ^{ij} has dimension $D \times D$, where D is the total dimension of the density field $D = N_{\text{grid}}^3$, a matrix inversion is impractical in reality. In addition, a matrix of this size cannot be stored in memory.

2.3.2 All-orders-in-current terms but only unit operator $O = \mathbb{1}$

Let us now study a second instructive case, the limit in which we keep only the coefficients $C_{\mathbb{1}}^{\{m\}}$, i.e. the subset of kernels in Eq. (1.22) given by $C_{\mathbb{1}}^{\{m\}} \hat{\delta}_{\text{D}}^{j_1 \dots j_m}$ for $m \geq 2$. That is, we keep arbitrary powers of the current, but neglect the coupling with δ_{in} in the stochastic ($m \geq 2$) contributions, while keeping the full set of deterministic ($m = 1$) terms. At the level of n -point functions, this corresponds to keeping only the purely stochastic “shot-noise” terms in Eq. (2.4).

In this limiting case, the posterior is given by

$$\begin{aligned} \mathcal{P}'[\hat{\delta}_{g, k_{\text{max}}}] &= (Z[\mathbf{J} = \mathbf{0}])^{-1} \int \mathcal{D}(\mathbf{i} \mathbf{J}_{k_{\text{max}}}) \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[-J_{k_{\text{max}}, i} Y^i[\hat{\delta}_{g, k_{\text{max}}}, \{b_O\}, \delta_{\text{in}}] \right. \\ &\quad \left. + \sum_{m=2}^{\infty} \frac{1}{m!} C_{\mathbb{1}}^{\{m\}} \hat{\delta}_{\text{D}}^{j_1 \dots j_m} J_{k_{\text{max}}, j_1} \cdots J_{k_{\text{max}}, j_m} \right], \end{aligned} \quad (2.19)$$

which we can write as

$$\mathcal{P}'[\hat{\delta}_{g, k_{\text{max}}}] = \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \mathcal{L}_{k_{\text{max}}} \left[\hat{\delta}_{g, k_{\text{max}}} - \mu_{k_{\text{max}}}[\{b_O\}, \delta_{\text{in}}], \{C_{\mathbb{1}}^{\{m \geq 2\}}\} \right], \quad (2.20)$$

where

$$\begin{aligned} \mathcal{L}_{k_{\text{max}}} \left[Y, \{C_{\mathbb{1}}^{\{m \geq 2\}}\} \right] &\equiv (Z[\mathbf{J} = \mathbf{0}])^{-1} \int \mathcal{D}(\mathbf{i} \mathbf{J}_{k_{\text{max}}}) \\ &\quad \times \exp \left[-J_{k_{\text{max}}, i}^i Y_i + \sum_{m=2}^{\infty} \frac{1}{m!} C_{\mathbb{1}}^{\{m\}} \hat{\delta}_{\text{D}}^{j_1 \dots j_m} J_{k_{\text{max}}, j_1} \cdots J_{k_{\text{max}}, j_m} \right]. \end{aligned} \quad (2.21)$$

Eq. (2.21) shows explicitly that, also for this limiting case, a single noise field with PDF $\mathcal{L}_{k_{\text{max}}}[\cdot]$ captures all stochastic contributions $C_{\mathbb{1}}^{\{m\}}$ to the tracer density field. Moreover, we can identify the coefficients $C_{\mathbb{1}}^{\{m\}}$ as the cumulants of the noise field (recall that Y is the residual between the data $\hat{\delta}_{g, k_{\text{max}}}$ and mean-field prediction $\mu_{k_{\text{max}}}$). Therefore, the partition function, with the linear-in-current term factored out, corresponds to the *characteristic functional* of the stochastic process. Notice that the contraction between Y and $J_{k_{\text{max}}}$ implies that $\mathcal{L}_{k_{\text{max}}}[Y]$ does not depend on momenta in Y that are beyond k_{max} .

Unfortunately, even if the coupling of stochasticity and initial conditions could be neglected as done here, Eq. (2.21) is not very practical. The EFT expansion demands that we employ a non-Gaussian probability distribution that allows for a flexible specification of an arbitrary number of higher-order cumulants (skewness, kurtosis, ...), depending on the order of the expansion considered, but no such closed-form distribution exists. Nevertheless, Eq. (2.21) could be interesting as a starting point for a resummation of noise contributions to an approximate, closed-form non-Gaussian PDF.

2.4 General expansion around the Gaussian likelihood

So far, we have considered the likelihood in the limits of only $C_O^{\{\leq 2\}} \neq 0$ (Sec. 2.3.1), and only $\{C_O^{\{1\}}, C_{\mathbb{1}}^{\{m\}}\} \neq 0$ (Sec. 2.3.2), respectively. In both cases, we could arrive at all-order expressions for the likelihood, and we found that all noise terms could be captured by a single non-Gaussian stochastic field (with the cumulants given by $C_{\mathbb{1}}^{\{m\}}$ in the latter case).

In this section, we push the results from Sec. 2.3.1 further, by perturbatively including the higher-in-current terms without neglecting the coupling to δ_{in} . Concretely, we perform a

perturbative expansion in the current $\mathbf{J}$, which corresponds to an Edgeworth-like expansion in the cumulants of the stochastic field. We emphasize that the expressions are valid at all orders in δ_{in} . We can write

$$\begin{aligned}\mathcal{P}'[\hat{\delta}_{g,k_{\text{max}}}] &= (Z[\mathbf{J} = \mathbf{0}])^{-1} \int \mathcal{D}(\mathbf{i} \mathbf{J}_{k_{\text{max}}}) \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[-J_{k_{\text{max}},i} Y^i + \frac{1}{2} J_{k_{\text{max}},i} J_{k_{\text{max}},j} \Sigma^{ij} \right] \\ &\quad \times \exp \left[\sum_{m=3}^{\infty} \frac{1}{m!} \mathcal{K}^{\{m\},j_1 \dots j_m} J_{k_{\text{max}},j_1} \dots J_{k_{\text{max}},j_m} \right] \\ &= (Z[\mathbf{J} = \mathbf{0}])^{-1} \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \int \mathcal{D}(\mathbf{i} \mathbf{J}_{k_{\text{max}}}) \exp \left[-J_{k_{\text{max}},i} Y^i + \frac{1}{2} J_{k_{\text{max}},i} J_{k_{\text{max}},j} \Sigma^{ij} \right] \\ &\quad \times \left[1 + \sum_{L=3}^{\infty} \frac{1}{L!} \mathcal{C}^{\{L\},j_1 \dots j_L} J_{k_{\text{max}},j_1} \dots J_{k_{\text{max}},j_L} \right],\end{aligned}\tag{2.22}$$

where in the second equation we expanded the exponential and we have defined (see App. C for the complete derivation)

$$\mathcal{C}^{\{L\},j_1 \dots j_L}[\{\mathcal{C}_O^{\{\geq 3\}}\}, \delta_{\text{in}}] = \sum_{\substack{a_3, \dots, a_L \geq 0 \\ 3a_3 + \dots + La_L = L}} \frac{L!}{a_3! \dots a_L!} \prod_{m=3}^L \frac{1}{(m!)^{a_m}} \left(\mathcal{K}^{\{m\}}[\{\mathcal{C}_O^{\{m\}}\}, \delta_{\text{in}}] \right)^{a_m},\tag{2.23}$$

such that $\mathcal{C}^{\{L\},j_1 \dots j_L}$, appearing at order $\mathbf{J}^L$, contains products of the kernels $\mathcal{K}^{\{m\}}$ (and the contributions are correspondingly multiplied by products of $\mathcal{C}_O^{\{m\}}$). For examples, see Eq. (C.6).

Our aim is to perform the $\mathbf{J}$ integral at fixed δ_{in} . In the following, we drop the arguments and the k_{max} subscript for clarity. We now shift the integration variable as⁵

$$J^i \rightarrow \tilde{J}^i = J^i - (\Sigma^{-1})^i_j Y^j.\tag{2.24}$$

This completes the square in Eq. (2.22), and yields

$$\begin{aligned}\mathcal{P}'[\hat{\delta}_{g,k_{\text{max}}}] &= (Z[\mathbf{J} = \mathbf{0}])^{-1} \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \int \mathcal{D}(\mathbf{i} \tilde{\mathbf{J}}) \exp \left[-\frac{1}{2} \mathbf{Y}^T \Sigma^{-1} \mathbf{Y} + \frac{1}{2} \tilde{\mathbf{J}}^T \Sigma \tilde{\mathbf{J}} \right] \\ &\quad \times \left[1 + \sum_{L=3}^{\infty} \frac{1}{L!} \mathcal{C}^{\{L\},j_1 \dots j_L} (\tilde{\mathbf{J}} + \Sigma^{-1} \mathbf{Y})_{j_1} \dots (\tilde{\mathbf{J}} + \Sigma^{-1} \mathbf{Y})_{j_L} \right].\end{aligned}\tag{2.25}$$

Now we have a sum over Gaussian integrals over $\tilde{\mathbf{J}}$, which can be done to yield (see App. C)

$$\mathcal{P}'[\hat{\delta}_{g,k_{\text{max}}}] = \mathcal{N}_{\mathcal{L}} \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[-\frac{1}{2} \mathbf{Y}^T \Sigma^{-1} \mathbf{Y} \right] \left[1 + \sum_{m=3}^{\infty} \frac{1}{m!} \tilde{\mathcal{C}}^{\{m\},j_1 \dots j_m} Y_{j_1} \dots Y_{j_m} \right],\tag{2.26}$$

where the coefficients

$$\begin{aligned}\tilde{\mathcal{C}}^{\{m\},j_1 \dots j_m} &= \sum_{\substack{L=m \\ L-m \text{ even}}}^{\infty} \frac{L!}{2^{(L-m)/2} (L-m)!} \mathcal{C}_{\text{symm}}^{\{L\},l_1 \dots l_L} (\Sigma^{-1})_{l_{m+1} l_{m+2}} \dots (\Sigma^{-1})_{l_{L-1} l_L} \\ &\quad \times (\Sigma^{-1})_{l_1 j_1} \dots (\Sigma^{-1})_{l_m j_m}\end{aligned}\tag{2.27}$$

⁵Note that, in Fourier space, J has dimension 0, Y has dimension -3 , Σ^{-1} has dimension 6, and the integration over j has dimension 3.

involve contractions of the fully-symmetrized versions $\mathcal{C}_{\text{symm}}^{\{L \geq m\}}$ [Eq. (C.7)] with powers of Σ^{-1} . Note that

$$\mathbf{Y} = \mathbf{Y}[\hat{\delta}_{g, k_{\max}}, \{b_O\}, \delta_{\text{in}}]; \quad \Sigma = \Sigma[\{C_O^{\{2\}}\}, \delta_{\text{in}}]; \quad \tilde{\mathcal{C}}^{\{m\}} = \tilde{\mathcal{C}}^{\{m\}}[\{C_O^{\{\geq 2\}}\}, \delta_{\text{in}}]. \quad (2.28)$$

As in Sec. 2.3.2, the expansion in J of the partition function leads to an expansion in Y of the likelihood. This expansion is valid if higher cumulants of the residual Y between data and mean-field prediction are suppressed. This is the case if the noise is perturbatively close to Gaussian (see also the discussion in [27]). The leading contribution to Eq. (2.27), in the sense of the expansion discussed, is given by $L = m$, such that every $C_O^{\{m\}}$ has a unique, leading-order contribution to $\tilde{\mathcal{C}}^{\{m\}}$ with distinct shape, indicating that the $C_O^{\{m\}}$ yield non-degenerate contributions to the field-level likelihood, just as they do in the partition function Eq. (2.1). Note that this equation also reproduces Eq. (2.16) if $\tilde{\mathcal{C}}^{\{m\}} = 0$ ($m \geq 3$), which holds of course if $\mathcal{C}^{\{L\}} = 0$ for $L \geq 3$.

Eq. (2.26) yields a consistent perturbative expansion in the EFT context, however it is not practical for real inference applications. The reason is twofold. First, the inversion of Σ still presents the same obstacle as discussed in Sec. 2.3.2. Second, when truncating at finite order m , the probability density becomes ill-defined (negative) in some regions of parameter space, and is not normalizable in general. This is typical of Edgeworth-like expansions in terms of cumulants.⁶ While higher-order terms in m should in principle be suppressed following the discussion above, a numerical sampler will explore the tails of the distribution, and eventually encounter the regime of ill-defined probability. To conclude, the EFT likelihood Eq. (2.13) is difficult to evaluate for the most general noise contributions.

