

---

# Simulation-based inference for galaxy clustering

Beatriz Tucci Schiewaldt

---



München 2025



---

# Simulation-based inference for galaxy clustering

Beatriz Tucci Schiewaldt

---

Dissertation  
an der Fakultät für Physik  
der Ludwig-Maximilians-Universität  
München

vorgelegt von  
Beatriz Tucci Schiewaldt  
aus São Paulo, Brasilien

München, den 22.04.2025

Erstgutachter: Prof. Dr. Eiichiro Komatsu

Zweitgutachter: Prof. Dr. Daniel Grün

Tag der mündlichen Prüfung: 26.05.2025

# Contents

|                                                            |            |
|------------------------------------------------------------|------------|
| <b>Zusammenfassung</b>                                     | <b>xix</b> |
| <b>1 Introduction</b>                                      | <b>1</b>   |
| <b>2 Galaxy clustering</b>                                 | <b>9</b>   |
| 2.1 Gravity . . . . .                                      | 10         |
| 2.2 Perturbation Theory for LSS . . . . .                  | 12         |
| 2.2.1 Gravitational fluid . . . . .                        | 12         |
| 2.2.2 EPT . . . . .                                        | 12         |
| 2.2.3 The Fluid Equations . . . . .                        | 12         |
| 2.2.4 SPT . . . . .                                        | 13         |
| 2.2.5 EFTofLSS . . . . .                                   | 16         |
| 2.2.6 LPT . . . . .                                        | 17         |
| 2.3 Galaxy bias . . . . .                                  | 19         |
| 2.3.1 The leading local gravitational observable . . . . . | 19         |
| 2.3.2 EFT description . . . . .                            | 20         |
| 2.3.3 Time factorization and finite basis . . . . .        | 21         |
| 2.3.4 Lagrangian Bias . . . . .                            | 22         |
| 2.3.5 Eulerian Bias . . . . .                              | 23         |
| 2.3.6 Gravity and non-locality . . . . .                   | 24         |
| 2.3.7 Advection of operators . . . . .                     | 25         |
| 2.3.8 Higher-order derivative terms . . . . .              | 25         |
| 2.3.9 Stochasticity . . . . .                              | 26         |
| 2.4 Wilson-Polchinski for galaxies . . . . .               | 27         |
| 2.4.1 Wilson-Polchinski RG . . . . .                       | 27         |
| 2.4.2 The partition function of galaxies . . . . .         | 28         |
| 2.4.3 Galaxy bias RG equations . . . . .                   | 30         |
| 2.4.4 The field-level galaxy likelihood . . . . .          | 32         |
| 2.5 LEFTfield . . . . .                                    | 33         |
| <b>3 Cosmological inference</b>                            | <b>37</b>  |
| 3.1 Bayesian inference . . . . .                           | 37         |
| 3.2 Standard cosmological inference . . . . .              | 38         |

|          |                                                                                             |            |
|----------|---------------------------------------------------------------------------------------------|------------|
| 3.3      | Why SBI for galaxy clustering? . . . . .                                                    | 38         |
| 3.4      | Simulation-based inference . . . . .                                                        | 41         |
| 3.4.1    | Approximate Bayesian Computation . . . . .                                                  | 41         |
| 3.4.2    | Neural density estimators . . . . .                                                         | 41         |
| 3.4.3    | Normalizing flows . . . . .                                                                 | 44         |
| 3.4.4    | Simulation-based calibration . . . . .                                                      | 48         |
| 3.5      | Data compression techniques . . . . .                                                       | 49         |
| 3.5.1    | MOPED . . . . .                                                                             | 49         |
| 3.5.2    | MSE . . . . .                                                                               | 49         |
| 3.5.3    | VMIM . . . . .                                                                              | 50         |
| 3.6      | Field-level inference . . . . .                                                             | 52         |
| 3.6.1    | Hamiltonian Monte Carlo . . . . .                                                           | 53         |
| 3.6.2    | Marginalization over galaxy bias . . . . .                                                  | 57         |
| 3.7      | Inferring $\sigma_8$ from galaxies . . . . .                                                | 58         |
| <b>4</b> | <b>EFTofLSS meets simulation-based inference: <math>\sigma_8</math> from biased tracers</b> | <b>61</b>  |
| 4.1      | Introduction . . . . .                                                                      | 61         |
| 4.2      | Methods . . . . .                                                                           | 62         |
| 4.2.1    | LEFTfield and SBI . . . . .                                                                 | 62         |
| 4.2.2    | Summary statistics . . . . .                                                                | 63         |
| 4.3      | Results . . . . .                                                                           | 65         |
| 4.3.1    | Linear forward model . . . . .                                                              | 65         |
| 4.3.2    | 2LPT forward model . . . . .                                                                | 68         |
| 4.4      | Tests of the inference . . . . .                                                            | 80         |
| 4.5      | Conclusions . . . . .                                                                       | 84         |
| <b>5</b> | <b>How much information can be extracted from galaxy clustering at the field level?</b>     | <b>87</b>  |
| 5.1      | Introduction . . . . .                                                                      | 87         |
| 5.2      | Methods . . . . .                                                                           | 88         |
| 5.3      | Results . . . . .                                                                           | 90         |
| 5.4      | Tests of the inference: FBI . . . . .                                                       | 91         |
| 5.5      | Tests of the inference: SBI . . . . .                                                       | 98         |
| 5.6      | Summary and Discussion . . . . .                                                            | 102        |
| <b>6</b> | <b>The information content in the galaxy trispectrum</b>                                    | <b>105</b> |
| 6.1      | Trispectrum . . . . .                                                                       | 105        |
| 6.2      | Estimators . . . . .                                                                        | 106        |
| 6.2.1    | Trispectrum estimator . . . . .                                                             | 106        |
| 6.2.2    | Integrated trispectrum . . . . .                                                            | 107        |
| 6.3      | Validation of the trispectrum estimator . . . . .                                           | 108        |
| 6.3.1    | $T_{1122}$ . . . . .                                                                        | 109        |
| 6.3.2    | $T_{1113}$ . . . . .                                                                        | 110        |

---

|          |                                                                     |            |
|----------|---------------------------------------------------------------------|------------|
| 6.4      | Trispectrum covariance . . . . .                                    | 111        |
| 6.5      | Results . . . . .                                                   | 115        |
| 6.5.1    | LEFTfield settings . . . . .                                        | 115        |
| 6.5.2    | Trispectrum information content . . . . .                           | 115        |
| 6.6      | Conclusion & next steps . . . . .                                   | 126        |
| <b>7</b> | <b>Conclusion</b>                                                   | <b>129</b> |
| <b>A</b> | <b>Bispectrum and trispectrum stochasticity</b>                     | <b>131</b> |
| A.1      | Leading order stochastic bispectrum . . . . .                       | 131        |
| A.2      | Leading order stochastic trispectrum . . . . .                      | 132        |
| <b>B</b> | <b>Appendix C: Bias relations</b>                                   | <b>135</b> |
| <b>C</b> | <b>Optimizing SBI</b>                                               | <b>137</b> |
| C.1      | Convergence with simulation budget . . . . .                        | 137        |
| C.2      | Hyperparameter optimization . . . . .                               | 138        |
| C.3      | Data compression . . . . .                                          | 140        |
| <b>D</b> | <b>On the non-Gaussianity of the <math>n</math>-point functions</b> | <b>147</b> |
| <b>E</b> | <b>Field-level inference of the BAO scale</b>                       | <b>151</b> |
| E.1      | Fixed initial conditions . . . . .                                  | 151        |
| E.2      | Free initial conditions . . . . .                                   | 153        |
| E.3      | Conclusions and next steps . . . . .                                | 157        |
|          | <b>Danksagung</b>                                                   | <b>173</b> |



# List of Figures

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |    |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Schematic visualization of galaxy $n$ -point functions (power-spectrum, bispectrum and trispectrum) on Hubble Ultra Deep Field (UDF). Note that galaxies are positioned on the vertices of each $n$ -point function, although stars could be located at the vertices in this image instead. Furthermore, galaxy clustering analysis are focused on spectroscopic galaxy surveys instead of Hubble UDF. These surveys typically map the positions of millions of galaxies over 3D volumes extending billions of light years. <i>Credit for Hubble UDF: NASA, ESA, and S. Beckwith (STScI) and the HUDF Team.</i>                                                                                                                                                                                                 | 3  |
| 1.2 | <i>Left:</i> Diagram showing the comparison between FBI, where the full galaxy density field is used for inference using a field-level likelihood, and SBI, where the galaxy power-spectrum and bispectrum (two- and three- point functions) are used. Both inference methods are based on LEFTfield, which evolves the Universe initial conditions (ICs) to the final 3D distribution of galaxies given cosmological parameters. <i>Right:</i> comparison of constraints of the cosmological parameter $\sigma_8$ for one particular dark-matter halo and scale cut considered in the analysis, where the galaxy density field shows to provide 5 times more information than the combined two- and three-point functions. <i>Credit for Hubble UDF: NASA, ESA, and S. Beckwith (STScI) and the HUDF Team.</i> | 6  |
| 2.1 | Galaxy formation picture, happening on large time scales and small spatial scales. Adapted from [66].                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 21 |
| 2.2 | Normalized linear power spectrum at different redshifts and corresponding non-linear scale $k_{\text{NL}}$ , which increases with redshift due to non-linear clustering. The cutoff scale $\Lambda$ is also indicated, where the blue shaded region correspond to the small-scales that are integrated over in the EFTofLSS. Adapted from [69].                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 29 |
| 2.3 | Schematic representation of $\delta_{\Lambda'}^{(1)}$ : the shell is given by the part in between $\Lambda$ and $\Lambda'$ .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 30 |
| 2.4 | Diagram for the construction of Eulerian bias operators in LEFTfield.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 34 |
| 3.1 | Different neural density estimation algorithms and their sequential versions. From [137].                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 42 |