However, we have seen in Sec. 2.3 that the stochastic contributions can be captured by *a single degree of freedom* both when restricted to the Gaussian case (but with mean and variance both coupled to δ_{in}) and the non-Gaussian case (with moments other than the mean not coupled to δ_{in}). The generalization when including subleading noise contributions, as we just saw, indicates that one can consider expansions around a Gaussian field. In the next section, we will build on this to develop an alternative route to evaluating the EFT likelihood in the fully general case, with the goal of obtaining a formulation that does not suffer from the above-mentioned problems, but yields an expansion that is perturbatively equivalent to Eq. (2.26).

3 The noise-field formulation of stochasticity

As we have seen in the previous section, the formal likelihood obtained directly from the partition function is difficult to evaluate numerically and may lead to an ill-defined probability density when expanded in the cumulants (corresponding to an expansion in powers of the current in the partition function). In this section we take a different approach, starting instead with a field-level ansatz for the noise. Our goal is to show that this field-level ansatz is equivalent to the partition function approach, order by order in perturbation theory. First,

⁶The Edgeworth expansion in terms of cumulants κ_ℓ can be written as

$$\mathcal{P}(\delta) = \frac{e^{-\delta^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \left[1 + \sum_{\ell} \kappa_\ell \frac{1}{\ell! \sigma^\ell} H_\ell(\delta) \right], \quad (2.29)$$

with H_ℓ being the Hermite polynomials. In general, when truncated at any finite order ℓ , there are regions in δ -space that have negative probability density.

this provides a crucial conceptual clarification of galaxy stochasticity, namely the statement that it is described precisely and unambiguously by Eq. (1.3), rather than Eq. (1.2). Second, thanks to the straightforward field-level formulation, this approach can be directly used in field-level inference analyses to incorporate coupled and non-Gaussian stochasticity at any order. We present results in Sec. 4.

3.1 Definition

Let us repeat the model of Eq. (1.3). In the following, we will shorten $\epsilon_G \rightarrow \epsilon$ for clarity of notation, since all instances of ϵ in this section correspond to the single Gaussian noise field introduced in Eq. (1.3):

$$\begin{aligned} \delta_g(\mathbf{x}, \tau) &= \delta_{g,\text{det}} + \sum_{m=1}^{\infty} \sum_{\mathbb{1}, O} b_O^{\{m\}}(\tau) [\epsilon(\mathbf{x})]^m O(\mathbf{x}, \tau), \\ \text{with } \delta_{g,\text{det}} &= \sum_O b_O^{\{0\}}(\tau) O(\mathbf{x}, \tau). \end{aligned} \quad (3.1)$$

We again keep the smoothing of the initial fields implicit, but in practice one would typically choose the same filtering scale Λ for both ϵ and δ_{in} . Precisely, the joint prior on ϵ and δ_{in} is given by

$$\mathcal{P}[\delta_{\text{in}}, \epsilon] = \mathcal{N}[\mathbf{0}, \text{diag}\{P_L(k)\}][\delta_{\text{in}}] \times \mathcal{N}[\mathbf{0}, \text{diag}\{P_\epsilon\}][\epsilon], \quad (3.2)$$

where P_ϵ is the constant power spectrum of the Gaussian field ϵ . The numerical value of P_ϵ depends on the normalization convention which we leave unspecified here (see Sec. 4 for the lattice implementation). The noise field ϵ can be understood as a “nuisance field” that captures how small-scale stochastic fluctuations affect the large-scale galaxy density field. In the following, we will again drop the time arguments as we did after Eq. (1.3).

The coefficients $b_O^{\{m\}}$ in Eq. (3.1) are not the same as the $C_O^{\{m\}}$. While the dimensionless $b_O^{\{m\}}$ are defined via the field-level formulation Eq. (3.1) and the index m corresponds to the number of contracted stochastic fields (starting at zero, for the deterministic part), the dimensionful $C_O^{\{m\}}$ are defined by the partition function of Sec. 2.1 and m corresponds to the number of contracted currents J (starting at $m = 1$ for the deterministic part). Specifically, we have

$$b_O^{\{0\}} = C_O^{\{1\}} \quad (3.3)$$

for the usual deterministic bias coefficients, while the Gaussian stochasticity contribution to the galaxy power spectrum is at leading order described by

$$(b_{\mathbb{1}}^{\{1\}})^2 P_\epsilon = C_{\mathbb{1}}^{\{2\}}, \quad (3.4)$$

respectively in the two formulations. The leading non-Gaussian stochasticity is controlled by $b_{\mathbb{1}}^{\{2\}}$ vs. $C_{\mathbb{1}}^{\{3\}}$. Generally, the term $\propto C_O^{\{m\}}$ in the partition function is indeed captured by $b_O^{\{m-1\}}$ in the noise-field formulation, but with additional corrections.

We can now rewrite Eq. (3.1) using the index notation Eq. (1.22) and Eq. (1.23) as

$$\begin{aligned} \delta_g^{-k} &= \delta_{g,\text{det}}^{-k} + \sum_{m=1}^{\infty} \sum_{\mathbb{1}, O} b_O^{\{m\}} \sum_{n=n(O)}^{\infty} \hat{\delta}_{D, i_1 \dots i_n}^{k j_1 \dots j_m} K_O^{(n)}(\mathbf{k}_{i_1}, \dots, \mathbf{k}_{i_n}) \delta_{\text{in}}^{i_1} \dots \delta_{\text{in}}^{i_n} \epsilon_{j_1} \dots \epsilon_{j_m} \\ &= \delta_{g,\text{det}}^{-k} + \sum_{m=1}^{\infty} \mathcal{K}^{\{m+1\}, k j_1 \dots j_m} [\{b_O^{\{m\}}\}, \delta_{\text{in}}] \epsilon_{j_1} \dots \epsilon_{j_m}, \end{aligned} \quad (3.5)$$

where the deterministic part is given by $m = 0$. Note that $\epsilon(\mathbf{k})$ has the same dimensions as $\delta_{g,\text{det}}(\mathbf{k})$ and $\delta_{\text{in}}(\mathbf{k})$. The kernels here are precisely the same as those in the expansion of the partition function in $\mathbf{J}$, we essentially just change coefficients from $\{C_O^{\{m\}}\}$ to $\{b_O^{\{m-1\}}\}$.⁷ Correspondingly, the contribution $\sim b_O^{\{m\}}$ involves a kernel with $m + 1$ upper indices, analogous to the contribution to $Z[\mathbf{J}]$ controlled by $C_O^{\{m+1\}}$. We point out that the recent Ref. [7] considered a similar expansion; we discuss the relation further below.

3.2 Expansion around the Gaussian likelihood

After having written down the field-level formulation Eq. (3.5), we can use the Gaussian prior (or free action) Eq. (3.2) for δ_{in} and ϵ to express the PDF of δ_g in general as

$$\begin{aligned} \mathcal{P}'[\hat{\delta}_{g,k_{\text{max}}}] &= \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \mathcal{N}_{\epsilon}[P_{\epsilon}] \int \mathcal{D}\epsilon \exp \left[-\frac{1}{2} P_{\epsilon}^{-1} \epsilon^T \epsilon \right] \\ &\quad \times \delta_D^{[0,k_{\text{max}}]} \left(Y^k[\hat{\delta}_{g,k_{\text{max}}}, \{b_O^{\{0\}}\}, \delta_{\text{in}}] - \sum_{m=1}^{\infty} \mathcal{K}^{\{m+1\},kj_1 \dots j_m}[\{b_O^{\{m\}}\}, \delta_{\text{in}}] \epsilon_{j_1} \dots \epsilon_{j_m} \right), \end{aligned} \quad (3.6)$$

where the variable Y is defined in Eq. (2.17), and $\mathcal{N}_{\epsilon}[P_{\epsilon}]$ is the normalization of the prior over ϵ (fixed, since P_{ϵ} is kept fixed). We have used k as placeholder index inside the field-level Dirac delta, in order to make the index structure clear. Our goal now is to integrate out the noise field ϵ , and make a connection to the expanded general PDF derived in the previous section, Eq. (2.26). The idea is that we can circumvent the complications found in Sec. 2 of terms nonlinear in the current by introducing the effective (Gaussian) noise field ϵ .

We solve the integral over ϵ recursively in m , inserting the solution for the zero of the Dirac delta at each order into the next-order calculation. We start by setting $m = 1$, i.e. considering only the first, linear-in- ϵ term in the sum inside the Dirac delta in Eq. (3.6). For this, we restrict the $\mathbf{k}$ support of ϵ to $(0, k_{\text{max}})$, the same range as for the current $\mathbf{J}_{k_{\text{max}}}$ in Sec. 2.4, and correspondingly restrict the matrix Σ to this range. Then, the Dirac delta fixes the solution for ϵ at linear order. If we were to allow for higher $\mathbf{k}$ support of ϵ , some modes in ϵ would remain unconstrained; however, we expect that the effect of integrating out these additional modes would simply shift the coefficients $b_{\mathbb{1},O}^{\{m\}}$. We obtain

$$\epsilon_{m=1}^k[\hat{\delta}_{g,k_{\text{max}}}, \{b_O^{\{0\}}\}, \{b_O^{\{1\}}\}, \delta_{\text{in}}] \equiv \left[\left(\mathcal{K}^{\{2\}}[\{b_O^{\{1\}}\}, \delta_{\text{in}}] \right)^{-1} \right]_{kj} Y^j[\hat{\delta}_{g,k_{\text{max}}}, \{b_O^{\{0\}}\}, \delta_{\text{in}}]. \quad (3.7)$$

Note that the arguments of Y are the same as in Sec. 2, since they refer to the deterministic component described by the bias parameters proper $b_O^{\{0\}}$ [see Eq. (3.3)]. In contrast, the arguments of $\mathcal{K}^{\{2\}}$ differ: while the partition function formulation is defined with $\Sigma[C_O^{\{2\}}, \delta_{\text{in}}] \equiv \mathcal{K}^{\{2\}}[C_O^{\{2\}}, \delta_{\text{in}}]$, the field-level description in this section uses $\mathcal{K}^{\{2\}}[b_O^{\{1\}}, \delta_{\text{in}}]$. Hereafter, we usually drop the arguments in Y and $\mathcal{K}^{\{m\}}$ for clarity. We also drop parameter-independent constants as they are irrelevant for the desired field-level posterior. Performing the integral over ϵ , we have

$$\mathcal{P}'[\hat{\delta}_{g,k_{\text{max}}}]|_{m=1} = \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \hat{\mathcal{N}}_{\mathcal{L}} \exp \left[-\frac{1}{2} \mathbf{Y}^T \hat{\Sigma}^{-1} \mathbf{Y} \right], \quad (3.8)$$

⁷Note that $\mathcal{K}^{\{m\}}[\{b_O^{\{m\}}\}]$ and $\mathcal{K}^{\{m\}}[\{C_O^{\{m\}}\}]$ inherit different dimensionality from the coefficients $b_O^{\{m\}}$ vs $C_O^{\{m\}}$.

where

$$\begin{aligned}\hat{\Sigma}^{ij}[\{b_O^{\{1\}}\}, \delta_{\text{in}}] &\equiv P_\epsilon \mathcal{K}^{\{2\}ik} \mathcal{K}^{\{2\}kj}[\{b_O^{\{1\}}\}, \delta_{\text{in}}] \\ \hat{\mathcal{N}}_{\mathcal{L}}[\{b_O^{\{1\}}\}, \delta_{\text{in}}] &\equiv N_\epsilon [P_\epsilon] |\hat{\Sigma}|^{-1/2}.\end{aligned}\quad (3.9)$$

While not immediately obvious, it is straightforward to show that $\hat{\Sigma}$ is equivalent to Σ in Eq. (2.16). This is because for any O_1 and O_2 in the basis of bias operators, the real-space product $O_1 O_2$ is also in the basis. The momentum structure of Eq. (3.9) is such that it precisely contains such real-space products (see App. A for a related argument at the n -point-function level). In addition to Eq. (3.4), we obtain

$$C_O^{\{2\}} = 2b_{\mathbb{1}}^{\{1\}} b_O^{\{1\}} P_\epsilon, \quad (3.10)$$

for any elementary operator O , i.e. one that cannot be written as a product $O_1 O_2$, and

$$C_{O_1 O_2}^{\{2\}} = \left[2b_{\mathbb{1}}^{\{1\}} b_{O_1 O_2}^{\{1\}} + b_{O_1}^{\{1\}} b_{O_2}^{\{1\}} \right] P_\epsilon, \quad (3.11)$$

otherwise.