- 3.2 Diagram of how normalizing flows work, with the specific example of Masked Autoregressive Flows. The samples from the vector  $\mathbf{z}_0 = \mathbf{u}$ , sampled from the simple distribution  $\pi(\mathbf{u})$ , are deformed through the sequence of transformations  $f = f_T \circ \dots \circ f_1$  into those of  $\mathbf{z}_T = \mathbf{x}$ , which follow a more complex distribution  $p(\mathbf{x})$ . In the lower panel, we illustrate the conditioner that “masks out” the connections between  $z_i$  and  $\mathbf{h}_{<i}$ , as well as the affine functions applied to the vector components. . . . . 46
- 3.3 Schematic representation of field-level inference through the forward model `LEFTfield`. The initial conditions  $\delta_\Lambda^{(1)}$  are evolved with the forward model to obtain  $\delta_{g,\text{det}}$ , which is further used in the EFT likelihood together with  $\delta_{g,\text{obs}}$ . New parameters are proposed and the process is repeated to sample the posterior until convergence. *Credits: Julia Stadler*. . . . . 52
- 3.4  $L = 100$  Leapfrog steps with size  $\varepsilon = 0.25$ . The initial state is  $q = (-1.50, -1.55)$  and  $p = (-0.78, 0.94)$ . The rejection was 0.05. . . . . 56
- 3.5  $L = 20$  Leapfrog steps with size  $\varepsilon = 0.18$ . The initial state is  $q = (-1.50, -1.55)$ . The rejection was 0.025. . . . . 57
- 3.6  $L = 100$  Leapfrog steps with size  $\varepsilon = 0.25$ . The initial state is  $q = (-1.50, -1.55)$ . The rejection was 0.39. . . . . 57
- 4.1 Comparison of power spectrum and bispectrum estimators to theory for the linear forward model in Eq. (4.6), where we use  $\Lambda = 0.3h\text{Mpc}^{-1}$ ,  $k_{\text{min}} = 2\pi L^{-1}h\text{Mpc}^{-1}$  and  $k_{\text{max}} = 0.2h\text{Mpc}^{-1}$  for both cases. The blue lines show the  $n$ -point function estimator from `LEFTfield` as described in this section, which agree with the theory predictions indicated by the red curves. *Left*: we set  $b_1 = b_2 = 1$  and use 50 linear bins for the power spectrum. The linear power spectrum  $P_L$  is displayed in black for reference. *Right*: we set  $b_1 = 1$  and  $b_2 = 0.1$ , in order to suppress next-to-leading-order corrections, and use 150 triangle bins for a better visualization of the bispectrum, which are constructed from 10 linear bins in  $k$ . The lowest triangle bin indexes contain the smaller  $k$ -modes, hence their large variance, where  $k_1 \leq k_2 \leq k_3$ , where  $k_1$  corresponds to the outermost loop index, while  $k_3$  is the innermost index. . . . . 64
- 4.2 Flowchart of our pipeline for SBI, assuming an  $n$ LPT forward model. The cosmological, bias and noise parameters  $\boldsymbol{\theta}$  are sampled from the prior and used as input for `LEFTfield` (Sec. 2.5), which constructs a model for the galaxy overdensity field  $\delta_g$  through  $n$ LPT (2LPT in this work). We then measure the summary statistics  $\mathbf{x}$ , which in our case is the galaxy power-spectrum concatenated with the bispectrum (Sec. 4.2.2). We repeat this process  $N_{\text{sim}}$  times, and use the resulting pairs  $\{(\boldsymbol{\theta}_n, \mathbf{x}_n)\}_{n=1}^{N_{\text{sim}}}$  as input to the SBI neural density estimation (Sec. 3.4.2), which then gives us the posterior  $p(\boldsymbol{\theta}|\mathbf{x})$  that can in turn be evaluated at a given observed data  $\mathbf{x}_o$ . . . . . 66

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.3  | Posterior standard deviation of the linear bias parameter squared $b_1^2$ obtained by SBI divided by its analytical prediction as a function of the total simulation budget $N_{\text{sim}}$ . Continuous lines show the results obtained from ABC, while the dashed line corresponds to NPE. The value of $f$ denotes the fractional percentage of samples which were accepted in ABC, where we use $f \times N_{\text{sim}}$ for posterior estimation. The error bars correspond to 10 different realizations of the observational mock data $\mathbf{x}_o$ . . . . .                                                                                    | 67 |
| 4.4  | Parameter posterior for the case where the cosmological parameter $\alpha$ is fixed. The left and right contour plots correspond to the models where the data vector is sampled from a Gaussian with analytical and sample covariance, respectively. For all cases, the neural density estimator method NPE was used with a simulation budget of $N_{\text{sim}} = 10^5$ . . . . .                                                                                                                                                                                                                                                                         | 70 |
| 4.5  | Parameter posterior for the case where the cosmological parameter $\alpha$ is fixed. Blue and purple contours correspond to the models where the data vector is sampled from a Gaussian likelihood with analytical and sample covariance, respectively, where $N_{\text{cov}} = 10^4$ . The red contours show the full SBI results, with no Gaussian assumption. Dotted lines indicate the Fisher prediction with sample covariance for reference. For all cases, the method NPE was used from a simulation budget of $N_{\text{sim}} = 10^5$ , scale cut of $k_{\text{max}} = \Lambda = 0.1h\text{Mpc}^{-1}$ and data vector dimension $D = 33$ . . . . . | 72 |
| 4.6  | Same as Fig. 4.5, but for $k_{\text{max}} = \Lambda = 0.2h\text{Mpc}^{-1}$ and $D = 49$ . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 73 |
| 4.7  | SBC (simulation-based calibration) tests for the full-SBI case where the cosmological parameter $\alpha$ is fixed with $N_{\text{sim}} = 10^5$ . The upper panels show the CDF of the ranks distributions for each parameter, where the grey area show the 95% confidence interval of a uniform distribution, while the lower panels show the rank distribution, where the grey areas denote the 99% confidence interval of a uniform distribution. . . . .                                                                                                                                                                                                | 74 |
| 4.8  | Convergence of the full-SBI parameter posterior with the number of simulations $N_{\text{sim}}$ for the case where the cosmological parameter $\alpha$ is fixed. The standard deviation of the posterior samples are normalized by their respective Fisher prediction for better comparison. Note that it should not necessarily converge to one. The errors indicate posterior evaluation at 10 different data observations, with same fiducial parameters but different initial conditions realizations. . . . .                                                                                                                                         | 75 |
| 4.9  | Parameter posterior for the case where the cosmological parameter $\alpha$ is sampled together with the bias and noise parameters. The contour colors, simulation budget and data vector size are the same as in Fig. 4.5. For all cases, the method NPE was used from a simulation budget of $N_{\text{sim}} = 10^5$ , scale cut of $k_{\text{max}} = \Lambda = 0.1h\text{Mpc}^{-1}$ and data vector dimension $D = 33$ . . . . .                                                                                                                                                                                                                         | 77 |
| 4.10 | Same as Fig. 4.9, but for $k_{\text{max}} = \Lambda = 0.2h\text{Mpc}^{-1}$ and $D = 49$ . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 78 |
| 4.11 | Same as Fig. 4.7 but for the case where the cosmological parameter $\alpha$ is sampled as well, with $N_{\text{sim}} = 10^5$ . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 79 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.12 | Same as Figure 4.8, but showing convergence for the full case where the cosmological parameter $\alpha$ is sampled. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 79 |
| 4.13 | <i>Left</i> : Convergence of the standard deviation of the posterior of each parameter with respect to the simulation budget $N_{\text{sim}}$ normalized by their Fisher prediction. Solid lines denote the case where the data vector corresponds to the raw spectra, while dashed lines display the results using the normalized spectra. <i>Right</i> : Standard deviation of the parameters from SNPE starting from the prior of Eq. (4.14) as a function of its rounds, where $10^3$ simulations are sampled at each round, divided by the NPE standard deviation of each parameter obtained from $N_{\text{sim}} = 10^5$ . . . . . | 80 |
| 4.14 | <i>Left</i> : posteriors obtained with NLE, NPE and SNPE (at round 22). <i>Right</i> : Parameter posterior comparing the prior range in grey with the constraints from power spectrum only (dark cyan) and the power spectrum combined with the bispectrum (cyan). . . . .                                                                                                                                                                                                                                                                                                                                                               | 81 |
| 4.15 | <i>Left</i> : impact of the number of simulations for covariance estimation on the final posterior for the Gaussian-likelihood case with sample covariance. <i>Right</i> : impact of doubling the number of hidden units and number of transforms. . . . .                                                                                                                                                                                                                                                                                                                                                                               | 82 |
| 4.16 | Same as Fig. 4.9, but using an ensemble of networks instead of a single posterior estimation. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 83 |
| 5.1  | Posterior constraints on $\alpha = \sigma_8/\sigma_{8,\text{true}}$ from the SNG sample (see text) at $k_{\text{max}} = [0.10, 0.12]h\text{Mpc}^{-1}$ . Vertical bands denote 68% confidence limits. The upper right corners indicate the ratio of 1- $\sigma$ constraints between FBI (blue) and SBI P+B (yellow). . . . .                                                                                                                                                                                                                                                                                                              | 92 |
| 5.2  | Similar to Fig. 5.1, but for the Uchuu sample (see text) at $k_{\text{max}} = [0.10, 0.12]h\text{Mpc}^{-1}$ . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 93 |
| 5.3  | Posterior distributions comparing the fiducial $\{b_O\}$ -marginalized likelihood (blue, open) and the $b_\delta$ -unmarginalized likelihood (yellow, filled). The figure presents 1D and 2D constraints on $\alpha = \sigma_8/\sigma_{8,\text{fid.}}$ , the linear bias parameter $b_\delta$ , and the galaxy stochasticity parameters $[\sigma_{\epsilon,0}, \sigma_{\epsilon,k^2}]$ . Contours indicate 68% and 95% confidence intervals. . . . .                                                                                                                                                                                     | 94 |
| 5.4  | Comparison of posterior constraints obtained using the Eulerian bias (blue, open) and Lagrangian bias (yellow, filled). The figure displays constraints on the amplitude rescaling parameter $\alpha = \sigma_8/\sigma_{8,\text{fid.}}$ and the effective noise amplitude $\sigma_\epsilon^{\text{eff.}} = \sigma_{\epsilon,0}[1 + \sigma_{\epsilon,k^2}k_{\text{max}}^2]$ . Contours denote 68% and 95% confidence regions of the 2D marginal posteriors, while vertical bands represent the 68% intervals of the 1D marginal posteriors. . . . .                                                                                       | 95 |
| 5.5  | Trace plot of the amplitude rescaling parameter $\alpha = \sigma_8/\sigma_{8,\text{fid.}}$ from TPI and RPI chains in the FBI analysis. Different colors indicate distinct initial conditions. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                   | 96 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |     |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.6  | Trace plots of various statistics for the posterior initial conditions $\hat{s}$ from the TPI chain in Fig. 5.5. The warm-up phase is omitted, and samples are thinned by a factor of 100 for clarity. Grey vertical bands mark scales beyond the cutoff $k_{\max} = 0.1h\text{Mpc}^{-1}$ . (Top) Histogram of $\hat{s} - \hat{s}_{\text{true}}$ . (Middle) Square root of the ratio between inferred and true initial power spectra. (Bottom) Variance of $\hat{s} - \hat{s}_{\text{true}}$ . . . . . | 97  |
| 5.7  | Posterior consistency of MCMC chains with different initializations. Colors correspond to the chains in Fig. 5.6. Contours represent the 68% and 95% credible intervals of the 1D and 2D marginal posteriors for $[\alpha, \sigma_{\epsilon}^{\text{eff}}]$ . . . . .                                                                                                                                                                                                                                  | 98  |
| 5.8  | Rank distributions from the SBC tests for the SBI posteriors shown in Fig. 5.1 and Fig. 5.2. The grey shaded area indicates the 99% confidence interval of an uniform distribution. Upper, middle and lower panels correspond to the SNG ( $k_{\max} = 0.1h\text{Mpc}^{-1}$ ), Uchuu ( $k_{\max} = 0.1h\text{Mpc}^{-1}$ ), and SNG ( $k_{\max} = 0.12h\text{Mpc}^{-1}$ ) analyses, respectively. . . . .                                                                                               | 100 |
| 5.9  | Convergence of the standard deviations of $\alpha$ posteriors in SBI P+B analyses with increasing simulation budget $N_{\text{sim}}$ used for posterior estimation. Each value is normalized by the corresponding $\alpha$ constraint from a Fisher analysis, for each of the cases considered. . . . .                                                                                                                                                                                                | 100 |
| 5.10 | Parameter posteriors obtained in the SBI P+B analysis of the Uchuu halo sample at redshift $z = 0.5$ with $k_{\max} = 0.1h\text{Mpc}^{-1}$ . Contours indicate 68% and 95% credible intervals. Dashed lines correspond to the Fisher analysis for this case, using sample covariance for the data vector. . . . .                                                                                                                                                                                      | 101 |
| 5.11 | Comparison of $\alpha$ posteriors in SBI P+B analyses, incorporating non-Gaussian stochasticity, as detailed in Section 5.5. Vertical bands denote 68% credible intervals. Upper right corners show the ratio of 1- $\sigma$ constraints between FBI (blue) and SBI with non-Gaussian stochasticity (pink). . . . .                                                                                                                                                                                    | 102 |
| 6.1  | Schematic representation of the galaxy trispectrum, where each galaxy is in one of the vertices of the tetrahedrum. . . . .                                                                                                                                                                                                                                                                                                                                                                            | 106 |
| 6.2  | Total trispectrum estimator measured in a Gaussian random field. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                               | 108 |
| 6.3  | Maximum and minimum values that the diagonal $k_{14}$ can take. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                | 110 |
| 6.4  | Trispectrum estimator averaged over 2000 realizations (in black) and theoretical prediction $b_1^2 b_2^2 T_{1122}$ (in red) as a function of tetrahedrum bin. The standard deviation over the realizations is indicated by the shaded grey area. . . . .                                                                                                                                                                                                                                               | 111 |
| 6.5  | Normalized histogram of $(\langle \hat{T} \rangle - b_1^3 b_3 T_{1113}) / \sigma(\hat{T})$ , where $\hat{T}$ are the estimated trispectra for $b_1 = 1$ and $b_3 = 0.1$ and $T_{1113}$ is the prediction. . . . .                                                                                                                                                                                                                                                                                      | 112 |