We now incorporate higher-order-in-noise terms $m \geq 2$. For ϵ close to Gaussian, we can expand the Dirac delta in terms of its derivatives using its Fourier representation. Considering the field-level Dirac delta for a single mode with index k , we have

$$\begin{aligned}\delta_{\text{D}}^{[1]} \left(Y^k - \mathcal{K}^{\{2\},kj} \epsilon_j - \sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\},kj_1 \dots j_m} \epsilon_{j_1} \dots \epsilon_{j_m} \right) \\ = \int ds \exp \left[i s \left(Y^k - \mathcal{K}^{\{2\},kj} \epsilon_j \right) \right] \exp \left[i s \left(- \sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\},kj_1 \dots j_m} \epsilon_{j_1} \dots \epsilon_{j_m} \right) \right] \\ = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!} \left[\sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\},kj_1 \dots j_m} \epsilon_{j_1} \dots \epsilon_{j_m} \right]^\ell \left[\int ds (i s)^\ell \exp \left[i s \left(Y^k - \mathcal{K}^{\{2\},kj} \epsilon_j \right) \right] \right] \\ = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!} \left[\sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\},kj_1 \dots j_m} \epsilon_{j_1} \dots \epsilon_{j_m} \frac{\partial}{\partial Y^k} \right]^\ell \delta_{\text{D}}^{[1]} \left(Y^k - \mathcal{K}^{\{2\},kj} \epsilon_j \right).\end{aligned}\quad (3.12)$$

Note that the expansion in ℓ here corresponds to an expansion in powers of the *non-Gaussian* stochastic contribution (powers of $\mathcal{K}$ with $m+1 \geq 3$ upper indices), while the expansion in m continues to denote the order in the Gaussian noise field ϵ . The latter expansion is parametrically equivalent to that discussed in Sec. 2.4.

Using this result, the integral over ϵ becomes a sum over Gauss integrals with polynomial integrands:

$$\begin{aligned}\mathcal{P}'[\hat{\delta}_{g,k_{\text{max}}}] &= \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \mathcal{N}_\epsilon \int \mathcal{D}\epsilon \exp \left[-\frac{1}{2} P_\epsilon^{-1} \epsilon^T \epsilon \right] \\ &\times \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!} \prod_k \left\{ \left[\sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\},kj_1 \dots j_m} \epsilon_{j_1} \dots \epsilon_{j_m} \frac{\partial}{\partial Y^k} \right]^\ell \delta_{\text{D}}^{[1]} \left(Y^k - \mathcal{K}^{\{2\},kj} \epsilon_j \right) \right\},\end{aligned}\quad (3.13)$$

and, turning the product of component Dirac deltas into a field-level Dirac again,

$$\begin{aligned}
\mathcal{P}'[\hat{\delta}_{g,k_{\max}}] &= \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \mathcal{N}_{\epsilon} \int \mathcal{D}\epsilon \exp \left[-\frac{1}{2} P_{\epsilon}^{-1} \epsilon^T \epsilon \right] \\
&\times \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell}}{\ell!} \prod_{\kappa=1}^{\ell} \left[\sum_{m_{\kappa}=2}^{\infty} \mathcal{K}^{\{m_{\kappa}+1\}, k_{\kappa} j_{\kappa,1} \dots j_{\kappa, m_{\kappa}}} \epsilon_{j_{\kappa,1}} \dots \epsilon_{j_{\kappa, m_{\kappa}}} \frac{\partial}{\partial Y^{k_{\kappa}}} \right] \delta_{\text{D}}^{[0, k_{\max}]} \left(\mathbf{Y} - \mathcal{K}^{\{2\}} \epsilon \right) \\
&= \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \hat{\mathcal{N}}_{\mathcal{L}} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell}}{\ell!} \prod_{\kappa=1}^{\ell} \frac{\partial}{\partial Y^{k_{\kappa}}} \\
&\left\{ \sum_{m_{\kappa}=2}^{\infty} \mathcal{K}^{\{m_{\kappa}+1\}, k_{\kappa} j_{\kappa,1} \dots j_{\kappa, m_{\kappa}}} ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_{\kappa,1}} \dots ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_{\kappa, m_{\kappa}}} \exp \left[-\frac{1}{2} \mathbf{Y}^T \hat{\Sigma}^{-1} \mathbf{Y} \right] \right\} \\
&\equiv \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \hat{\mathcal{N}}_{\mathcal{L}} \exp \left[-\frac{1}{2} \mathbf{Y}^T \hat{\Sigma}^{-1} \mathbf{Y} \right] \left[1 + \sum_{m=1}^{\infty} \frac{1}{m!} \mathcal{B}^{\{m\}, i_1 \dots i_m} Y_{i_1} \dots Y_{i_m} \right].
\end{aligned} \tag{3.14}$$

Notice that contributions with m_{κ} involve the kernels $\mathcal{K}^{\{m_{\kappa}+1\}}$. In particular, the corrections to the Gaussian likelihood again start at $\mathcal{K}^{\{3\}}$, as expected. In the last line, we defined the coefficients⁸

$$\mathcal{B}^{\{m\}, i_1 \dots i_m} \equiv \mathcal{B}^{\{m\}, i_1 \dots i_m} [\{b_O^{\{\geq 1\}}\}, \delta_{\text{in}}] \tag{3.15}$$

as the sum of all the terms containing m powers of Y in this expansion, which are derived in App. D. The term linear in Y ($m = 1$) can be absorbed by shifting the mean μ inside Y [see Eq. (D.5)], and correspondingly the $m = 2$ contribution can be absorbed by shifting $\hat{\Sigma}$. These are examples of higher-order stochastic terms shifting lower-order contributions, a point to which we return below.

3.3 Discussion

The correspondence of Eq. (3.14) with Eq. (2.26), the main result of the partition function expansion in the current $\mathbf{J}$ derived in Sec. 2 is now clear, which we repeat for convenience:

$$\mathcal{P}'[\hat{\delta}_{g,k_{\max}}] = \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \mathcal{N}_{\mathcal{L}} \exp \left[-\frac{1}{2} \mathbf{Y}^T \Sigma^{-1} \mathbf{Y} \right] \left[1 + \sum_{m=3}^{\infty} \frac{1}{m!} \tilde{\mathcal{C}}^{\{m\}, i_1 \dots i_m} Y_{i_1} \dots Y_{i_m} \right]. \tag{3.16}$$

In particular, at leading order, there is a unique mapping between the $\mathcal{B}^{\{m\}}$ in Eq. (3.14) and the kernels $\mathcal{K}^{\{m\}}$:

$$\mathcal{B}^{\{m\}, j_1 \dots j_m} \supset m! \mathcal{K}^{\{m\}, k j'_1 \dots j'_{m-1}} [\{b_O^{\{m-1\}}\}, \delta_{\text{in}}] ([\mathcal{K}^{\{2\}}]^{-1})_{j'_1 j_1} \dots ([\mathcal{K}^{\{2\}}]^{-1})_{j'_{m-1} j_{m-1}} (\hat{\Sigma}^{-1})_{k j_m},$$

which is derived in App. D. As we have shown in Sec. 2.4 and App. C, the $\mathcal{K}^{\{m\}}$ can in turn be related unambiguously to the $\tilde{\mathcal{C}}^{\{m\}}$ in Eq. (3.16). As in the latter case, there are higher-order corrections. In particular, $\mathcal{B}^{\{m\}}$ also receives contributions from $\mathcal{K}^{\{m+2\}}$ [see Eq. (D.3)], and from products of kernels $\mathcal{K}$ (in fact, each ℓ yields precisely ℓ factors of $\mathcal{K}$, and

⁸Note that the $\mathcal{B}^{\{m\}}$ have the same dimensions as the $\tilde{\mathcal{C}}^{\{m\}}$.

the above relation was derived at $\ell = 1$). To be explicit, we give just the leading example ($m = 3$) here:

$$\mathcal{B}^{\{3\}, j_1 j_2 j_3} = \mathcal{K}^{\{3\}, k j'_1 j'_2}[\{b_O^{\{2\}}\}, \delta_{\text{in}}]([\mathcal{K}^{\{2\}}]^{-1})_{j'_1 j_1}([\mathcal{K}^{\{2\}}]^{-1})_{j'_2 j_2}([\mathcal{K}^{\{2\}}]^{-1})_{k j_3}. \quad (3.17)$$

The fact that higher-order contributions in the $b_O^{\{m\}}$ expansion correct lower-order terms is already clear from Eq. (3.1). Consider the effective large-scale shot-noise variance, quantified either by the coefficient of the term quadratic in Y ($m = 2$) in Eq. (3.14) or by computing the power spectrum from Eq. (3.1), which is given by

$$(b_{\perp}^{\{1\}})^2 + (b_{\delta}^{\{1\}})^2 \langle \delta^2 \rangle + \dots + 2(b_{\perp}^{\{2\}})^2 + \dots \quad (3.18)$$

in the noise-field formulation. In contrast, in the likelihood obtained from the partition function Eq. (2.1), this contribution is directly given by $C_{\perp}^{\{2\}}$.⁹ This mixing of contributions is a downside of the noise-field formulation, since this can create correlations between parameters that negatively affect the sampling efficiency. Notice that the same higher-order corrections appeared in the formal expansion of the general likelihood Eq. (3.16), via the additional higher-order contributions to the $\tilde{\mathcal{C}}^{\{m\}}$ (cf. the discussion in App. C). Thus, this feature appears generic to expanding the EFT field-level likelihood around the Gaussian approximation. Order by order, it can be remedied by a reparametrization, denoting the combination in Eq. (3.18) as $\propto P_{\epsilon, \text{eff}}$, with the parameter $P_{\epsilon, \text{eff}}$ replacing $b_{\perp}^{\{1\}}$, and similarly for the higher-order coefficients. We emphasize again that no new contributions are introduced by the higher-order corrections; they can all be absorbed by existing, lower-order $b_O^{\{m\}}$.

Recently, Ref. [7] similarly considered an expansion of the galaxy density including stochasticity as

$$\delta_g(\mathbf{k}) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \int_{\mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{p}'_1, \dots, \mathbf{p}'_m} \hat{\delta}_{\text{D}}(\mathbf{k} - \mathbf{p}_{1\dots n} - \mathbf{p}'_{1\dots m}) \mathcal{K}^{(n,m)}(\mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{p}'_1, \dots, \mathbf{p}'_m) \times \delta_{\text{in}}(\mathbf{p}_1) \cdots \delta_{\text{in}}(\mathbf{p}_n) \epsilon(\mathbf{p}'_1) \cdots \epsilon(\mathbf{p}'_m), \quad (3.19)$$

for a generic $\mathcal{K}^{(n,m)}$ kernel, and integrated over ϵ including terms up to order ϵ^2 ($m \leq 2$), in a very similar way as done here, to obtain non-Gaussian corrections to the likelihood. Note that Eq. (3.19) likewise only involves a single stochastic field for a single tracer (the multi-tracer generalization is also given there). In our case, the kernels $\mathcal{K}^{(n,m)}$ are explicitly defined in terms of the bias operator kernels as

$$\mathcal{K}^{(n,m)}(\mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{p}'_1, \dots, \mathbf{p}'_m) \rightarrow \sum_{O: n(O) \leq n} b_O^{\{m\}} K_O^{(n)}(\mathbf{p}_1, \dots, \mathbf{p}_n). \quad (3.20)$$

Crucially, we showed that this set of kernels fully reproduces the partition function Eq. (2.1).

At this point, we should emphasize that we have performed an expansion of the field-level likelihood in the noise-field formulation, Eq. (3.6), in order to connect it to the similar expansion of the likelihood derived from the partition function, Eq. (3.16). Crucially, unlike the latter case, Eq. (3.6) provides an explicit, normalized probability distribution at arbitrary order in m , i.e. fully incorporating non-Gaussian stochasticity.

⁹We ignore loop corrections resulting from the integration over δ_{in} such as the $P^{(22)}$ contribution here, focusing solely on the stochastic contributions at a fixed cutoff.