- 6.6 Reduced covariance matrix for the galaxy power-spectrum, bispectrum and trispectrum bins. The left columns show the covariance for  $z = 0.5$  and bias parameters fixed to the MAP from SNG halos posterior, while the right column shows the same for the Uchuu halos at  $z = 1.03$ . The colorbar on the upper panels show the value of the reduced covariance matrix for each bin, while the lower panels have a maximum value of 0.3 to contrast the difference between both redshifts. In general, the data vector at  $z = 0.5$  is more strongly correlated among between different bins than at  $z = 1.03$ . . . 113
- 6.7 Diagonal part of the sample covariance  $\hat{C}$  of the galaxy power-spectrum (*upper left*), bispectrum (*upper right*) and trispectrum (*bottom*) measured from LEFTfield, compared to the corresponding theoretical covariance  $C^T$ . Red lines correspond to redshift  $z = 0.5$  while blue lines indicate  $z = 1.03$ . 114
- 6.8 Posterior distribution of all model parameters inferred using Uchuu halos at redshift  $z = 1.03$  as the “observed” data vector. Red and blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For both cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 8 \times 10^5$ , scale cut of  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$  and the baseline prior given in Eq. (6.28). . . . . 116
- 6.9 Posterior distribution of all model parameters inferred using SNG halos at redshift  $z = 0.5$  as the “observed” data vector. Red and blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For both cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 8 \times 10^5$ , scale cut of  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$  and the baseline prior given in Eq. (6.28). . . . . 117
- 6.10 Same as Fig. 6.8, but using a simulation budget of  $N_{\text{sim}} = 5 \times 10^5$  and the narrower prior given in Eq. (6.29). . . . . 119
- 6.11 Same as Fig. 6.9, but using a simulation budget of  $N_{\text{sim}} = 5 \times 10^5$  and the narrower prior given in Eq. (6.29). . . . . 120
- 6.12 Posterior distribution of the model parameters using mock data generated from the same model at redshift  $z = 0.5$  when only sampling  $\{b_\delta, b_{\delta^2}, b_{\delta^3}\}$  and using the trivial forward model. Dark and light blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For both cases, the method SNPE was used using 15 rounds and a simulation budget of  $N_{\text{sim}} = 10^4$  per round, and scale cut of  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$ . . . . . 122

- 6.13 Posterior distribution of the model parameters inferred using mock data generated from `LEFTfield` when only sampling  $\{\alpha, b_\delta\alpha, b_{\delta^2}\alpha^2, b_{\delta^3}\alpha^3\}$  and using 2LPT. *Left*: Inference at redshift  $z = 1.0$ . Light and dark pink contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. *Right*: Inference at redshift  $z = 0.5$ . Dark and light blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For all cases, the method SNPE was used using 15 rounds and a simulation budget of  $N_{\text{sim}} = 10^4$  per round, and scale cut of  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$ . . . . . 123
- 6.14 Posterior distribution of the model parameters inferred using mock data from SNG FTP parameters at redshift  $z = 0.5$  as the “observed” data vector consisting of the power-spectrum, bispectrum and trispectrum. In this case, the parameter space is reduced to include the cosmological parameter  $\alpha$  and deterministic bias parameters. The data vector generated for training is sampled from a Gaussian likelihood. Grey and orange contours correspond to when we use an analytical or sample covariance. For both cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 5 \times 10^5$ , scale cut of  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$ . . . . . 125
- C.1 *Left*: standard deviation of  $b_1^2$  obtained by NPE as a function of simulation budget used in training. *Right*: logarithm of the best validation loss over all epochs obtained by NPE as a function of simulation budget used in training. 138
- C.2 *Left*: logarithm of the best validation loss over all epochs obtained by NPE as a function of simulation budget used in training. *Right*: logarithm of the validation (training) loss over batches obtained by NPE in continuous (dashed) lines as a function of training epoch for different simulation budgets. . . . . 139
- C.3 Hyperparameter optimization for NPE when using  $N_{\text{sim}} = 5 \times 10^4$ . Grey (black) horizontal lines indicates NPE results using baseline hyperparameters and  $N_{\text{sim}} = 10^5$  ( $N_{\text{sim}} = 5 \times 10^5$ ). See the hyperparameters list C.2 for a description of specific hyperparameters chosen by `OPTUNA` in each label. *Top left*: logarithm of the best validation loss over all epochs obtained by NPE as a function of `OPTUNA` trial. No model presents a better validation loss than the baselines with more simulations. *Top right*: standard deviation of  $\alpha$  obtained by NPE as a function of `OPTUNA` trial. The models with lower error on  $\alpha$  than the baseline with  $N_{\text{sim}} = 5 \times 10^5$  (in grey) are displayed with a star mark. *Bottom*: correlation of standard deviation of  $\alpha$  and best validation loss over all epochs obtained by NPE for each `OPTUNA` trial. . . . 142
- C.4 Same as Fig. C.3, but the hyperparameter optimization for NPE is done when using  $N_{\text{sim}} = 10^5$ . The models with higher validation loss than the baseline with  $N_{\text{sim}} = 10^5$  (in black) are displayed with a star mark. . . . . 143

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |     |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| C.5 | Same as Fig. C.3, but the hyperparameter optimization for NPE is done when using $N_{\text{sim}} = 5 \times 10^5$ . The models with higher validation loss than the baseline with $N_{\text{sim}} = 5 \times 10^5$ (in grey) are displayed with a star mark. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 144 |
| C.6 | Data compression impact on convergence properties of NPE. Solid and dashed lines correspond to different runs with exact same configurations and indicate the intrinsic network error. <i>Left</i> : logarithm of the best validation loss over all epochs obtained by each method as a function of simulation budget used in training. <i>Right</i> : standard deviation of $\alpha$ obtained by NPE as a function of simulation budget used in training. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 145 |
| D.1 | Distribution of galaxy $n$ -point functions over initial conditions realizations measured in LEFTfield (in colored lines) compared to a Gaussian distribution (in grey). <i>Top left</i> : measured power-spectrum distribution $\mathcal{P}$ (in green) and Gaussian distribution $\mathcal{P}_{\text{Gaussian}}$ (in grey) for different $k$ -bins. <i>Top right</i> : measured bispectrum distribution $\mathcal{B}$ (in blue) and Gaussian distribution $\mathcal{B}_{\text{Gaussian}}$ (in grey) for different triangle bins. We only select equilateral bins in this illustration ( $k_1 = k_2 = k_3$ ). <i>Bottom</i> : measured integrated trispectrum distribution $\mathcal{T}$ (in red) and Gaussian distribution $\mathcal{T}_{\text{Gaussian}}$ (in grey) for different quadrilateral bins. We only select equilateral bins in this illustration ( $k_1 = k_2 = k_3 = k_4$ ). The bump observed in the first panel is due to discreteness effects (i.e, it would not be present if we used more simulations). | 148 |
| D.2 | KL divergence between the full and Gaussian distribution of each $n$ -point function. The vertical lines in the bispectrum and trispectrum figures correspond to when the innermost index starts running again, i.e., when we go to large scales at the left parts of the plots. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 149 |
| E.1 | Ratio of the uncertainty on the BAO scale inferred from the power spectrum likelihood, $\sigma_{PS}(\hat{\beta})$ , to that from the field-level likelihood, $\sigma_F(\hat{\beta})$ , as a function of cutoff for different redshifts. Each panel corresponds to a different halo mass range. From From [23]. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 154 |
| E.2 | Flowchart of the two methods for BAO inference employed in this work: field-level inference (top) and power spectrum-based inference with and without reconstruction (bottom). The reconstruction step includes the generation of a discrete tracer catalog from $\delta_g$ , as described in the text. From [24].                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 156 |
| E.3 | Inferred error bar $\sigma_{PS}(\beta)$ on the BAO scale using pre- and post-reconstruction power spectrum, relative to that in the field-level inference $\sigma_F$ , in the case of Mock A. This mock was generated (and sampled, in case of field-level inference) using Lagrangian bias. Pre-reconstruction results are depicted using squares, whereas post-reconstruction results are represented by stars. It is evident that, even comparing to power spectrum after BAO reconstruction, the field-level BAO scale inference is more precise, by up to a factor of 1.35. From [24]. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                        | 157 |

# List of Tables

|     |                                                                                                                                                                                                                                                                                                                                                                                           |     |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.1 | ESS estimates for $\alpha$ in various FBI analyses. . . . .                                                                                                                                                                                                                                                                                                                               | 99  |
| 6.1 | Results of inference on the cosmological parameter $\alpha$ obtained from dark-matter halos, with values for the mean and 68% confidence level (CL) error on $\alpha$ from by the posterior samples obtained by NPE. We compare results using the baseline and narrower priors, for PB (power-spectrum and bispectrum) and PBT (power-spectrum, bispectrum and trispectrum) data vectors. | 121 |



# Zusammenfassung

Die großräumige Galaxienverteilung enthält grundlegende physikalische Informationen, die sich durch kosmologische Inferenz erschließen lassen. Parameter wie die Amplitude der Materiedichtefluktuationen  $\sigma_8$  ermöglichen Einblicke in die Dunkle Energie und testen die Allgemeine Relativitätstheorie. In dieser Arbeit entwickeln wir neue Inferenzmethoden, um aus spektroskopischen Galaxienvermessungen mehr Information zu extrahieren und robustere kosmologische Aussagen zu treffen.