Finally, it is worth pointing out another possible route to connecting the noise-field formulation and the general EFT partition function. We have focused so far on the connection at the level of the likelihood. Conversely, one can also obtain a partition function from the noise-field formulation, by integrating over the field ϵ . Eq. (3.5) implies that we can write a partition function of the form

$$Z_{\text{nf}}[\mathbf{J}] \equiv \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \int \mathcal{D}\epsilon \mathcal{P}[\epsilon] \exp \left(J_k \left[\delta_{g,\text{det}}^{-k} [\{b_O^{\{0\}}\}, \delta_{\text{in}}] + \mathcal{K}^{\{2\},kj} [\{b_O^{\{1\}}\}, \delta_{\text{in}}] \epsilon_j \right. \right. \quad (3.21) \\ \left. \left. + \sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\},kj_1 \dots j_m} [\{b_O^{\{m\}}\}, \delta_{\text{in}}] \epsilon_{j_1} \dots \epsilon_{j_m} \right] \right),$$

where the first line contains the deterministic term and the linear coupling to the Gaussian field ϵ , while the second line contains the nonlinear couplings. One can then expand the exponential in these coupling terms. Noting that $\mathcal{P}[\epsilon]$ is a Gaussian with diagonal (and fixed) covariance given by P_ϵ , we again obtain a Gaussian integral over ϵ , which leads exactly to the partition function up to order $\mathbf{J}^2$ [Eq. (2.1) or Eq. (2.8)], following the matching in Eq. (3.4) and Eq. (3.10). The higher powers in ϵ from the expansion of the last line correspondingly lead to the higher powers of $\mathbf{J}$ in the partition function.

To summarize, the noise-field approach, where the galaxy density field is described via Eq. (3.1), is equivalent to Eq. (2.1) both at the partition function and likelihood levels. However, the resulting likelihood is explicitly normalized, defined everywhere in the joint field space $(\delta_{\text{in}}, \epsilon)$, and allows for a robust numerical evaluation. We describe such an implementation and first results in the next section.

4 Numerical implementation and results

We mentioned in the last sections that the field-level formulation of the galaxy density Eq. (3.1) is well suited for practical inference applications. We now discuss the implementation in `LEFTfield` and first results. We show field-level inference results of σ_8 from dark matter halos in the rest frame, precisely the case considered in [26, 35], but now including both non-Gaussian stochasticity and the coupling between stochasticity and density perturbations, via the bias terms $b_\delta^{\{1\}}$ and $b_\perp^{\{2\}}$. Moreover, we include the leading higher-derivative stochastic correction via a term $b_{\nabla^2}^{\{1\}} \nabla^2 \epsilon$.

Field-level inference proceeds by numerically sampling from $\mathcal{P}[\{\theta\}, \{b_O^{\{m\}}\} \mid \hat{\delta}_{g,k_{\text{max}}}]$, the posterior, where we again denote the observed data by $\hat{\delta}_{g,k_{\text{max}}}$. The forward model described by Eqs. (3.1)–(3.2) corresponds to using Eq. (3.6) instead of Eq. (2.10):

$$\mathcal{P}[\{\theta\}, \{b_O^{\{m\}}\} \mid \hat{\delta}_{g,k_{\text{max}}}] \propto \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}} \mid \{\theta\}] \int \mathcal{D}\epsilon \mathcal{P}[\epsilon \mid c_\epsilon] \\ \times \delta_D^{[0,k_{\text{max}}]} \left[\hat{\delta}_g(\mathbf{k}) - \sum_{m=0}^{\infty} \sum_{\mathbb{1}, O} b_O^{\{m\}} (\epsilon^m O)(\mathbf{k}) \right] \\ \times \mathcal{P}(\{b_O^{\{m\}}\}, \{\theta\}). \quad (4.1)$$

Notice that the likelihood is replaced by a Dirac delta, since all stochastic contributions are explicitly accounted for by the terms involving ϵ . The Dirac delta however only considers the

data up to the momentum cut $k_{\max}$, as indicated by the superscript. We employ a Gaussian prior with diagonal covariance on the grid where ϵ is discretized (see below),

$$\mathcal{P}(\epsilon|c_\epsilon) \propto \prod_{\mathbf{x}} \mathcal{N}(\epsilon(\mathbf{x})|0, c_\epsilon^2), \quad (4.2)$$

which thus multiplies all stochastic contributions.¹⁰ We fix $b_{\perp}^{\{1\}} = 1$, as it is degenerate with c_ϵ (conversely, one could choose $c_\epsilon = 1$ fixed and leave $b_{\perp}^{\{1\}}$ free). The Dirac likelihood can only be approximated numerically, and in the actual implementation we replace it with a Gaussian with fixed variance σ_0^2 :

$$\delta_{\text{D}}^{[0, k_{\max}]}[X(\mathbf{k})] \longrightarrow \exp \left[-\frac{1}{2} \sum_{\mathbf{k} \neq 0}^{k_{\max}} \left(\frac{|X(\mathbf{k})|^2}{\sigma_0^2} + \sigma_0^2 \right) \right], \quad (4.3)$$

which asymptotes to the desired Dirac distribution in the limit $\sigma_0 \rightarrow 0$. The choice of σ_0 corresponds to a tradeoff between accuracy of the likelihood approximation (smaller σ_0) and numerical efficiency (larger σ_0). We have tested that the precise value of σ_0 has limited significance if it is much smaller than the physical noise contribution described by c_ϵ . In practice, we choose σ_0^2 to be $\lesssim 0.25$ of the noise variance given by c_ϵ^2 . The Gaussian likelihood Eq. (4.3) further allows for analytical marginalization over all $b_O^{\{m\}}$ (cf. [36]). We employ this marginalization for all bias coefficients, listed here for completeness (see [26] for the definitions)

$$\left\{ b_O^{\{0\}} : O \in [\delta, \delta^2, K^2, \delta^3, K^3, \delta K^2, O_{\text{td}}, \nabla^2 \delta] \right\}, \quad b_{\nabla^2}^{\{1\}}, \quad b_{\delta}^{\{1\}}, \quad b_{\perp}^{\{2\}}, \quad (4.4)$$

except for $b_{\delta}^{\{0\}} \equiv b_1$ and $b_{\perp}^{\{1\}}$, the latter of which is fixed.¹¹ The deterministic bias expansion is chosen to match that of [26], so that only the stochastic part of the model changes. Here, we decide to keep the leading stochastic terms that appear in the galaxy power spectrum and bispectrum, as well as the subleading stochastic term in the power spectrum ($b_{\nabla^2}^{\{1\}}$). The priors on σ_8 and the bias coefficients are chosen as in [26], with wide priors for the additional stochastic coefficients,

$$\begin{aligned} \mathcal{P}(c_\epsilon) &= \mathcal{U}(0.05, 0.5); & \mathcal{P}(b_{\nabla^2}^{\{1\}} [(h^{-1} \text{Mpc})^2]) &= \mathcal{N}(0, 5^2); \\ \mathcal{P}(b_{\delta}^{\{1\}}) &= \mathcal{N}(0, (0.5)^2); & \mathcal{P}(b_{\perp}^{\{2\}}) &= \mathcal{N}(0, (0.2)^2). \end{aligned} \quad (4.5)$$

To sample from the posterior Eq. (4.1), we employ joint Hamiltonian Monte Carlo (HMC) sampling of the two fields $\{\hat{s}, \epsilon\}$, which both have unit Gaussian priors, where $\hat{s}$ is related to δ_{in} via

$$\delta_{\text{in}}(\mathbf{k}) \propto \sqrt{P_{\text{L}}(k)} \hat{s}(\mathbf{k}), \quad (4.6)$$

and the normalization depends on the grid size (see e.g. [26]). Both $\hat{s}$ and ϵ are filtered at a scale $\Lambda > k_{\max}$ and represented on grids that are sized appropriately to have Nyquist frequency just above Λ . Here, we choose $\Lambda = 0.14 h \text{Mpc}^{-1}$ and $k_{\max} = 0.12 h \text{Mpc}^{-1}$, the higher cutoff values considered in [26]. We employ a block-diagonal mass matrix, consisting

¹⁰The parameter c_ϵ is related to the power spectrum P_ϵ introduced in the previous section via $P_\epsilon = (L/N_g)^3 c_\epsilon^2$, where L is the box and N_g the grid size.

¹¹Elsewhere, the notation $b_{\epsilon\delta} \equiv b_{\delta}^{\{1\}}$, $b_{\epsilon^2} \equiv b_{\perp}^{\{2\}}$, $b_{\nabla^2\epsilon} \equiv b_{\nabla^2}^{\{1\}}$ is also used.

of dense 2×2 blocks for each pair $\{\hat{s}(\mathbf{k}), \epsilon(\mathbf{k})\}_{\mathbf{k}}$, with components derived from a linear forward model, the only case where the posterior can be computed analytically [20]. The HMC sampling steps are interleaved with slice sampling steps for the parameters $\{b_{\delta}^{\{0\}}, c_{\epsilon}, \sigma_8\}$ in a block-sampling fashion.

Given a proposal for $\{\hat{s}(\mathbf{k}), \epsilon(\mathbf{k})\}_{\mathbf{k}}, b_{\delta}^{\{0\}}, c_{\epsilon}, \sigma_8\}$, the forward model and field-level likelihood evaluation proceed as follows:

1. δ_{in} is computed via Eq. (4.6), and the grids for $\delta_{\text{in}}, \epsilon$ are zero-padded to avoid aliasing to modes below k_{max} in the computation of the nonlinear operators.
2. Lagrangian perturbation theory and bias expansions are performed to construct the Eulerian operators O . The details can be found in [37]. In the present case, we employ second-order LPT (2LPT) and a third-order Eulerian bias expansion. This was the matter/bias model employed in [26].
3. In parallel, the fields ϵ^m are constructed (here we restrict to $m \leq 2$) and the $\epsilon(\mathbf{x})O(\mathbf{x})$ are constructed in Eulerian space. Zero-padding is again performed as necessary to avoid aliasing to modes below k_{max} .
4. Finally, the likelihood Eq. (4.3) is evaluated, where σ_0 is fixed, i.e. not varied in the inference.

Conceptually, this sampling approach is straightforward, can easily be extended to any desired order in the expansion Eq. (3.1), and also allows for the accurate incorporation of other physical effects that are beyond the scope of this paper, such as redshift-space distortions [38, 39]. The major drawback is the need to sample two correlated fields, and thus not only doubling the dimensionality of the inference problem, but also adding significant correlations. In the present case, we have $2 \times 90^3 \approx 1.5$ million free parameters. The increased dimensionality and correlations lead to less efficient exploration of the posterior space, i.e. longer correlations between samples.

We first consider the parameter $\alpha \equiv \sigma_8/\sigma_{8,\text{fid}}$. Fig. 1 shows parameter traces from four independent sampling chains (left panel) and the autocorrelation function averaged over the four samples after removing burn-in (right panel). Estimating the correlation length τ as the first crossing of $\rho(\tau) = 0.1$, we obtain $\tau \simeq 32,600$, with a range of $26,900 - 40,300$ estimated from cross-chain variance (reducing the threshold further from 0.1 does not affect τ significantly, as can be gleaned from the figure). Clearly, a very large Monte Carlo sample size is required to obtain converged statistics (in this case ~ 2 million after removing 400,000 burn-in samples in total). The total effective sample size is estimated to be 61 (range 49–74), and the Gelman-Rubin statistic is $R(\alpha) = 1.06$. Given this effective sample size, it is justified to report the mean and 68% CL error bar for α :

$$\alpha = 0.984 \pm 0.035. \quad (4.7)$$

This result corresponds to the first field-level cosmology inference that uses the proper EFT-based, non-Gaussian noise model. The error bar is very mildly increased over that reported for Gaussian noise in [26], who obtained $\alpha = 1.013 \pm 0.033$, while the posterior mean is consistent with the latter within $< 1\sigma$. It is worth emphasizing that this is a high number density, i.e. low-noise, halo sample with $\bar{n} \simeq 1.3 \cdot 10^{-3} (h^{-1}\text{Mpc})^{-3}$. This explains why the detailed noise model does not affect the inference of the power spectrum amplitude σ_8

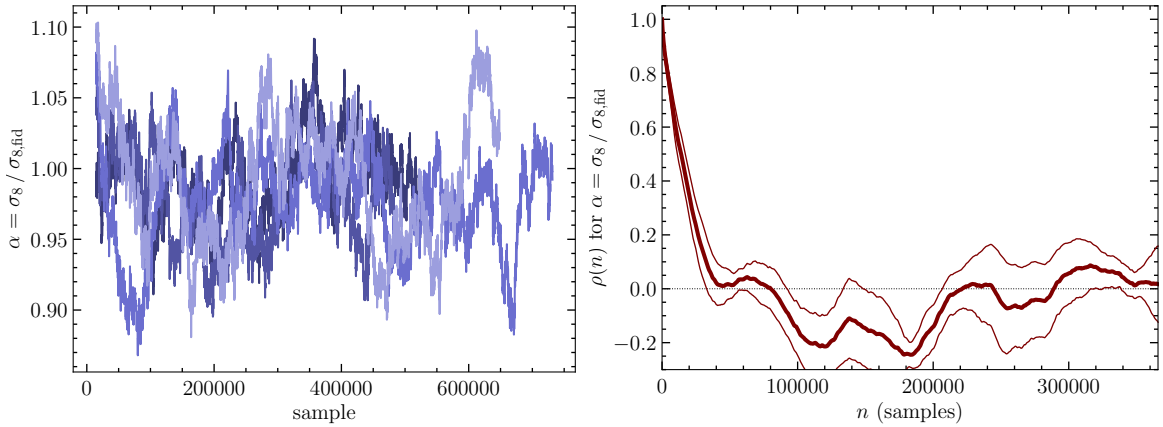


Figure 1: Trace plot (left panel) and normalized autocorrelation function $\rho(n) = \xi(n)/\xi(0)$ for the parameter $\alpha \equiv \sigma_8/\sigma_{8,\text{fid}}$ in four independent FLI chains for the “SNG” rest-frame dark matter halo sample from [26]. The autocorrelation for each chain is estimated after removing a burn-in phase of 100,000 samples. The thick line in the right panel shows the mean autocorrelation across the four chains, while the thin lines indicate the error on the mean estimated from the sample variance across correlation functions.

significantly.¹² While this finding could still be considered preliminary given the limited effective sample size (the results in [26] were based on an effective sample size greater than 100), it indicates that a non-Gaussian noise model does not substantially affect field-level inference error bars on σ_8 , at least for high-number-density samples.