Im Zentrum steht die Anwendung von *Simulationsbasierter Inferenz* (SBI), ein Deep-Learning-Ansatz zur Parameterschätzung ohne angenäherte Likelihoods oder Kovarianzmatrizen. Durch das direkte Lernen der Posteriorverteilung aus Simulationen umgeht SBI typische Schwächen klassischer Verfahren, etwa die systematische Unterschätzung von Unsicherheiten, und bietet ein leistungsfähiges Rahmenwerk für die Analyse von Galaxienverteilungen.

Die zugrunde liegenden Simulationen werden mit `LEFTfield` erzeugt – einem schnellen Lagrangeschen Vorwärtsmodell, das auf der Effektiven Feldtheorie großräumiger Strukturen (EFTofLSS) und der Bias-Expansion basiert und dadurch eine konsistente und kontrollierte Inferenz ermöglicht. Als primäre Zusammenfassungsgrößen verwenden wir  $n$ -Punkt-Funktionen der Galaxienverteilung, wie sie aus perturbativen Ansätzen zur Strukturbildung motiviert sind. Darüber hinaus erlaubt `LEFTfield` eine *Feld-basierte Inferenz*, bei der das vollständige dreidimensionale Galaxienfeld direkt ausgewertet wird, anstatt auf komprimierte Statistiken zurückzugreifen.

Zunächst demonstrieren wir die Anwendung von SBI mit `LEFTfield` anhand von Mock-Daten, indem wir kosmologische Parameter aus dem Galaxienleistungs- und Bispektrum ableiten und die Methode anschließend an Halovertelungen testen. Dabei zeigen wir, dass das Galaxienfeld bis zu fünfmal mehr Information über  $\sigma_8$  enthält als das Leistungs- und Bispektrum allein.

Abschließend widmen wir uns der Analyse des Galaxientrispektrums – einer Aufgabe, die mit klassischen Methoden aufgrund der hohen Dimensionalität bisher nicht praktisch umsetzbar ist. Wir zeigen, wie SBI eine gemeinsame Auswertung von Leistungs-, Bis- und Trispektrum ermöglicht und so den Zugang zu bislang unerschlossener kosmologischer Information eröffnet. Diese Ergebnisse unterstreichen das Potenzial simulationsbasierter Inferenzverfahren, den wissenschaftlichen Ertrag zukünftiger Galaxiendurchmusterungen erheblich zu steigern und unser Verständnis fundamentaler Physik zu vertiefen.



# Abstract

The large-scale distribution of galaxies encodes valuable information about fundamental physics, which can be accessed by means of cosmological inference. The cosmological parameters such as the amplitude of matter fluctuations  $\sigma_8$  offer a unique window to test the nature of dark energy and the validity of General Relativity. In this thesis, we build on novel inference methods to enable the extraction of more information from spectroscopic galaxy surveys, leading to tighter and more reliable cosmological constraints.

In particular, we employ *simulation-based inference* (SBI), a deep-learning framework that allows for robust parameter estimation without relying on approximate analytical likelihoods or covariance matrices. By directly learning the posterior distribution from simulations, SBI avoids common problems of traditional methods, such as underestimation of errors, and represents a powerful framework for galaxy clustering analysis.

The simulations are generated using `LEFTfield`, a fast Lagrangian forward model based on the Effective Field Theory of Large-Scale Structure (EFTofLSS) and the bias expansion, ensuring robustness in the inference process. We adopt galaxy  $n$ -point functions as our primary summary statistics, motivated by perturbative approaches to LSS. Furthermore, `LEFTfield` enables *field-level inference*, where we directly analyze the three-dimensional galaxy density field instead of compressed summary statistics.

We first demonstrate SBI with `LEFTfield` using mock data, performing cosmological inference with the galaxy power spectrum and bispectrum, and subsequently validating the method on dark matter halos. We use SBI to demonstrate that the galaxy density field contains up to five times more information on  $\sigma_8$  when compared to the power spectrum and bispectrum.

Finally, we tackle the analysis of the galaxy trispectrum, a task that remains computationally infeasible with traditional methods due to its high dimensionality. We show how to use SBI to perform a joint analysis of the galaxy power-spectrum, bispectrum and trispectrum to reveal the cosmological information in the trispectrum. Together, these results demonstrate the promise of simulation-based inference methods for maximizing the scientific return of future galaxy surveys and advancing our understanding of fundamental physics.



# Chapter 1

## Introduction

There are still many unresolved questions in physics, such as the ultimate theory of gravity, the nature of dark matter and dark energy, the hierarchy of neutrino masses, and the physics of the early Universe – all of which shape the distribution of galaxies in the sky. In this thesis, we explore how to tackle the challenge of maximizing the information extracted from observations of the large-scale distribution of galaxies, while ensuring a robust analysis. Our ultimate goal is to provide a trustworthy framework capable of revealing, and reliably interpreting, potential signatures of new physics.

We believe that the initial conditions for structure formation were set by a period of accelerated expansion in the first  $\sim 10^{-32}$  seconds in the Universe, known as inflation. After inflation, the Universe underwent a phase of reheating and became nearly homogeneous, filled with particles described by the Standard Model of particle physics, as well as dark matter. Inflation generated small density fluctuations, which, in the absence of primordial non-Gaussianities (PNGs), are statistically described by a Gaussian random field. The peaks of this field mark regions of higher density, which grow over time as matter flows in under the influence of gravity, forming increasingly deep potential wells. When the density in such a region crosses a critical threshold, it decouples from the overall expansion and collapses to form a gravitationally bound object known as a dark matter halo. These halos emerge early in the history of the Universe from the dark matter component, while baryons are still tightly coupled to photons. Only after recombination, roughly 400,000 years after the end inflation, do baryons decouple from photons and begin to fall into the gravitational wells seeded by dark matter, allowing gas to cool and the first galaxies to form.

A key feature imprinted in the matter and galaxy distribution is the baryon acoustic oscillation (BAO) scale. In the early Universe, sound waves propagated through the photon-baryon fluid, generating a characteristic scale in the distribution of matter after recombination. This leads to a subtle, yet measurable, excess in the clustering of galaxies at a preferred separation of about 150 Mpc. The BAO feature appears as a distinct bump in the correlation function and a series of oscillations (“wiggles”) in the power spectrum. It serves as a standard ruler for measuring distances and provides a useful observational tool for probing the expansion history of the Universe.

In the coming years, a vast amount of data will be collected by a new generation of

surveys that aim to map the large-scale structure (LSS) of the Universe with unprecedented precision. The LSS refers to the distribution of galaxies and other extragalactic objects over distances larger than  $\sim 10^2$  Mpc, forming a “cosmic web” pattern that encodes valuable information about the underlying physics that influenced its formation. Among these efforts, spectroscopic surveys will play a particularly central role and are the focus of this thesis. By measuring the redshift of galaxies through their spectra, spectroscopic surveys reconstruct their three-dimensional positions, enabling precise maps of the galaxy distribution. The analysis of such data falls under the domain of *galaxy clustering*. Upcoming spectroscopic experiments include DESI [8, 82], Euclid [40, 43], and PFS [209], which have improved spectral resolution, wider sky coverage, and significantly larger sample size compared to its predecessor BOSS [63]. Together, these surveys promise to enhance our understanding of fundamental physics by placing tighter constraints on the cosmological parameters.

The standard model of cosmology, known as  $\Lambda$ CDM, describes the Universe using six cosmological parameters that characterize its composition and evolution. These include the densities of baryons and cold dark matter, for example. The model also assumes a spatially flat Universe dominated at late times by a cosmological constant  $\Lambda$ , which drives accelerated expansion. While  $\Lambda$ CDM successfully explains a wide range of observations, several of its components point to unresolved questions in fundamental physics. The nature of dark matter and dark energy remains unknown, the small observed value of  $\Lambda$  poses a fine-tuning problem, and tensions between independent measurements of cosmological parameters suggest either new physics or unknown systematics. Thus, precision measurements of these parameters not only test the consistency of the model but also offer a window into possible extensions beyond  $\Lambda$ CDM.

One key cosmological parameter is  $\sigma_8$ , the root-mean-square amplitude of linear matter fluctuations on  $8 h^{-1}$ Mpc scales. Accurate measurements of  $\sigma_8$  are essential for understanding the growth of structure, the properties of dark energy, and possible deviations from General Relativity. Currently, a notable tension has emerged between measurements of  $\sigma_8$  from different cosmological probes within the standard  $\Lambda$ CDM framework.

The cosmic microwave background (CMB), the photons we detect today from the last-scattering surface, provides a precise measurement of the initial conditions for structure formation. High-resolution observations from missions such as *Planck* [9], ACT [10, 139], and SPT [26] infer the value of  $\sigma_8$  by modeling how these initial perturbations evolve under gravity. On the other hand, weak gravitational lensing (the distortion of the shapes of background galaxies due to the gravitational influence of foreground mass) offers a direct probe of the matter distribution at late times. Recent lensing measurements from surveys such as KiDS [97], DES [1], and HSC [2, 119, 208] yield consistently lower values of  $\sigma_8$  compared to those inferred from the CMB.

This discrepancy has sparked considerable interest, as it could hint at new physics beyond  $\Lambda$ CDM. It remains an open question whether the tension originates from a genuine mismatch between early- and late-time cosmology [132, 155, 176] or from scale-dependent systematic effects [19, 178, 184]. Independent constraints on  $\sigma_8$  from galaxy clustering provide a crucial complementary probe to help disentangle these possibilities, and thereby

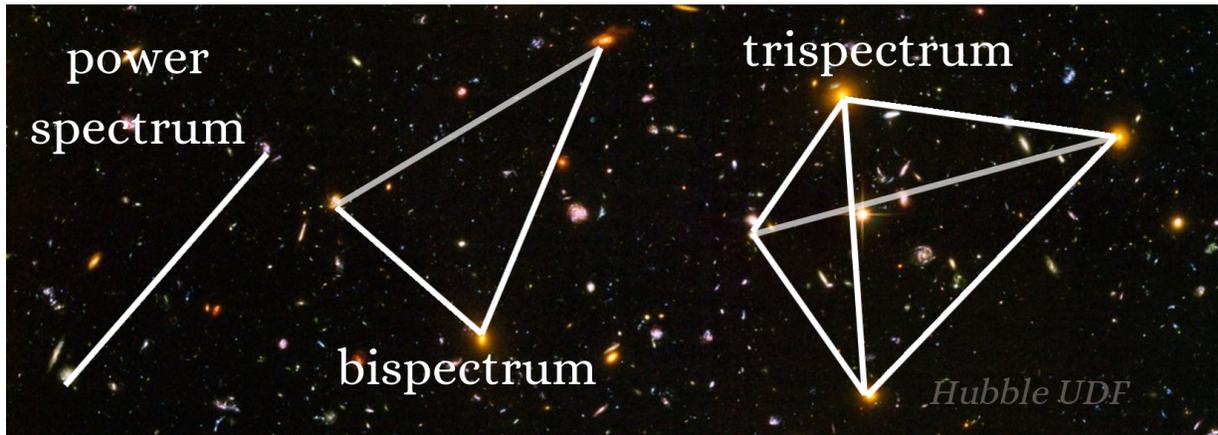


Figure 1.1: Schematic visualization of galaxy  $n$ -point functions (power-spectrum, bispectrum and trispectrum) on Hubble Ultra Deep Field (UDF). Note that galaxies are positioned on the vertices of each  $n$ -point function, although stars could be located at the vertices in this image instead. Furthermore, galaxy clustering analysis are focused on spectroscopic galaxy surveys instead of Hubble UDF. These surveys typically map the positions of millions of galaxies over 3D volumes extending billions of light years. *Credit for Hubble UDF: NASA, ESA, and S. Beckwith (STScI) and the HUDF Team.*

play a crucial role in testing the consistency of the current cosmological model.

Furthermore, the LSS encodes the growth history of matter perturbations during the epoch when dark energy begins to dominate the energy budget of the Universe, providing a unique opportunity to investigate the physics behind the observed accelerated expansion. This is intimately connected to the long-standing cosmological constant problem [6], the enormous discrepancy between the theoretically predicted and observed values of vacuum energy, which is considered by many to be one of the greatest unsolved problems in theoretical physics. While it is unlikely that cosmological observations alone will fully resolve this problem, obtaining a precise and robust characterization of the accelerated expansion remains essential to guide theoretical progress.

**Galaxy  $n$ -point functions.** Focusing on the case of a galaxy catalog from a certain cosmological survey, it is standard to compress the information contained in this catalog and infer the cosmological parameters using summary statistics as the data vector, such as only the BAO “wiggles” in the galaxy power spectrum or its “full-shape” [4,5]. LSS contains rich 3D information, offering greater constraining power on cosmological parameters than the CMB. While the power-spectrum is sufficient for most of CMB analysis due to its predominately Gaussian nature, LSS is inherently more non-Gaussian due to gravitational evolution, which requires going beyond traditional analyses based on the power spectrum.

The first natural question then is how to perform the galaxy field compression, in order to maximize the amount of information extracted. Many alternatives have been proposed in the literature to go beyond two-point statistics, such as higher  $n$ -point functions (bispec-

trum and trispectrum – see Fig. 1.1), Minkowski functionals, wavelets, k-nearest neighbor cumulative distribution functions, peak counts, void statistics, among many others (see, e.g., [74, 166, 171]). Of course, a given cosmological parameter, such as  $\sigma_8$  or dark energy equation of state  $w$ , affects observables differently, so that the optimal compression choice is highly dependent on the parameter to be inferred.

Throughout this thesis, we defer the difficult question of optimal compression, and instead focus on the galaxy  $n$ -point functions, namely the galaxy power spectrum, bispectrum and trispectrum. These summaries are well motivated on quasilinear scales by perturbative approaches to LSS and have been extensively investigated in the literature. They offer a systematic approach to probing the information content in galaxy clustering, with higher-order statistics expected to progressively capture more cosmological information.

**Simulation-based inference.** Having fixed the definition of the data vector, the next problem consists in predicting the distribution of each of the summary statistics. The current standard cosmological inference procedure consists in providing an analytical *likelihood* for the data vector considered, which together with parameter priors and sampling methods such as Monte-Carlo Markov Chain (MCMC) yields the posteriors of the parameters of interest given the observed data [215, 220] (see details in Sec. 3.1 and Sec. 3.2). For essentially all relevant LSS observables, only approximate expressions for these distributions are known. Even if an exact likelihood expression could be written down and used for inference, there would remain additional issues with this standard explicit-likelihood approach which we discuss in Sec. 3.3.

These issues become progressively more challenging as we move to higher-order statistics. For example, obtaining cosmological constraints using the full-shape of the galaxy trispectrum has never been achieved due to the complexities inherent in traditional analysis. Although modeling, analyzing, and combining these complex statistics is challenging, in this thesis I demonstrate how to tackle these obstacles using simulation-based inference (SBI) [57, 137]. SBI is a deep learning approach that learns the full distribution of data vectors from simulations, bypassing the need for conventional analytical likelihoods or covariance matrices. We explain in detail how this algorithm works in Sec. 3.4. In essence, SBI fits the posterior distribution by transforming complexly distributed data into a tractable form, enabling efficient sampling and accurate estimation of the cosmological parameters.

Using the true distribution of data vectors is critical for robust analysis, as traditional likelihoods often serve as mere approximations of the true data distribution, which can introduce bias and inaccuracies in cosmological inference. When evaluating discoveries of new physics, it is essential that our error bars are well-calibrated to avoid drawing misleading conclusions. For instance, miscalculated covariance can result in overly tight error bars, potentially leading to false claims of new physics. By eliminating the need for an explicit likelihood and instead focusing on the generative processes of physics to create simulations, we can move away from the analytical approximations that accompany likelihoods of data vectors. Therefore, SBI offers a direct pathway to achieve more reliable cosmological constraints.

---

## Outline of the results

**Perturbative techniques for galaxy clustering.** The large-scale distribution of galaxies, dark matter halos, and quasars traces the underlying matter density field, but not in a one-to-one manner. These objects are referred to as *biased tracers* of large-scale structure, meaning their spatial distribution is a biased reflection of the total matter field [110]. The statistical relationship between tracers and the matter field is described by the *bias expansion*, which can be applied to any tracer population. Throughout this thesis, we use the term “galaxy” generically to refer to any biased tracer, including dark matter halos, unless otherwise stated.

In Chapter 2, we introduce the perturbative techniques developed over the past decades that will be used throughout this thesis. Given the complexities and uncertainties of galaxy formation on small scales, we adopt the galaxy bias expansion, an effective field theory (EFT) approach to model its impact on galaxy clustering without requiring full modeling of the unknown physics. By parametrizing our ignorance of small-scale physics through a controlled set of free coefficients, the bias expansion offers a robust and flexible foundation for cosmological inference. For a comprehensive review, we refer the reader to [66]. This framework provides a theoretically robust connection between galaxy clustering observables, such as the  $n$ -point functions, and the cosmological parameters.

In particular, in Sec. 2.5, we conclude with a detailed explanation of how to use these perturbative tools to build `LEFTfield`, a heavily parallelized forward model that evolves possible initial conditions of the Universe into the late-time 3D distribution of galaxies within seconds. Part of my work has been devoted to co-developing `LEFTfield`, such as the implementation of the galaxy bispectrum and trispectrum estimators to enable cosmological inference using SBI. Since it is entirely based on the bias expansion, it provides a robust theoretical foundation for modeling the large-scale distribution of galaxies, while being orders of magnitude faster to evaluate than traditional N-body simulations. It is worthy noting here that, although galaxies are observed in redshift space, in this work we restrict our analysis to restframe and do not incorporate observational systematics or survey masks.

**Cosmological inference.** In Chapter 3, review Bayesian inference and standard techniques used in cosmological inference. We then motivate the use of SBI for galaxy clustering and explain in detail the novel inference methods that will be used throughout this thesis: simulation-based inference (SBI) and field-level inference (see below). We also review data compression schemes for optimizing SBI, which will be employed in Appendix C.

**SBI for power-spectrum and bispectrum.** In Chapter 4, we demonstrate how to perform SBI with `LEFTfield`, while allowing for cosmological volumes analysis, rigorous posterior diagnostics, and convergence tests. In fact, the work presented in this thesis is unique in the literature in performing SBI within the framework of the bias expansion. My first work in this context (a first-author publication in JCAP [217]) used both the

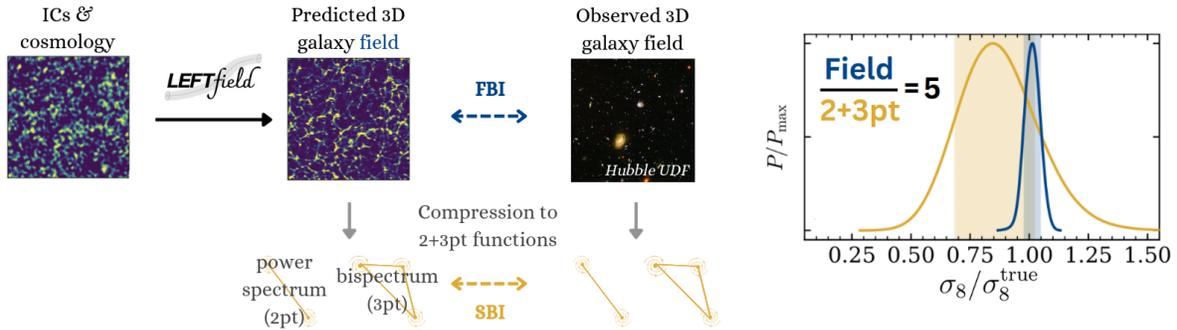


Figure 1.2: *Left*: Diagram showing the comparison between FBI, where the full galaxy density field is used for inference using a field-level likelihood, and SBI, where the galaxy power-spectrum and bispectrum (two- and three- point functions) are used. Both inference methods are based on `LEFTfield`, which evolves the Universe initial conditions (ICs) to the final 3D distribution of galaxies given cosmological parameters. *Right*: comparison of constraints of the cosmological parameter  $\sigma_8$  for one particular dark-matter halo and scale cut considered in the analysis, where the galaxy density field shows to provide 5 times more information than the combined two- and three-point functions. *Credit for Hubble UDF: NASA, ESA, and S. Beckwith (STScI) and the HUDF Team.*

power spectrum and bispectrum on Euclid-like mock data, resulting in successful analysis and demonstrating that we need around  $10^5$  simulations for convergence. I am currently working to improve this number with more advanced SBI techniques, and I provide a few examples in Appendix C. Since our forward model is based on the perturbative bias expansion, I introduced model mismatch by analyzing dark-matter halos from N-body simulations as the first proxy of real data, successfully recovering accurate cosmological constraints despite the increased complexity [157]. These results are presented in Chapter 5.

**Galaxy density field information content.** Beyond high-order statistics, the next frontier in cosmological analysis is field-level Bayesian inference (FBI), where we directly analyze the galaxy density field instead of compressing it to summary statistics (a detailed explanation of how it works is given in Sec. 3.6). FBI is extremely challenging, as it involves working with millions of parameters corresponding to the voxels of the galaxy density grid, besides requiring a theoretically robust model for the galaxy distribution since FBI is a very sensitive analysis. This requires using sampling techniques suitable for high dimensions and a sophisticated field-level likelihood (derived in Sec. 2.4.4), which can be computationally expensive. Recently, exciting developments are being made towards field-level inference [23, 105, 115, 118, 120, 156, 173, 192, 224].

In Chapter 5, it is discussed one key result from this work using SBI, which was to demonstrate that the galaxy density field contains up to five times more information on  $\sigma_8$  than what can be extracted from the power spectrum and bispectrum [157], as illustrated in Fig. 1.2. This finding translates into a significant increase in the effective survey volume.