We now turn to the inference of the noise amplitude parameter c_ϵ . Ref. [41], and more recently [26, 40], pointed out that the noise amplitude drifts to unphysically small values in FLI inferences applied to nonlinear tracers such as halos [26, 41], HOD-based catalogs [35] or mock data with non-Gaussian noise [40], when assuming Gaussian stochasticity in the inference. The term “sigma collapse” was coined for this phenomenon [41], given that the variance in the Gaussian field-level likelihood is usually denoted as σ^2 . We can test for this phenomenon in the noise-field formulation as well, by performing an inference where $b_\delta^{\{1\}}$ and $b_\perp^{\{2\}}$ are fixed to zero. The resulting parameter traces for c_ϵ are shown as maroon lines in the left panel of Fig. 2, which clearly reproduce “sigma collapse,” i.e. the drift of the noise amplitude c_ϵ to small values (0.05 being the arbitrary lower bound imposed on the parameter here).

The blue lines in the left panel of Fig. 2 instead show the traces for the same chains as in Fig. 1, i.e. allowing for density-dependent and non-Gaussian stochasticity. Evidently, “sigma collapse” does not occur in FLI chains that consistently incorporate non-Gaussian stochasticity. Instead, c_ϵ remains at a physically expected level; the pure Poisson-noise expectation, $P_\epsilon = 1/\bar{n}$, for this halo catalog corresponds to $c_\epsilon^{\text{Poisson}} = 0.15$. Note that one expects a somewhat larger effective noise from integrating out modes above the cutoff Λ for a nonlinearly biased tracer [4, 20].

Nevertheless, sampling c_ϵ remains challenging due to the long correlations within samples. A residual dependence on the starting value of c_ϵ is still clearly visible in the parameter

¹²Note that the recent Ref. [40] considered mock datasets with much higher noise amplitude.

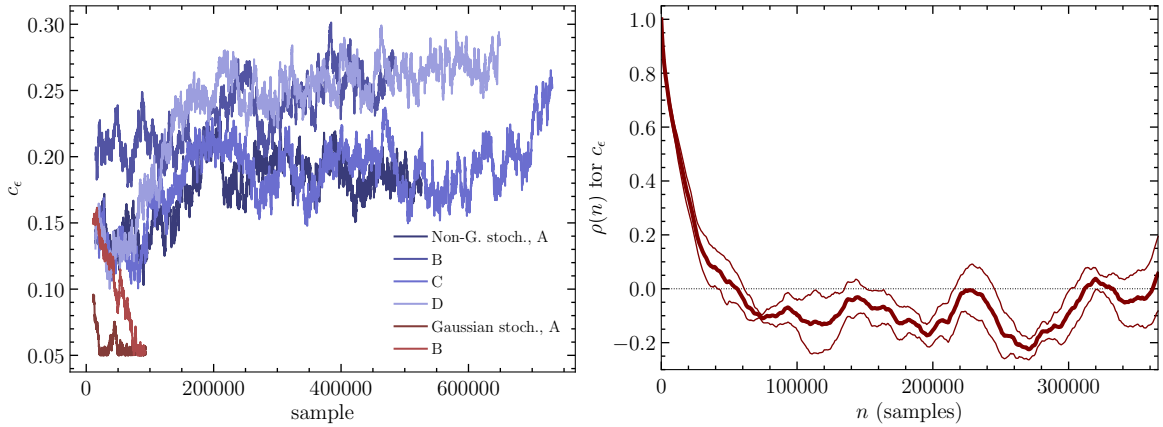


Figure 2: Trace plot (left panel) and normalized autocorrelation function $\rho(n) = \xi(n)/\xi(0)$ for the noise-amplitude parameter c_ϵ for the same chains as in Fig. 1 (blue shades, labeled A–D). In the trace plot we also show the results of two chains using the same noise-field formulation but with non-Gaussian and density-dependent noise turned off ($b_\delta^{\{1\}} = 0 = b_\perp^{\{2\}}$; maroon shades). These latter chains show that c_ϵ drifts to its lower limit 0.05, a trend previously found for Gaussian likelihoods in [26, 40, 41].

traces. Hence, we do not quote any posterior mean or error bar here. While the overdispersion in c_ϵ fortunately does not affect the posterior for α strongly (as evidenced by the Gelman-Rubin statistic for α reported above), proper robust cosmology posteriors require converged posteriors in all parameters. Future work must thus aim at improving the sampling efficiency. This could involve an improved mass matrix, different marginalization schemes, and/or joint HMC sampling of parameters and fields.

5 Conclusions

In this paper, we have built on the general EFT partition function for galaxy clustering from [4], Eq. (2.1), to investigate how stochastic contributions to galaxy clustering can be described at the field level. Our results are at two levels. First, we derive the general expression for the field-level likelihood in the EFT, which is given by the functional Fourier transform of the partition function (Sec. 2.2). This Fourier transform cannot be computed in closed form. However, one can expand around the Gaussian limit of stochasticity (Sec. 2.4), which is a valid expansion within the perturbative EFT context. Second, we establish that the reduced model with a single Gaussian field, Eq. (1.3), is a sufficient description of galaxy stochasticity within the EFT, by deriving the same likelihood in this formulation (Sec. 3). This description is significantly more restrictive than the non-minimal model in Eq. (1.2) which has so far been assumed as the standard. App. A compares the two approaches at the n -point function level.

The formal likelihood obtained from the EFT partition function has two major drawbacks: it is not normalized, showing unphysical behavior in the tails (i.e. negative probability densities), as is common for Edgeworth-like expansions of probability distributions around a Gaussian; and it involves field-level matrix inversions which are intractable in practice (see also App. B). Instead, the *noise-field formulation* of the EFT likelihood presented in Sec. 3

avoids unphysical probabilities and is suited for numerical sampling, while at the same time capturing the full EFT likelihood, order by order.

Finally, Sec. 4 presents first results of field-level cosmology inference using this noise-field formulation. The downside of this approach is that the dimensionality of parameters to be sampled is doubled, yielding a slower exploration of the parameter space. Nevertheless, we show converged results for σ_8 (Fig. 1) jointly inferred with bias and stochastic parameters from a dark matter halo sample in real space. This represents the first time that the full EFT model for galaxy bias and stochasticity has been employed in field-level inference. Moreover, the non-Gaussian noise formulation leads to a stable inferred noise amplitude at a physically expected value, and thus solves the problem of the “collapse” of the noise variance observed for the Gaussian-noise case (Fig. 2).

In the future, it will be imperative to explore more efficient sampling techniques, perhaps including more refined analytical likelihoods that correctly capture the stochastic mode couplings which the existing real-space likelihood unfortunately does not (App. B). It would also be interesting and important to generalize the results of our work to predict the stochastic contributions to the galaxy velocity field, which are necessary for redshift-space distortions.

Acknowledgments

The authors acknowledge Minh Nguyen and Zvonimir Vlah for useful conversations, and Francesco Conteddu and Mathias Garny for comments on the draft.

A Comparison of standard (non-minimal) and minimal stochastic contributions for n -point functions

We here compare the predictions for stochastic contributions to galaxy n -point functions predicted by Eq. (1.2) on the one hand, and Eq. (1.3) on the other. Consider a “Gaussian stochastic” contribution to the galaxy n -point correlation function in real space,

$$\langle \delta_g(\mathbf{x}_1) \cdots \delta_g(\mathbf{x}_n) \rangle, \quad (\text{A.1})$$

defined as involving precisely two powers of stochastic fields. We choose to work in real space here, in order to make the local products of operators simple to express. In the non-minimal formulation [Eq. (1.2)], such a contribution can be written as

$$\begin{aligned} & \langle \epsilon_O^{\text{n-min}}(\mathbf{x}_1) O(\mathbf{x}_1) \epsilon_{O'}^{\text{n-min}}(\mathbf{x}_2) O'(\mathbf{x}_2) O_3(\mathbf{x}_3) \cdots O_n(\mathbf{x}_n) \rangle \\ &= \langle \epsilon_O^{\text{n-min}}(\mathbf{x}_1) \epsilon_{O'}^{\text{n-min}}(\mathbf{x}_2) \rangle \langle O(\mathbf{x}_1) O'(\mathbf{x}_2) O_3(\mathbf{x}_3) \cdots O_n(\mathbf{x}_n) \rangle \\ &= P_{\epsilon_O \epsilon_{O'}} \delta_D(\mathbf{x}_1 - \mathbf{x}_2) \langle O(\mathbf{x}_1) O'(\mathbf{x}_1) O_3(\mathbf{x}_3) \cdots O_n(\mathbf{x}_n) \rangle, \end{aligned} \quad (\text{A.2})$$

where O, O' could be either the unit operator or a nontrivial bias operator, while $O_3, \dots, O_n$ stand for arbitrary bias operators (otherwise, we trivially reduce to a lower order n -point function). We have also introduced

$$\langle \epsilon_O^{\text{n-min}}(\mathbf{k}) \epsilon_{O'}^{\text{n-min}}(\mathbf{k}') \rangle' = P_{\epsilon_O \epsilon_{O'}}. \quad (\text{A.3})$$

A k^2 -dependence capturing higher-derivative stochastic terms following Eq. (1.12) can also be included here, corresponding to derivatives acting on the Dirac delta in Eq. (A.2).

On the other hand, Eq. (1.3) predicts¹³

$$b_O^{\{1\}} b_{O'}^{\{1\}} P_\epsilon \delta_D(\mathbf{x}_1 - \mathbf{x}_2) \langle O(\mathbf{x}_1) O'(\mathbf{x}_1) O_3(\mathbf{x}_3) \cdots O_n(\mathbf{x}_n) \rangle, \quad (\text{A.4})$$

where P_ϵ is introduced in Eq. (3.2). Clearly, the two expressions are equivalent in terms of structure. To understand the apparent reduction in coefficients in the two formulations, consider a pair of operators, such as $\mathbb{1}$ and O . This pair is characterized by three coefficients in the non-minimal formulation, $P_{\epsilon_{\mathbb{1}}\epsilon_{\mathbb{1}}}, P_{\epsilon_{\mathbb{1}}\epsilon_O}, P_{\epsilon_O\epsilon_O}$. Indeed, three coefficients are needed to describe the covariance of two fields. On other hand, the same pair is only controlled by two coefficients in Eq. (A.4), $b_{\mathbb{1}}^{\{1\}}, b_O^{\{1\}}$. However, the contribution to Eq. (A.2) that comes with $P_{\epsilon_O\epsilon_O}$ is precisely degenerate with that controlled by $P_{\epsilon_{\mathbb{1}}\epsilon_{OO}}$, which has an analogous coefficient $b_{OO}^{\{1\}}$ in Eq. (A.4). This degeneracy is due to (1) the locality of the stochastic process; and (2) the completeness of the bias expansion, which for any O_1, O_2 in the basis also includes $O_1 O_2$. In a practical analysis, one would thus eliminate either $P_{\epsilon_O\epsilon_O}$ or $P_{\epsilon_{\mathbb{1}}\epsilon_{OO}}$. Thanks to these facts, Eq. (A.4) can capture the full set of non-degenerate contributions in Eq. (A.2).