In fact, this was the first successful field-level inference of cosmological parameters and initial conditions presented in the literature. The analysis directly compared the same dark-matter halos as data, the same forward model `LEFTfield`, and identical scale cuts. We utilized two dark-matter halo samples varying in number density and redshift, revealing differences in the information gained from each sample and scale cut.

In Appendix E, we provide a discussion on the main findings of two papers I co-authored on field-level inference of the BAO scale, highlighting the advantages of FBI in extracting cosmological information. Since these results have already been discussed in Ivana Babić’s dissertation thesis, the first author, I briefly discuss the main results and my contributions here. In our first paper [23], we demonstrated that FBI can yield more information than traditional power spectrum analysis using fixed initial conditions. In the second paper [24], we extended our analysis to sample initial conditions in FBI, and found that FBI provides greater information content not only compared to the power spectrum but also the reconstructed power spectrum. Decades of cosmologists and generations of surveys have led to the consensus that the reconstructed power spectrum is the optimal BAO analysis; however, we were the first to show that the galaxy density field can carry more cosmological information.

**SBI for power-spectrum, bispectrum and trispectrum.** In Chapter 6, the first results from a cosmological analysis that jointly utilizes the galaxy power spectrum, bispectrum, and trispectrum are presented — an approach deemed infeasible due to the dimensionality of the trispectrum. In principle, SBI combined with `LEFTfield` should provide a more straightforward route than traditional analysis, since we do not need to explicitly estimate and invert its high dimensional covariance, and instead simply used the trispectrum measurements from `LEFTfield` to conduct cosmological inference with SBI. In practice, SBI with the galaxy trispectrum introduces its own challenges, as discussed in that chapter. Although this is a work in progress, we present our current results along with future plans, given the complexity of the problem. This will allow for the first demonstration of the cosmological information encoded in the galaxy trispectrum on  $\sigma_8$ , besides demonstrating the additional information gained from the trispectrum when compared to the galaxy density field.

We conclude in Chapter 7, providing an overview of the results obtained in this work and outlining the next steps to be taken in this line of research. We demonstrate that the methods used in this work offer a promising path toward achieving tighter and more reliable constraints, potentially reducing statistical uncertainties while preserving theoretical rigor. These represent important next steps toward applying these techniques to real data. Ultimately, such developments pave the way for more precise and robust analyses of galaxy clustering, with the potential to uncover signatures of new physics.



# Chapter 2

## Galaxy clustering

The large-scale distribution of galaxies in the Universe encodes valuable information about fundamental physics, including gravity and the expansion history. To extract this information, it is essential to understand how the initially small density fluctuations present after recombination evolved under gravity to form the complex large-scale structure (LSS) we observe today. This is a central question in cosmology, as galaxy clustering arises from a highly non-linear and complex process. Several approaches can be employed to tackle this challenge. Hydrodynamical simulations can model galaxy formation [154, 222], but they are dependent on specific prescriptions that we know are incomplete. Ideally, we would marginalize over all possible models of galaxy formation, but the computational cost of running these simulations is prohibitive, and current models do not fully reproduce observations. Alternatively, N-body dark-matter-only simulations are comparably less computationally expensive [160, 203], allowing galaxies to be painted into dark-matter halos using halo occupation distribution (HOD) models [34]. However, these models rely on highly simplified treatment of small-scale physics, for which we lack accurate theoretical descriptions.

The most robust approach, therefore, is to abandon explicitly modeling small scales and instead focus on the Effective Field Theory of the Large-Scale Structure (EFTofLSS). This framework allows us to extract robust cosmological information by employing perturbative techniques where density fluctuations are small and well understood. By applying general principles such as the equivalence principle from General Relativity, we construct a bias expansion that includes all operators allowed by symmetry [66], each with an associated free bias parameter that must be constrained by data. By marginalizing over these bias parameters, we are accounting for the marginalization over the reasonable unknown models of galaxy formation on small scales.

This chapter develops the theoretical framework needed to model galaxy clustering in the framework of the EFTofLSS, beginning with a review of key aspects of General Relativity (GR) that will be important for later discussions (Sec. 2.1). Building on this, we explore how to use perturbation theory to model the nonlinear growth of structure, both in Eulerian and Lagrangian frames, and how to construct the EFTofLSS (Sec. 2.2). We then discuss how to construct the bias expansion to model the distribution of galaxies,

connecting theoretical predictions to actual observations (Sec. 2.3). We also introduce Wilson-Polchinski renormalization scheme for galaxies (Sec. 2.4), from which we will derive important results that set the foundations of `LEFTfield`, the forward model that will be used throughout this thesis (Sec. 2.5). It is solely built upon the perturbative techniques discussed in this chapter, and will allow us to relate observable quantities to the cosmological parameters and initial conditions, setting the stage for cosmological inference (Chapter 3).

## 2.1 Gravity

In General Relativity (GR), physical observables must be independent of coordinate choices. This requirement implies that the only meaningful quantities are those that are invariant under diffeomorphisms. Physical observables must be constructed from quantities that remain invariant under coordinate transformations  $x^\mu \rightarrow x'^\mu(x)$ . The metric itself transforms as [50]

$$g'_{\mu\nu}(x') = \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta}(x). \quad (2.1)$$

Since this transformation depends on arbitrary coordinate choices, the metric  $g_{\mu\nu}$  itself is not directly observable. The Christoffel symbols,

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2} g^{\lambda\sigma} (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}), \quad (2.2)$$

describe the connection on spacetime but are not tensorial quantities; they change in a non-tensorial way under coordinate transformations. Therefore, they are not observable. The first truly observable quantity constructed from the metric is the Riemann curvature tensor,

$$R_{\sigma\mu\nu}^\rho = \partial_\mu \Gamma_{\nu\sigma}^\rho - \partial_\nu \Gamma_{\mu\sigma}^\rho + \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\sigma}^\lambda - \Gamma_{\nu\lambda}^\rho \Gamma_{\mu\sigma}^\lambda. \quad (2.3)$$

This tensor is invariant under coordinate transformations and contains all the information about spacetime curvature. It depends explicitly on the second derivatives of the metric tensor. Only the second derivatives of the metric, appearing in the Riemann tensor, provide truly invariant information about spacetime. This explains why all physical observables in GR, such as the Einstein tensor,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R, \quad (2.4)$$

and the Weyl tensor,

$$C_{\sigma\mu\nu}^\rho = R_{\sigma\mu\nu}^\rho - \frac{1}{2} (g_\mu^\rho R_{\sigma\nu} - g_\nu^\rho R_{\sigma\mu} + g_{\sigma\nu} R_\mu^\rho - g_{\sigma\mu} R_\nu^\rho) + \frac{1}{6} R g_\sigma^\rho g_{\mu\nu}, \quad (2.5)$$

are constructed from the second derivatives of the metric. All physical effects, such as gravitational waves, tidal forces, and spacetime curvature, are described by the Riemann

tensor and its contractions. This fundamental principle ensures that all measurements, whether using geodesic deviation, gravitational lensing, or the detection of gravitational waves, ultimately depend on the second derivative of the metric.

The Einstein field equations describe the relationship between the geometry of spacetime and the energy-momentum content,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu}. \quad (2.6)$$

This equation is *local* because it relates the Einstein tensor  $G_{\mu\nu}$  to the stress-energy tensor  $T_{\mu\nu}$  at the same point in spacetime. That is, at any given point, the curvature is determined solely by the energy and momentum present there.

While the Einstein equations are local, their solutions generally are *not*. This is because the field equations are differential equations whose solutions depend on boundary conditions and global properties of spacetime. The classic example is the Newtonian gravitational potential  $\Phi$  in the weak-field limit:

$$\nabla^2\Phi = 4\pi G\rho. \quad (2.7)$$

The solution to this equation involves  $\nabla^{-2}$  and therefore an integral over all mass sources:

$$\Phi(\mathbf{x}) = -G \int \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} d^3x'. \quad (2.8)$$

This shows that the gravitational potential at a given point depends on the entire distribution of matter, not just on local properties.

In General Relativity, this non-locality manifests in the full metric solution  $g_{\mu\nu}$ , which is influenced by the global distribution of mass-energy. A prime example is the *Schwarzschild solution*, which describes the external gravitational field of a spherically symmetric mass, but remains valid even far from the source.

While the Ricci tensor and the stress-energy tensor are related via Einstein equations, it seems that the Weyl tensor, which is the traceless part of the Riemann tensor, should be unconstrained. However, the components of the curvature tensor should be related by the Bianchi identity,

$$\nabla_{[\lambda}R_{\rho\sigma]\mu\nu} = 0, \quad (2.9)$$

which provides a relation for the (derivatives) of the Weyl tensor and the Ricci tensor. Since the latter can be related to  $T_{\mu\nu}$  via Einstein equations, we have that [50]

$$\nabla^\rho C_{\rho\sigma\mu\nu} = 8\pi G \left( \nabla_{[\mu}T_{\nu]\sigma} - \frac{1}{3}g_{\sigma[\mu}\nabla_{\nu]}T \right), \quad (2.10)$$

where we see that the Weyl tensor is related to the stress-energy tensor via a first-order differential equation. This can be thought of as an analogy of Maxwell's equations,  $\nabla_\mu F^{\nu\mu} = J^\nu$ , where on the right hand side we have a source term for the propagation of gravitational waves. Note however that, given a energy-momentum distribution, we

can have multiple solutions, as it should be. For example, for a spacetime where we have vacuum everywhere, we have that  $R_{\mu\nu} = 0$ . Both flat Minkowski spacetime and gravitational waves propagation through vacuum are a solution in this case, specified by different boundary conditions. This implies that the Weyl tensor propagates freely, much like the electromagnetic field in vacuum, carrying information about distant sources. In astrophysical settings, anisotropic distributions of matter around an object can source nonzero Weyl curvature, leading to observable tidal effects.

## 2.2 Perturbation Theory for LSS

In this section, we review the perturbative approach to describing the nonlinear gravitational evolution of dark matter density and velocity fields. On sufficiently large scales, where perturbative bias expansion is valid, vorticity as well as baryonic pressure effects are negligible, and the nonlinear evolution of the cosmic matter density field can be accurately modeled as a self-gravitating, pressureless, irrotational single fluid. We will neglect the aforementioned effects throughout this discussion (see EFTofLSS for their treatment, Sec. 2.2.5).

### 2.2.1 Gravitational fluid

Over a Hubble time,  $T = \mathcal{H}^{-1}$ , both matter and galaxies move distances no larger than the non-linear scale  $k_{\text{NL}}$ . This limited displacement is not due to strong interactions, as would be the case in a conventional fluid, but rather because peculiar velocities are small and the universe’s finite age constrains how far structures can evolve.

A key point in this framework is the distinction between the Hubble scale and the non-linear scale. Gravity, by setting a finite causal horizon and a finite time, naturally introduces a hierarchy among the multipole moments of the distribution function. This hierarchy is essential for the emergence of a fluid description: without gravity, the system could be arbitrarily old, no preferred scale would exist, and a coherent fluid-like approximation would break down. Thus, when we integrate out short-wavelength non-linearities in an expanding FRW universe, the effective dynamics at large scales resemble those of a “gravitational fluid”. This emphasizes the fundamental role of gravity not only in clustering matter but also in enabling a controlled effective field theory description of LSS [31].