To illustrate how this reasoning continues to higher order, consider a contribution with three powers of stochastic fields. Eq. (1.2) yields

$$\begin{aligned} & \langle \epsilon_O^{\text{n-min}}(\mathbf{x}_1) O(\mathbf{x}_1) \epsilon_{O'}^{\text{n-min}}(\mathbf{x}_2) O'(\mathbf{x}_2) \epsilon_{O''}^{\text{n-min}}(\mathbf{x}_3) O''(\mathbf{x}_3) O_4(\mathbf{x}_4) \cdots O_n(\mathbf{x}_n) \rangle \\ &= \langle \epsilon_O^{\text{n-min}}(\mathbf{x}_1) \epsilon_{O'}^{\text{n-min}}(\mathbf{x}_2) \epsilon_{O''}^{\text{n-min}}(\mathbf{x}_3) \rangle \langle O(\mathbf{x}_1) O'(\mathbf{x}_2) O''(\mathbf{x}_3) O_4(\mathbf{x}_4) \cdots O_n(\mathbf{x}_n) \rangle \\ &= B_{\epsilon_O\epsilon_{O'}\epsilon_{O''}} \delta_D(\mathbf{x}_1 - \mathbf{x}_2) \delta_D(\mathbf{x}_2 - \mathbf{x}_3) \langle O(\mathbf{x}_1) O'(\mathbf{x}_1) O''(\mathbf{x}_1) O_4(\mathbf{x}_4) \cdots O_n(\mathbf{x}_n) \rangle, \end{aligned} \quad (\text{A.5})$$

where $B_{\epsilon_O\epsilon_{O'}\epsilon_{O''}}$ is defined analogously to Eq. (A.3), while Eq. (1.3) predicts

$$\begin{aligned} & 2 \left[b_O^{\{2\}} b_{O'}^{\{1\}} b_{O''}^{\{1\}} + 2 \text{ perm.} \right] (P_\epsilon)^2 \delta_D(\mathbf{x}_1 - \mathbf{x}_2) \delta_D(\mathbf{x}_2 - \mathbf{x}_3) \\ & \times \langle O(\mathbf{x}_1) O'(\mathbf{x}_1) O''(\mathbf{x}_1) O_4(\mathbf{x}_4) \cdots O_n(\mathbf{x}_n) \rangle. \end{aligned} \quad (\text{A.6})$$

Again, both have the same structure. Moreover, since $B_{\epsilon_O\epsilon_{O'}\epsilon_{O''}}$ is totally symmetric (as the configuration and scale dependence of $B_{\epsilon_O\epsilon_{O'}\epsilon_{O''}}$ is trivial, it is invariant under interchange of any of the fields ϵ_O), and its contribution is degenerate with that of $B_{\epsilon_{\mathbb{1}}\epsilon_{OO'}\epsilon_{O''}}$ as well as other corresponding terms, Eq. (A.6) has sufficient flexibility to capture the non-degenerate contributions to Eq. (A.5).

B Issues with the real-space formulation of the likelihood

The real-space likelihood formulation introduced in [28] in principle offers a neat way of incorporating the stochastic terms $C_O^{\{2\}}$ or equivalently $b_O^{\{1\}}$, i.e. the leading coupling between stochasticity and long-wavelength modes while keeping the stochasticity Gaussian. This is precisely the case studied in Sec. 2.3.1. The likelihood in Eq. (2.16) cannot be simply computed in Fourier space, as the covariance Σ is dense. However, in real space we can, at least naively, write Eq. (2.15) as¹⁴

$$\Sigma_{\mathbf{x}, \mathbf{y}} = \left[C_{\mathbb{1}}^{\text{r}, \{2\}} + \sum_O C_O^{\text{r}, \{2\}} O[\delta_{\text{in}}](\mathbf{x}) \right] \delta_D(\mathbf{x} - \mathbf{y}), \quad (\text{B.1})$$

¹³At leading order; as discussed in Sec. 3, these contributions are corrected by higher-order terms, which however have the same structure and can be absorbed in redefined $b_O^{\{1\}}$.

¹⁴The $C_{\mathbb{1}, O}^{\text{r}, \{2\}}$ appearing here are normalized differently than those in Eq. (2.16), since they refer to real-space fields and the reduced grid N_g .

since the coupling between stochasticity and the $O[\delta_{\text{in}}]$ is local in real space. This is a diagonal covariance matrix which can be trivially inverted in real space. However, one still needs to implement the sharp- k_{max} cut in the likelihood in Eq. (2.16). For this, Refs. [14, 28] proposed to perform a grid reduction in Fourier space, effectively restricting the components $i = 1, 2, 3$ of all represented wavenumbers to $|k_i| \leq k_{\text{Ny}} = N_g \pi / L$, where k_{Ny} is the Nyquist frequency, N_g the grid resolution, and L the size of the box. By appropriately choosing N_g , one can ensure that $|k_i| \leq k_{\text{max}}$. The real-space likelihood is then given by [14, 28]

$$\mathcal{L}_{\text{real}}(\hat{\delta}_g | \{b_O\}, \{C_O^{\text{r},\{2\}}\}, \delta_{\text{in}}) \propto \exp \left[-\frac{1}{2} \sum_{\mathbf{x}}^{N_g^3} \left(\frac{(\hat{\delta}_g(\mathbf{x}) - \delta_{g,\text{det}}[\delta_{\text{in}}, \{b_O\}](\mathbf{x}))^2}{\sigma^2(\mathbf{x})} + \sigma^2(\mathbf{x}) \right) \right],$$

with $\sigma^2(\mathbf{x}) \equiv \sigma^2[\delta_{\text{in}}, \{C_O^{\text{r},\{2\}}\}](\mathbf{x}) = C_{\mathbf{1}}^{\text{r},\{2\}} + \sum_O C_O^{\text{r},\{2\}} O[\delta_{\text{in}}](\mathbf{x}), \quad (\text{B.2})$

where we have explicitly indicated the grid size in the real-space sum, and both $\hat{\delta}_g$ and $\delta_{g,\text{det}}$ are reduced to this grid in Fourier space before evaluating the likelihood (in an actual implementation, one would also ensure that $\sigma^2(\mathbf{x})$ is positive definite, see [14]).

In order to understand the issues with this implementation, imagine generating a mock dataset, i.e. tracer field from the real-space likelihood. For simplicity, we consider a fixed deterministic prediction $\delta_{g,\text{det}}(\mathbf{x})$, since this field does not play a role in this discussion. Clearly, such a mock dataset will again be represented on a grid of size N_g , and can be written as

$$\hat{\delta}_{g,\text{real}}(\mathbf{x}) = \delta_{g,\text{det}}(\mathbf{x}) + \epsilon_{\text{real}}(\mathbf{x}) \quad \text{with}$$

$$\epsilon_{\text{real}}(\mathbf{x}) \sim \mathcal{N}(0, \sigma^2(\mathbf{x})) = \left[C_{\mathbf{1}}^{\text{r},\{2\}} + \sum_O C_O^{\text{r},\{2\}} O(\mathbf{x}) \right]^{1/2} \epsilon_{N_g}(\mathbf{x}), \quad (\text{B.3})$$

where $\epsilon_{N_g}(\mathbf{x}) \sim \mathcal{N}(0, 1)$ is a unit Gaussian random field generated on the grid N_g (and again, all of $\hat{\delta}_{g,\text{real}}, \delta_{g,\text{det}}, O$ are represented on the same grid).

The last line in Eq. (B.3) bears out the issue. Note first that both ϵ_{N_g} and O have Fourier-support up to the Nyquist frequency of the grid, $k_{\text{Ny}} = k_{\text{max}}$. This is necessary in order to have the correct mean-field prediction $\delta_{g,\text{det}}$ and leading noise contribution up to k_{max} . On the other hand, Eq. (B.3) multiplies the fields ϵ_{N_g} and O in real space, thus exciting modes up to $2k_{\text{Ny}}$ in Fourier space (it is sufficient to expand the square-root in Eq. (B.3) to linear order in perturbations to see this). These modes cannot be represented directly on the grid of size N_g , and are instead aliased to lower- k modes. In fact, *all* Fourier modes on the likelihood grid are polluted by aliasing. This aliasing is unphysical, since the grid size is related to k_{max} , a scale which has no physical significance for the tracer or forward model. On the other hand, increasing the likelihood grid size while maintaining the k_{max} cut is not possible.

It thus appears impossible to represent the correct mode-coupling structure of the stochastic terms $\propto C_O^{\{2\}}$ in Eq. (2.16) in a closed-form real-space likelihood, while at the same time having a sharp filter on the data at some $k_{\text{max}} < \Lambda$.

C Derivation of coefficients $\mathcal{C}$ and $\tilde{\mathcal{C}}$

In the following derivations, we do not write dependencies of quantities on $C_O^{\{m\}}$ and δ_{in} for clarity, indicating the relevant dependencies at the end of each derivation. We will also drop

the explicit filtering of J at $k_{\max}$, as we always only encounter the $k_{\max}$ -filtered current.

Coefficients $\mathcal{C}$ in Eq. (2.22). We start with

$$\begin{aligned} & \exp \left[\sum_{m=3}^{\infty} \frac{1}{m!} \mathcal{K}^{\{m\}, i_1 \dots i_m} J_{i_1} \dots J_{i_m} \right] \\ &= \sum_{N=0}^{\infty} \frac{1}{N!} \left[\sum_{m=3}^{\infty} \frac{1}{m!} \mathcal{K}^{\{m\}, i_1 \dots i_m} J_{i_1} \dots J_{i_m} \right]^N. \end{aligned} \quad (\text{C.1})$$

In order to apply the multinomial formula, we restrict the sum over m to M ; we will specify M once we reorder the sum below. Further, $N = 0$ just yields the trivial 1 which we have pulled out in Eq. (2.22), so we start with $N = 1$. We have

$$\begin{aligned} & \sum_{N=1}^{\infty} \frac{1}{N!} \left[\sum_{m=3}^M \frac{1}{m!} \mathcal{K}^{\{m\}, i_1 \dots i_m} J_{i_1} \dots J_{i_m} \right]^N \\ &= \sum_{N=1}^{\infty} \sum_{\substack{a_3, \dots, a_M \geq 0 \\ a_3 + \dots + a_M = N}} \frac{1}{a_3! \dots a_M!} \prod_{m=3}^M \left[\frac{1}{m!} \mathcal{K}^{\{m\}, i_1 \dots i_m} J_{i_1} \dots J_{i_m} \right]^{a_m}. \end{aligned} \quad (\text{C.2})$$

Each term here has

$$N_J = \sum_{m=3}^M m a_m =: L \quad (\text{C.3})$$

powers of the current. We would now like to reorder the sum into an expansion in powers of the current with coefficients $\mathcal{C}$,

$$\exp \left[\sum_{m=3}^{\infty} \frac{1}{m!} \mathcal{K}^{\{m\}, i_1 \dots i_m} J_{i_1} \dots J_{i_m} \right] = \sum_{L=0}^{\infty} \frac{1}{L!} \mathcal{C}^{\{L\}, i_1 \dots i_L} J_{i_1} \dots J_{i_L}. \quad (\text{C.4})$$

Abbreviating the indices (these can easily be restored, but become clumsy to write), we have for the coefficients

$$\begin{aligned} \mathcal{C}^{\{L\}, i_1 \dots i_L} &= \sum_{N=1}^{\infty} \sum_{\substack{a_3, \dots, a_M \geq 0 \\ a_3 + \dots + a_M = N \\ 3a_3 + \dots + M a_M = L}} \frac{L!}{a_3! \dots a_M!} \prod_{m=3}^M \left(\frac{1}{m!} \right)^{a_m} (\mathcal{K}^{\{m\}, \dots})^{a_m} \\ &= \sum_{\substack{a_3, \dots, a_L \geq 0 \\ 3a_3 + \dots + L a_L = L}} \frac{L!}{a_3! \dots a_L!} \prod_{m=3}^L \left(\frac{1}{m!} \right)^{a_m} (\mathcal{K}^{\{m\}, \dots})^{a_m}, \end{aligned} \quad (\text{C.5})$$

where $L \geq 3$ and we have used that the highest M we need to consider is $M = L$, since the term with $a_L = 1$ and $a_{m \neq L} = 0$ already has the entire number of L currents. This is Eq. (2.23).