We motivate them below in Lagrangian Perturbation Theory (LPT) and Eulerian Perturbation Theory (EPT) before proceeding to the general bias operators construction.

### 2.2.2 EPT

### 2.2.3 The Fluid Equations

The distribution of particles in phase space is described by the distribution function  $f(\mathbf{x}, \mathbf{p}, \tau)$ . According to Liouville’s theorem, the phase space density is conserved along

particle trajectories. For collisionless dark matter, this leads to the collisionless Boltzmann equation, also known as the Vlasov equation

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{d\mathbf{x}}{d\tau} \frac{\partial f}{\partial \mathbf{x}} + \frac{d\mathbf{p}}{d\tau} \frac{\partial f}{\partial \mathbf{p}} = 0, \quad (2.11)$$

Since solving the full seven-dimensional Vlasov equation is highly challenging, we focus instead on moments of the distribution function, which correspond to fluid quantities such as density  $\rho(\mathbf{x}, \tau)$ , mean velocity  $v_i(\mathbf{x}, \tau)$ , and anisotropic stress  $\tau_{ij}(\mathbf{x}, \tau)$ , capturing deviations from a perfect single-stream flow.

To derive the evolution equations for the fluid quantities, we take moments of the Vlasov equation (2.11). The zeroth moment corresponds to the continuity equation. After integrating over momentum space and performing an integration by parts, we obtain

$$\frac{\partial}{\partial \tau} \delta(\mathbf{x}, \tau) + \nabla \cdot \{[1 + \delta(\mathbf{x}, \tau)]\mathbf{v}(\mathbf{x}, \tau)\} = 0, \quad (2.12)$$

where  $\delta = \rho/\bar{\rho} - 1$  is the density contrast. Taking the first moment and using the continuity equation, we obtain the Euler equation, or momentum conservation,

$$\frac{\partial}{\partial \tau} \mathbf{v}(\mathbf{x}, \tau) + [\mathbf{v}(\mathbf{x}, \tau) \cdot \nabla] \mathbf{v}(\mathbf{x}, \tau) + \mathcal{H}(\tau) \mathbf{v}(\mathbf{x}, \tau) = -\nabla \Phi(\mathbf{x}, \tau) - \frac{1}{\rho(\mathbf{x}, \tau)} \nabla \tau_{ij}(\mathbf{x}, \tau), \quad (2.13)$$

where  $\mathcal{H}(\tau)$  denotes the conformal Hubble parameter and  $\tau_{ij}$  again captures the effects of velocity dispersion.

The density also obeys Poisson equation from the Einstein equations,

$$\nabla^2 \Phi(\mathbf{x}, \tau) = \frac{3}{2} \mathcal{H}^2(\tau) \Omega_m(\tau) \delta(\mathbf{x}, \tau), \quad (2.14)$$

where  $\Omega_m(\tau)$  is the matter density parameter. The fluid velocity can be decomposed into a scalar and a vector part as  $\mathbf{v} = \mathbf{v}_k + \mathbf{v}_\perp$ , where  $\nabla \times \mathbf{v}_k = 0$  and  $\nabla \cdot \mathbf{v}_\perp = 0$ . Thus, the velocity field can be fully described by its vorticity  $\mathbf{w} = \nabla \times \mathbf{v}$  and its divergence  $\theta = \nabla \cdot \mathbf{v}$ .

## 2.2.4 SPT

To truncate this hierarchy, we assume that all moments beyond the velocity dispersion vanish, an approximation known as the *pressureless perfect fluid* assumption. In essence, in this approximation we will assume that the anisotropic stress vanishes. In Fourier space, one can eliminate the gravitational potential  $\Phi$  and describe the dynamics in terms of the velocity divergence  $\theta = \nabla \cdot \mathbf{v}$ . At linear order, the vorticity can be neglected (although it can be generated at higher orders), and the continuity equation (Eq. 2.12) becomes

$$\frac{\partial}{\partial \tau} \delta^{(1)}(\mathbf{x}, \tau) + \theta^{(1)}(\mathbf{x}, \tau) = 0, \quad (2.15)$$

while the Euler equation (Eq. 2.13) reads

$$\frac{\partial}{\partial \tau} \theta^{(1)}(\mathbf{x}, \tau) + \mathcal{H}(\tau) \theta^{(1)}(\mathbf{x}, \tau) = -\nabla^2 \Phi^{(1)}(\mathbf{x}, \tau) = -\frac{3}{2} \mathcal{H}^2(\tau) \Omega_m(\tau) \delta^{(1)}(\mathbf{x}, \tau), \quad (2.16)$$

where we used the Poisson equation (Eq. 2.14). Taking the time derivative of Eq. 2.15 and using Eq. 2.16, we obtain

$$\frac{\partial^2}{\partial \tau^2} \delta^{(1)}(\mathbf{k}, \tau) - \mathcal{H}(\tau) \theta^{(1)}(\mathbf{x}, \tau) - \frac{3}{2} \mathcal{H}^2(\tau) \Omega_m(\tau) \delta^{(1)}(\mathbf{x}, \tau) = 0, \quad (2.17)$$

where we can use Eq. 2.15 to obtain

$$\frac{\partial^2}{\partial \tau^2} \delta^{(1)}(\mathbf{k}, \tau) + \mathcal{H}(\tau) \frac{\partial}{\partial \tau} \delta^{(1)}(\mathbf{k}, \tau) - \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) \delta^{(1)}(\mathbf{k}, \tau) = 0. \quad (2.18)$$

The growth factor  $D(\tau)$  can be introduced by expressing the linear density contrast as

$$\delta^{(1)}(k, \tau) = \left[ \frac{D(\tau)}{D(\tau_0)} \right] \delta^{(1)}(k, \tau_0), \quad (2.19)$$

where  $\tau_0$  is a reference time and  $D(\tau)$  satisfies the differential equation

$$\frac{d^2 D(\tau)}{d\tau^2} + \mathcal{H} \frac{dD(\tau)}{d\tau} - \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) D(\tau) = 0, \quad (2.20)$$

where initial conditions are chosen so that  $D(\tau)$  follows the growing mode, typically scaling as  $D(\tau) \propto a(\tau)$  during matter domination. To move to a general solution, we note that in Fourier space

$$\mathbf{v}(\mathbf{k}) = -i \frac{\mathbf{k}}{k^2} \theta(\mathbf{k}), \quad (2.21)$$

to write the non-linear terms in Eq. (2.12) and Eq. (2.13) in the right-hand side of

$$\frac{\partial \delta(\mathbf{k}, \tau)}{\partial \tau} + \theta(\mathbf{k}, \tau) = - \int_{\mathbf{k}_1, \mathbf{k}_2} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}_{12}) \alpha(\mathbf{k}_1, \mathbf{k}_2) \theta(\mathbf{k}_1, \tau) \delta(\mathbf{k}_2, \tau), \quad (2.22)$$

$$\begin{aligned} \frac{\partial \theta(\mathbf{k}, \tau)}{\partial \tau} + \mathcal{H}(\tau) \theta(\mathbf{k}, \tau) + \frac{3}{2} \mathcal{H}^2(\tau) \Omega_m(\tau) \delta(\mathbf{k}, \tau) = \\ - \int_{\mathbf{k}_1, \mathbf{k}_2} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}_{12}) \beta(\mathbf{k}_1, \mathbf{k}_2) \theta(\mathbf{k}_1, \tau) \theta(\mathbf{k}_2, \tau), \end{aligned} \quad (2.23)$$

where we have introduced the shorthand  $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$  and the mode-coupling kernels

$$\alpha(\mathbf{k}_1, \mathbf{k}_2) = \frac{\mathbf{k}_{12} \cdot \mathbf{k}_1}{k_1^2}, \quad (2.24)$$

$$\beta(\mathbf{k}_1, \mathbf{k}_2) = \frac{k_{12}^2 (\mathbf{k}_1 \cdot \mathbf{k}_2)}{2k_1^2 k_2^2}. \quad (2.25)$$

In standard perturbation theory (SPT), the fields  $\delta$  and  $\theta$  are expanded perturbatively in powers of the linear density contrast  $\delta^{(1)}$ :

$$\delta(\mathbf{k}, \tau) = \sum_{n=1}^{\infty} \delta^{(n)}(\mathbf{k}, \tau), \quad (2.26)$$

$$\theta(\mathbf{k}, \tau) = \sum_{n=1}^{\infty} \theta^{(n)}(\mathbf{k}, \tau), \quad (2.27)$$

where  $\delta^{(n)}$  and  $\theta^{(n)}$  represent contributions of order  $n$  in  $\delta^{(1)}$ . The linear-theory velocity divergence field can then be obtained from the linearized continuity equation [Eq. (B.4)]:

$$\theta^{(1)}(k, \tau) = -\frac{d \ln D(\tau)}{d\tau} \delta^{(1)}(k, \tau) = -\mathcal{H}f(\tau) \delta^{(1)}(k, \tau), \quad (2.28)$$

where the logarithmic growth rate is defined as  $f(\tau) \equiv d \ln D / d \ln a$ . By looking at the structure of Eq. (2.22) and Eq. (2.23), we can infer a structure for the  $n$ -th order solutions,

$$\delta^{(n)}(\mathbf{k}, \tau) = \int_{\mathbf{k}_1} \cdots \int_{\mathbf{k}_n} (2\pi)^3 \delta_D(k - \mathbf{k}_{12\dots n}) F_n(\mathbf{k}_1, \cdots, \mathbf{k}_n; ) \delta^{(1)}(\mathbf{k}_1, \tau) \cdots \delta^{(1)}(\mathbf{k}_n, \tau), \quad (2.29)$$

$$\theta^{(n)}(\mathbf{k}, \tau) = -\mathcal{H}(\tau) f(\tau) \int_{\mathbf{k}_1, \dots, \mathbf{k}_n} (2\pi)^3 \delta_D(k - \mathbf{k}_{12\dots n}) G_n(\mathbf{k}_1, \cdots, \mathbf{k}_n) \delta^{(1)}(\mathbf{k}_1, \tau) \cdots \delta^{(1)}(\mathbf{k}_n, \tau), \quad (2.30)$$

where  $F_n$  and  $G_n$  are symmetrized kernels associated with the density and velocity divergence fields, respectively. From the linear solutions,  $F_1 = G_1 = 1$ . We have assumed an Einstein-de Sitter (EdS) universe ( $\Omega_m = f = 1$ ), where the kernels become time-independent, simplifying the calculations. To find the second order kernels, we rewrite. Using the power law ansatz and comparing to Eq. 2.30, we obtain

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{5}{7} + \frac{2}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right), \quad (2.31)$$

$$G_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{3}{7} + \frac{4}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right). \quad (2.32)$$

Rewriting  $F_2$  as

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{17}{21} + \frac{2}{7} \left[ \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} - \frac{1}{3} \right] + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right), \quad (2.33)$$

it becomes clear that the second-order density can also be written in real space as

$$\delta^{(2)}(\mathbf{x}, \tau) = \frac{17}{21} [\delta^{(1)}(\mathbf{x}, \tau)]^2 + \frac{2}{7} [K_{ij}^{(1)}(\mathbf{x}, \tau)]^2 - s_i^{(1)}(\mathbf{x}, \tau) \partial_i \delta^{(1)}(\mathbf{x}, \tau). \quad (2.34)$$

where  $K_{ij}^{(1)}$  is the tidal tensor constructed from the first-order density field, and  $s_i^{(1)}$  denotes the displacement field (see Sec. 2.2.6),

$$\mathbf{s}^{(1)}(\mathbf{q}, \tau) = \mathbf{x}_{\text{fl}}^{(1)}(\mathbf{q}, \tau) - \mathbf{q}, \quad (2.35)$$

where  $\mathbf{x}_\Pi(\tau)$  denotes the Eulerian coordinate of the fluid trajectory corresponding to a fixed Lagrangian position  $\mathbf{q} = \mathbf{x}_\Pi(\tau = 0)$ .