A few examples are given by

$$\begin{aligned} L = 3 : \quad & \mathcal{C}^{\{3\}, i_1 i_2 i_3} [\{C_O^{\{3\}}\}, \delta_{\text{in}}] = \mathcal{K}^{\{3\}, i_1 i_2 i_3} [\{C_O^{\{3\}}\}, \delta_{\text{in}}] \\ L = 4 : \quad & \mathcal{C}^{\{4\}, i_1 i_2 i_3 i_4} [\{C_O^{\{4\}}\}, \delta_{\text{in}}] = \mathcal{K}^{\{4\}, i_1 i_2 i_3 i_4} [\{C_O^{\{4\}}\}, \delta_{\text{in}}] \\ L = 6 : \quad & \mathcal{C}^{\{6\}, i_1 \dots i_6} [\{C_O^{\{3\}}\}, \{C_O^{\{6\}}\}, \delta_{\text{in}}] = \mathcal{K}^{\{6\}, i_1 \dots i_6} [\{C_O^{\{3\}}\}, \delta_{\text{in}}] \\ & + \frac{6!}{(3!)^2} \mathcal{K}^{\{3\}, i_1 i_2 i_3} [\{C_O^{\{3\}}\}, \delta_{\text{in}}] \mathcal{K}^{\{3\}, i_4 i_5 i_6} [\{C_O^{\{3\}}\}, \delta_{\text{in}}]. \end{aligned} \quad (\text{C.6})$$

In the following, we will also use the fully symmetrized version of the $\mathcal{C}^{\{L\}}$,

$$\mathcal{C}_{\text{symm}}^{\{L\}, i_1 \dots i_L} \equiv \frac{1}{L!} \sum_{\sigma \in S_L} \mathcal{C}^{\{L\}, \sigma(i_1) \dots \sigma(i_L)}, \quad (\text{C.7})$$

where S_L is the group of permutations of L elements (with cardinality $L!$). $\mathcal{C}^{\{3\}}$ and $\mathcal{C}^{\{4\}}$ are already symmetric thanks to the symmetry of the kernels $\mathcal{K}$, while for $L = 6$, for example, we need to symmetrize by summing over the partitions of 6 indices into two groups of 3.

Coefficients $\tilde{\mathcal{C}}$ in Eq. (2.26). We start from Eq. (2.25),

$$\begin{aligned} \mathcal{P}'[\hat{\delta}_{g, k_{\max}}] &= (Z[\mathbf{J} = \mathbf{0}])^{-1} \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[-\frac{1}{2} \mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \right] \int \mathcal{D}(\mathbf{i} \tilde{\mathbf{J}}) \exp \left[\frac{1}{2} \tilde{\mathbf{J}}^T \boldsymbol{\Sigma} \tilde{\mathbf{J}} \right] \\ &\times \left[1 + \sum_{L=3}^{\infty} \frac{1}{L!} \mathcal{C}^{\{L\}, i_1 \dots i_L} (\tilde{\mathbf{J}} + \boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_1} \dots (\tilde{\mathbf{J}} + \boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_L} \right]. \end{aligned} \quad (\text{C.8})$$

The main goal is to find $\tilde{\mathcal{C}}$, such that we can write it in the form of Eq. (2.26), which we repeat here for convenience:

$$\mathcal{P}'[\hat{\delta}_{g, k_{\max}}] = \mathcal{N}_{\mathcal{L}} \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[-\frac{1}{2} \mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \right] \left[1 + \sum_{m=3}^{\infty} \frac{1}{m!} \tilde{\mathcal{C}}^{\{m\}, i_1 \dots i_m} Y_{i_1} \dots Y_{i_m} \right]. \quad (\text{C.9})$$

We first isolate the integral over $\tilde{\mathbf{J}}$ for a single L . It is highly useful to employ the symmetrized kernels $\mathcal{C}_{\text{symm}}^{\{L\}}$, as we can then write the $\tilde{\mathbf{J}}$ integrals in Eq. (C.8) as

$$\int \mathcal{D}(\mathbf{i} \tilde{\mathbf{J}}) \exp \left[\frac{1}{2} \tilde{\mathbf{J}}^T \boldsymbol{\Sigma} \tilde{\mathbf{J}} \right] \frac{1}{L!} \mathcal{C}_{\text{symm}}^{\{L\}, i_1 \dots i_L} \sum_{\ell=0}^L \binom{L}{\ell} \tilde{J}_{i_1} \dots \tilde{J}_{i_{\ell}} (\boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_{\ell+1}} \dots (\boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_L}.$$

Now we perform the Gaussian integral over $\tilde{\mathbf{J}}$, and drop a common normalizing constant that is parameter independent, yielding

$$\begin{aligned} &|\boldsymbol{\Sigma}|^{-1/2} \frac{1}{L!} \mathcal{C}_{\text{symm}}^{\{L\}, i_1 \dots i_L} \sum_{\ell=0, \ell \text{ even}}^L \binom{L}{\ell} \frac{1}{2^{\ell/2} (\ell/2)!} \sum_{\sigma \in S_{\ell}} (\boldsymbol{\Sigma}^{-1})_{\sigma(i_1)\sigma(i_2)} \dots (\boldsymbol{\Sigma}^{-1})_{\sigma(i_{\ell-1})\sigma(i_{\ell})} \\ &\times (\boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_{\ell+1}} \dots (\boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_L}. \end{aligned} \quad (\text{C.10})$$

Given the symmetry of $\mathcal{C}_{\text{symm}}^{\{L\}}$ in the L indices, all permutations lead to the same result, so we can cancel the factor of $(\ell/2)!$ to obtain, again for a fixed L ,

$$\begin{aligned} &|\boldsymbol{\Sigma}|^{-1/2} \frac{1}{L!} \sum_{\ell=0, \ell \text{ even}}^L \frac{L!}{2^{\ell/2} \ell! (L-\ell)!} \mathcal{C}_{\text{symm}}^{\{L\}, i_1 \dots i_L} (\boldsymbol{\Sigma}^{-1})_{i_1 i_2} \dots (\boldsymbol{\Sigma}^{-1})_{i_{\ell-1} i_{\ell}} \\ &\times (\boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_{\ell+1}} \dots (\boldsymbol{\Sigma}^{-1} \mathbf{Y})_{i_L}. \end{aligned} \quad (\text{C.11})$$

Noting that $|\boldsymbol{\Sigma}|^{-1/2}$ is factored out into $\mathcal{N}_{\mathcal{L}}$, and collecting all terms with m powers of Y , and again using symmetry of the $\mathcal{C}_{\text{symm}}^{\{L\}}$, we obtain for the coefficient

$$\begin{aligned} \tilde{\mathcal{C}}^{\{m\}, i_1 \dots i_m} &= \sum_{\substack{L=m \\ L-m \text{ even}}}^{\infty} \frac{L!}{2^{(L-m)/2} (L-m)!} \mathcal{C}_{\text{symm}}^{\{L\}, j_1 \dots j_L} (\boldsymbol{\Sigma}^{-1})_{j_{m+1} j_{m+2}} \dots (\boldsymbol{\Sigma}^{-1})_{j_{L-1} j_L} \\ &\times (\boldsymbol{\Sigma}^{-1})_{j_1 i_1} \dots (\boldsymbol{\Sigma}^{-1})_{j_m i_m}, \end{aligned} \quad (\text{C.12})$$

where recall that $\mathcal{C}_{\text{symm}}^{\{L\}} = \mathcal{C}_{\text{symm}}^{\{L\}}[\{C^{\geq 3}\}, \delta_{\text{in}}]$ while $\Sigma = \Sigma[\{C_O^{\{2\}}\}, \delta_{\text{in}}]$, so that $\tilde{\mathcal{C}}^{\{m\}} = \tilde{\mathcal{C}}^{\{m\}}[\{C^{\geq 2}\}, \delta_{\text{in}}]$ (and $\mathbf{Y} = \mathbf{Y}[\hat{\delta}_{g, k_{\text{max}}}, \{b_O\}, \delta_{\text{in}}]$ as always).

At a given m , we can distinguish leading and higher-order contributions. The leading order contribution is

$$\tilde{\mathcal{C}}^{\{m\}, i_1 \dots i_m} \stackrel{\text{LO}}{=} \mathcal{C}_{\text{symm}}^{\{m\}, j_1 \dots j_m} (\Sigma^{-1})_{j_1 i_1} \dots (\Sigma^{-1})_{j_m i_m}, \quad (\text{C.13})$$

i.e. it is given by $\mathcal{C}_{\text{symm}}^{\{m\}}$ contracted with m instances of Σ^{-1} . The next higher-order contribution is

$$\tilde{\mathcal{C}}^{\{m\}, i_1 \dots i_m} \stackrel{\text{NLO}}{=} \frac{(m+2)(m+1)}{2} \mathcal{C}_{\text{symm}}^{\{m+2\}, j_1 \dots j_{m+2}} (\Sigma^{-1})_{j_{m+1} j_{m+2}} (\Sigma^{-1})_{j_1 i_1} \dots (\Sigma^{-1})_{j_m i_m}, \quad (\text{C.14})$$

consisting of $\mathcal{C}_{\text{symm}}^{\{m+2\}}$ with two indices contracted with Σ^{-1} . This similarly continues to higher order.

D Expansion of likelihood in the noise-field formulation

Here we provide more explicit results and discussions on the posterior expression in Eq. (3.14). Consider first the contribution from $\ell = 1$ to the second equality, dropping the normalization $\hat{\mathcal{N}}_{\mathcal{L}}$ for convenience:

$$\begin{aligned} \mathcal{P}'[\hat{\delta}_{g, k_{\text{max}}}] \Big|_{\ell=1} &\propto - \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \frac{\partial}{\partial Y^k} \left\{ \sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\}, k j_1 \dots j_m} ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_1} \dots ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_m} \right. \\ &\quad \left. \times \exp \left[-\frac{1}{2} \mathbf{Y}^T \hat{\Sigma}^{-1} \mathbf{Y} \right] \right\} \\ &= - \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \sum_{m=2}^{\infty} \mathcal{K}^{\{m+1\}, k j_1 \dots j_m} \left[m ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_1} \dots ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_{m-1}} ([\mathcal{K}^{\{2\}}]^{-1})_{j_m k} \right. \\ &\quad \left. - ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_1} \dots ([\mathcal{K}^{\{2\}}]^{-1} \mathbf{Y})_{j_m} (\hat{\Sigma}^{-1} \mathbf{Y})_k \right] \exp \left[-\frac{1}{2} \mathbf{Y}^T \hat{\Sigma}^{-1} \mathbf{Y} \right]. \end{aligned} \quad (\text{D.1})$$

Here we have used the symmetry of $\mathcal{K}^{\{m+1\}, j_1 \dots j_{m+1}}$. Comparing the contribution at $m+1$ with the definition of $\mathcal{B}$,

$$\mathcal{P}'[\hat{\delta}_{g, k_{\text{max}}}] \propto \int \mathcal{D}\delta_{\text{in}} \mathcal{P}[\delta_{\text{in}}] \exp \left[-\frac{1}{2} \mathbf{Y}^T \hat{\Sigma}^{-1} \mathbf{Y} \right] \left[1 + \sum_{m=1}^{\infty} \frac{1}{m!} \mathcal{B}^{\{m\}, i_1 \dots i_m} Y_{i_1} \dots Y_{i_m} \right], \quad (\text{D.2})$$

we see that at $\ell = 1$, we obtain contributions from $\mathcal{K}^{\{m+1\}}$ to $\mathcal{B}^{\{m+1\}}$ [second term in Eq. (D.1)] and $\mathcal{B}^{\{m-1\}}$ [first term in Eq. (D.1)]. Shifting m by one, we can write

$$\begin{aligned} \mathcal{B}^{\{m\}, j_1 \dots j_m} &\supset m! \mathcal{K}^{\{m\}, k j'_1 \dots j'_{m-1}} ([\mathcal{K}^{\{2\}}]^{-1})_{j'_1 j_1} \dots ([\mathcal{K}^{\{2\}}]^{-1})_{j'_{m-1} j_{m-1}} (\hat{\Sigma}^{-1})_{k j_m}, \\ \mathcal{B}^{\{m-2\}, j_1 \dots j_{m-2}} &\supset -(m-2)!(m-1) \mathcal{K}^{\{m\}, k j'_1 \dots j'_{m-1}} ([\mathcal{K}^{\{2\}}]^{-1})_{j'_1 j_1} \dots ([\mathcal{K}^{\{2\}}]^{-1})_{j'_{m-2} j_{m-2}} \\ &\quad \times ([\mathcal{K}^{\{2\}}]^{-1})_{j'_{m-1} k}, \end{aligned} \quad (\text{D.3})$$

where $[\mathcal{K}^{\{2\}}]^{-1} = [\mathcal{K}^{\{2\}}]^{-1}[\{b_O^{\{1\}}\}, \delta_{\text{in}}]$. The first line here provides a one-to-one mapping between $\mathcal{K}^{\{m\}}$ and $\mathcal{B}^{\{m\}}$.¹⁵ The contractions with $[\mathcal{K}^{\{2\}}]^{-1}$ are analogous to the contractions with Σ^{-1} appearing in the expansion around the Gaussian likelihood [cf. the definition of the $\tilde{\mathcal{C}}^{\{m\}}$, Eq. (C.12)]. In addition, there are contributions to $\mathcal{B}^{\{m-2\}}$. These correspond to a shift of lower-order contributions by higher-order ones, analogous to Eq. (C.14). Notice that, for the lowest non-Gaussian stochastic term with $m = 3$, we obtain a unique contribution to $\mathcal{B}^{\{1\}}$:

$$\mathcal{B}^{\{1\},j} = -2\mathcal{K}^{\{3\},kj'j'_2}([\mathcal{K}^{\{2\}}]^{-1})_{j'j}([\mathcal{K}^{\{2\}}]^{-1})_{j'_2k}, \quad (\text{D.4})$$

where $\mathcal{K}^{\{3\}} = \mathcal{K}^{\{3\}}[\{b_O^{\{2\}}\}, \delta_{\text{in}}]$. As mentioned in Sec. 3, this contribution can be removed by redefining

$$Y^l \rightarrow Y^l - 2([\mathcal{K}^{\{2\}}]^{-1})^{lj}([\mathcal{K}^{\{2\}}]^{-1})_{jj'}\mathcal{K}^{\{3\},kj'j'_2}([\mathcal{K}^{\{2\}}]^{-1})_{j'_2k}, \quad (\text{D.5})$$

where the shifted $\mathbf{Y}$ is now a function of $\{b_O^{\{1\}}, b_O^{\{2\}}\}$ in addition to $\{b_O^{\{0\}}\}$. Similarly, the term $\mathcal{B}^{\{2\}}$ can be absorbed by a shift in $\hat{\Sigma}$.

Finally, we turn to $\ell > 1$. First, notice that contributions at order ℓ involve ℓ powers of the kernels $\mathcal{K}^{\{m \geq 3\}}$, so that they will provide subleading corrections to the relation between the $\mathcal{K}^{\{m\}}$ and $\mathcal{B}^{\{m\}}$. Further, it is straightforward to see that, due to the same additional factors of $\mathcal{K}^{\{m \geq 3\}}Y^{m-1}$ that come in at each ℓ , the lowest polynomial order in Y that can be reached at a given ℓ is ℓ . That is, in order to derive the expression for the coefficient $\mathcal{B}^{\{m\}}$ it is sufficient to consider $1 \leq \ell \leq m$.

References

- [1] V. Desjacques, D. Jeong and F. Schmidt, *Large-Scale Galaxy Bias*, *Phys. Rept.* **733** (2018) 1–193, [[1611.09787](#)].
- [2] M. Mirbabayi, F. Schmidt and M. Zaldarriaga, *Biased tracers and time evolution*, *Journal of Cosmology and Astroparticle Physics* **7** (July, 2015) 30, [[1412.5169](#)].
- [3] 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*, [2211.17130](#).
- [4] H. Rubira and F. Schmidt, *The renormalization group for large-scale structure: origin of galaxy stochasticity*, *JCAP* **10** (2024) 092, [[2404.16929](#)].
- [5] N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith and T. Baldauf, *Minimizing the Stochasticity of Halos in Large-Scale Structure Surveys*, *Phys. Rev. D* **82** (2010) 043515, [[1004.5377](#)].
- [6] H. Rubira and F. Conteddu, *Multi-tracer beyond linear theory*, *JCAP* **10** (2025) 111, [[2504.18245](#)].
- [7] R. Voivodic, *Perturbative Likelihoods for Large-Scale Structure of the Universe*, [2505.23750](#).
- [8] J. Jasche and B. D. Wandelt, *Bayesian physical reconstruction of initial conditions from large-scale structure surveys*, *Mon. Not. Roy. Astron. Soc.* **432** (June, 2013) 894–913, [[1203.3639](#)].
- [9] F. S. Kitaura, *The initial conditions of the universe from constrained simulations.*, *Mon. Not. Roy. Astron. Soc.* **429** (Feb., 2013) L84–L88, [[1203.4184](#)].

¹⁵The additional $m!$ factor appearing here can be understood by looking at the stochastic contribution to the $(m+1)$ -point function, which is proportional to $mb_{\perp}^{\{m\}}$ in the noise-field formulation (cf. App. A), whereas it is proportional to $C_{\perp}^{\{m+1\}}$ in the partition function description [cf. Eq. (2.4)]. One could alternatively define the $b_O^{\{m\}}$ with an additional $1/m!$.

- [10] H. Wang, H. J. Mo, X. Yang and F. C. van den Bosch, *Reconstructing the Initial Density Field of the Local Universe: Methods and Tests with Mock Catalogs*, *Astrophys. J.* **772** (July, 2013) 63, [[1301.1348](#)].
- [11] H. Wang, H. J. Mo, X. Yang, Y. P. Jing and W. P. Lin, *ELUCID - Exploring the Local Universe with reConstructed Initial Density field I: Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics*, *Astrophys. J.* **794** (2014) 94, [[1407.3451](#)].
- [12] C. Modi, Y. Feng and U. Seljak, *Cosmological Reconstruction From Galaxy Light: Neural Network Based Light-Matter Connection*, *JCAP* **10** (2018) 028, [[1805.02247](#)].
- [13] F. Schmidt, G. Cabass, J. Jasche and G. Lavaux, *Unbiased Cosmology Inference from Biased Tracers using the EFT Likelihood*, *JCAP* **11** (2020) 008, [[2004.06707](#)].
- [14] F. Schmidt, *Sigma-Eight at the Percent Level: The EFT Likelihood in Real Space*, *JCAP* **04** (2021) 032, [[2009.14176](#)].
- [15] C. J. Shallue and D. J. Eisenstein, *Reconstructing cosmological initial conditions from late-time structure with convolutional neural networks*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 6256–6267, [[2207.12511](#)].
- [16] X. Chen, F. Zhu, S. Gaines and N. Padmanabhan, *Effective cosmic density field reconstruction with convolutional neural network*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 6272–6281, [[2306.10538](#)].
- [17] A. Andrews, J. Jasche, G. Lavaux and F. Schmidt, *Bayesian field-level inference of primordial non-Gaussianity using next-generation galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 5746–5763, [[2203.08838](#)].
- [18] C. Modi, Y. Li and D. Blei, *Reconstructing the universe with variational self-boosted sampling*, *JCAP* **03** (2023) 059, [[2206.15433](#)].
- [19] B. Dai and U. Seljak, *Translation and rotation equivariant normalizing flow (TRENF) for optimal cosmological analysis*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 2363–2373, [[2202.05282](#)].
- [20] A. Kostić, N.-M. Nguyen, F. Schmidt and M. Reinecke, *Consistency tests of field level inference with the EFT likelihood*, *JCAP* **07** (2023) 063, [[2212.07875](#)].
- [21] F. Qin, D. Parkinson, S. E. Hong and C. G. Sabiu, *Reconstructing the cosmological density and velocity fields from redshifted galaxy distributions using V-net*, *JCAP* **06** (2023) 062, [[2302.02087](#)].
- [22] V. Jindal, D. Jamieson, A. Liang, A. Singh and S. Ho, *Predicting the Initial Conditions of the Universe using Deep Learning*, [2303.13056](#).
- [23] T. Charnock, G. Lavaux, B. D. Wandelt, S. Sarma Boruah, J. Jasche and M. J. Hudson, *Neural physical engines for inferring the halo mass distribution function*, *Mon. Not. Roy. Astron. Soc.* **494** (2020) 50–61, [[1909.06379](#)].
- [24] L. Doeser, D. Jamieson, S. Stopyra, G. Lavaux, F. Leclercq and J. Jasche, *Bayesian inference of initial conditions from non-linear cosmic structures using field-level emulators*, *Mon. Not. Roy. Astron. Soc.* **535** (2024) 1258–1277, [[2312.09271](#)].
- [25] I. Babić, F. Schmidt and B. Tucci, *Straightening the Ruler: Field-Level Inference of the BAO Scale with LEFTfield*, [2407.01524](#).
- [26] N.-M. Nguyen, F. Schmidt, B. Tucci, M. Reinecke and A. Kostić, *How Much Information Can Be Extracted from Galaxy Clustering at the Field Level?*, *Phys. Review Lett.* **133** (Nov., 2024) 221006, [[2403.03220](#)].
- [27] G. Cabass and F. Schmidt, *The EFT Likelihood for Large-Scale Structure*, *JCAP* **04** (2020) 042, [[1909.04022](#)].

- [28] G. Cabass and F. Schmidt, *The Likelihood for LSS: Stochasticity of Bias Coefficients at All Orders*, *JCAP* **07** (2020) 051, [[2004.00617](#)].
- [29] H. Rubira and F. Schmidt, *Galaxy bias renormalization group*, *JCAP* **01** (2024) 031, [[2307.15031](#)].
- [30] C. Nikolis, H. Rubira and F. Schmidt, *The renormalization group for large-scale structure: primordial non-Gaussianities*, *JCAP* **08** (2024) 017, [[2405.21002](#)].
- [31] T. Bakx, M. Garny, H. Rubira and Z. Vlah, *Two-loop renormalization and running of galaxy bias*, [2507.13905](#).
- [32] B. Tucci and F. Schmidt, *EFTofLSS meets simulation-based inference: σ_8 from biased tracers*, *Journal of Cosmology and Astroparticle Physics* **2024** (May, 2024) 063, [[2310.03741](#)].
- [33] F. Schmidt, *On the connection between field-level inference and n-point correlation functions*, *JCAP* **09** (2025) 056, [[2504.15351](#)].
- [34] F. Schmidt, F. Elsner, J. Jasche, N. M. Nguyen and G. Lavaux, *A rigorous EFT-based forward model for large-scale structure*, *JCAP* **01** (2019) 042, [[1808.02002](#)].
- [35] BEYOND-2PT collaboration, E. Krause et al., *A Parameter-Masked Mock Data Challenge for Beyond-Two-Point Galaxy Clustering Statistics*, *Astrophys. J.* **990** (2025) 99, [[2405.02252](#)].
- [36] F. Elsner, F. Schmidt, J. Jasche, G. Lavaux and N.-M. Nguyen, *Cosmology inference from a biased density field using the EFT-based likelihood*, *Journal of Cosmology and Astroparticle Physics* **2020** (Jan., 2020) 029, [[1906.07143](#)].
- [37] J. Stadler, F. Schmidt and M. Reinecke, *Fast, accurate and perturbative forward modeling of galaxy clustering. Part I. Galaxies in the restframe*, *JCAP* **04** (2025) 089, [[2409.10937](#)].
- [38] G. Cabass, *The EFT Likelihood for Large-Scale Structure in Redshift Space*, *JCAP* **01** (2021) 067, [[2007.14988](#)].
- [39] J. Stadler, F. Schmidt, M. Reinecke and M. Esposito, *Fast, Accurate and Perturbative Forward Modeling of Galaxy Clustering Part II: Redshift Space*, [2411.04513](#).
- [40] K. Akitsu, M. Simonović, S.-F. Chen, G. Cabass and M. Zaldarriaga, *Cosmology inference with perturbative forward modeling at the field level: a comparison with joint power spectrum and bispectrum analyses*, [2509.09673](#).
- [41] N.-M. Nguyen, F. Schmidt, G. Lavaux and J. Jasche, *Impacts of the physical data model on the forward inference of initial conditions from biased tracers*, *Journal of Cosmology and Astroparticle Physics* **2021** (Mar., 2021) 058, [[2011.06587](#)].