We can use the solutions that we obtained to calculate the prediction for the matter power-spectrum. The non-trivial components are given by

$$P_{\text{SPT}}(k) = \langle \delta^{(2)}(\mathbf{k}) \delta^{(2)}(\mathbf{k}') \rangle' + \langle \delta^{(1)}(\mathbf{k}) \delta^{(3)}(\mathbf{k}') \rangle' = P^{(22)}(k) + P^{(13)}(k), \quad (2.36)$$

where

$$P^{(22)}(k) = 2 \int_{\mathbf{p}} [F_2(\mathbf{p}, \mathbf{k} - \mathbf{p})]^2 P_L(p) P_L(|\mathbf{k} - \mathbf{p}|) \quad (2.37)$$

$$P^{(13)}(k) = 3P_L(k) \int_{\mathbf{p}} F_3(\mathbf{p}, -\mathbf{p}, \mathbf{k}) P_L(p) \quad (2.38)$$

and  $F_3(\mathbf{p}, -\mathbf{p}, \mathbf{k})$  is one of the third-order kernels.

### 2.2.5 EFTofLSS

In Sec. 2.2.4, we assumed that dark matter behaves as a perfect, pressureless fluid. While this is a valid approximation on large scales, it breaks down on small scales, where pressure and anisotropic stress become non-negligible. In particular, the single-stream approximation and negligible velocity dispersion assumed in SPT fail in the nonlinear regime. As a result, one-loop integrals in SPT receive contributions from arbitrarily high wavenumbers  $k$ , where the fluid description is no longer valid.

Although imposing a sharp cutoff  $k < \Lambda$  can regulate these integrals, the results become sensitive to the arbitrary choice of cutoff, introducing unphysical dependence. Moreover, nonlinear phenomena such as baryonic feedback, galaxy formation, and exotic short-scale dynamics further complicate the modeling of small-scale physics.

This motivates the Effective Field Theory of Large-Scale Structure (EFTofLSS), which systematically incorporates the impact of small-scale nonlinear physics on large-scale observables. The central idea is to filter out the small scales and work instead with smoothed fields  $X = \{\delta_\Lambda, \mathbf{v}_\Lambda, \Phi_\Lambda\}$ ,

$$X_\Lambda(\mathbf{k}, \tau) = W_\Lambda(k) X(k, \tau), \quad (2.39)$$

where  $W_\Lambda$  is a smoothing kernel that filters out modes above a scale  $\Lambda$ .

Even though short-scale modes are integrated out, they influence the evolution of large-scale structures through an effective stress tensor  $\tau^{ij}$  that appears in the Euler equation, where now we have modified the structure of Eq. 2.13,

$$\frac{\partial}{\partial \tau} \mathbf{v}(\mathbf{x}, \tau) + [\mathbf{v}(\mathbf{x}, \tau) \cdot \nabla] \mathbf{v}(\mathbf{x}, \tau) + \mathcal{H}(\tau) \mathbf{v}(\mathbf{x}, \tau) = -\nabla \Phi(\mathbf{x}, \tau) - \frac{1}{\rho(\mathbf{x}, \tau)} \nabla \tau_\Lambda^{ij}(\mathbf{x}, \tau), \quad (2.40)$$

where  $\tau_\Lambda^{ij}$  encodes the backreaction of short-scale physics on the long-wavelength flow.

We can make progress by expanding the effective stress tensor  $[\tau^{ij}]_\Lambda$  in terms of a basis of *counter-terms*, that are constructed out from the leading gravitational observables along

the fluid trajectory. In fact, these are the same type as those employed in the general Eulerian bias expansion in Sec. 2.3.5 ( $\Pi^{[n]}$  there). The counter-terms are free coefficients which are left to be determined by data or simulations, and which encode small-scale physics processes that we do not know how to model. By leaving them as effective free coefficients, we are parametrizing our ignorance about the complicated small-scale physics.

At one-loop order, it is sufficient to retain only the linear term [25],

$$\frac{1}{\rho(\mathbf{x}, \tau)} \nabla \tau_{\Lambda}^{ij}(\mathbf{x}, \tau) = c_s^2(\tau) \partial^i \delta^{(1)}(\mathbf{x}, \tau), \quad (2.41)$$

where  $c_s^2(\tau)$  is a time-dependent counter-term. In practice, by introducing this counterterm structure, we are absorbing ultraviolet-sensitive contributions from the small-scale power that is integrated in the SPT power spectrum in Eq. (2.38). Schematically, the leading one-loop EFT correction to the matter power spectrum in the limit of  $\Lambda^{-1} \rightarrow 0$  at fixed time looks like

$$P_{\text{EFT}}(k) = P_{\text{SPT}}(k, ) - 2c_s^2 k^2 P_L(k, ), \quad (2.42)$$

$P_L(k)$  is the linear power spectrum. Overall, EFTofLSS provides a renormalized and systematic framework that extends the validity of perturbation theory, considerably improving agreement with  $N$ -body simulations. In the galaxy bias section, the concepts of the EFTofLSS presented here will be applied in a similar manner and explained in greater detail.

### 2.2.6 LPT

In order to study the dynamics of matter, we can write the geodesic equation for the Eulerian comoving position  $\mathbf{x}(\tau)$  of a Newtonian pressureless fluid element and relate it to the gravitational potential  $\Phi(\mathbf{x}, \tau)$ . If we denote the Lagrangian position of this fluid at  $\tau = 0$  as  $\mathbf{q}$ , we obtain the *advection equation*

$$\mathbf{x}_{\#}(\mathbf{q}, \tau) = \mathbf{q} + \mathbf{s}(\mathbf{q}, \tau), \quad (2.43)$$

where  $\mathbf{s}(\mathbf{q}, \tau)$  is the Lagrangian displacement along the fluid trajectory (see Fig. 2.1).

In the Lagrangian picture, the displacement  $\mathbf{s}(\mathbf{q}, \tau)$  is our dynamic variable. We can then write the geodesic equation for the displacement and relate it to the gravitational potential  $\Phi(\mathbf{x}, \tau)$ , which in turn is connected to the matter density contrast  $\delta(\mathbf{x}, \tau)$  via the Poisson equation. The continuity equation for matter,

$$[1 + \delta(\mathbf{x}_{\#})] d\mathbf{x} = d\mathbf{q}, \quad (2.44)$$

can be written as

$$1 + \delta(\mathbf{x}, \tau) = \left. \frac{d\mathbf{x}}{d\mathbf{q}} \right|_{\mathbf{x}=\mathbf{x}_{\#}}^{-1} = |\mathbf{1} + \mathbf{M}(\mathbf{q}, \tau)|^{-1}, \quad (2.45)$$

where we defined the Lagrangian deformation tensor  $M_{ij} \equiv \partial s_j / \partial q_i$ , allowing us to write the equation of motion of the displacement in terms of  $\mathbf{q}$  only [143].

Lagrangian Perturbation Theory (LPT) then proceeds by expanding the displacement as [44]

$$\mathbf{s}(\mathbf{q}, \tau) = \sum_{n=1}^{\infty} \mathbf{s}^{(n)}(\mathbf{q}, \tau), \quad (2.46)$$

and equivalently for the deformation tensor to obtain the LPT recursive relations [143, 180, 231]. It is also convenient to separate the displacement into a longitudinal  $\sigma \equiv \nabla \cdot \mathbf{s}$  and a transverse component  $\mathbf{t} \equiv \nabla \times \mathbf{s}$ ,

$$\mathbf{s} = \frac{\nabla}{\nabla^2} \sigma - \frac{1}{\nabla^2} \nabla \times \mathbf{t}, \quad (2.47)$$

where the resulting evolution equations show that the transverse contribution  $\mathbf{t}$  only becomes relevant at third order [181, 190]. The time component can be factorized out, which allows for writing evolution equations for  $\sigma$  and  $\mathbf{t}$  order by order as a function of their initial conditions, and also for the construction of a closed bias expansion (see below). Although this factorization is mostly used in the context of an Einstein-de Sitter (EdS) Universe, it is also possible for a general expansion history [71].

**Equation of motion.** We can write the equation of motion governing the fluid trajectory as [68]

$$\left( \frac{\partial^2}{\partial \tau^2} + \mathcal{H} \frac{\partial}{\partial \tau} \right) \mathbf{s}(\mathbf{q}, \tau) = -\nabla \Phi(\mathbf{q} + \mathbf{s}(\mathbf{q}, \tau), \tau). \quad (2.48)$$

We take the divergence

$$\nabla_x \cdot \left( \frac{\partial^2}{\partial \tau^2} + \mathcal{H} \frac{\partial}{\partial \tau} \right) \mathbf{s}(\mathbf{q}, \tau) = -\frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) \delta(\mathbf{q}, \tau), \quad (2.49)$$

with initial condition  $\mathbf{s}(\mathbf{q}, \tau = 0) = 0$ . We want to compute the Eulerian divergence of the displacement field  $\mathbf{s}$  in matrix notation:

$$\nabla_x \cdot \mathbf{s} = \frac{\partial s_j}{\partial x_j}. \quad (2.50)$$

Using the chain rule, we relate derivatives with respect to Eulerian and Lagrangian coordinates via

$$\frac{\partial}{\partial q_i} = \left( \delta_{ij}^K + M_{ij} \right) \frac{\partial}{\partial x_j}, \quad (2.51)$$

where we define the displacement gradient as

$$M_{ij} \equiv \frac{\partial s_j}{\partial q_i}. \quad (2.52)$$

Inverting the Jacobian, we obtain:

$$\frac{\partial}{\partial x_j} = \left[ (\delta^{Kj} + M)^{-1} \right]_{ji} \frac{\partial}{\partial q_i}. \quad (2.53)$$

Applying this to the divergence:

$$\frac{\partial s_j}{\partial x_j} = [(\delta^K + M)^{-1}]_{ji} \frac{\partial s_j}{\partial q_i} \quad (2.54)$$

$$= [(\delta^K + M)^{-1} \cdot M]_{ii} \quad (2.55)$$

$$= \text{Tr} [(\mathbf{1} + \mathbf{M})^{-1} \mathbf{M}]. \quad (2.56)$$

We can then write

$$\text{Tr} \left[ (\mathbf{1} + \mathbf{M})^{-1} \left( \frac{\partial^2}{\partial \tau^2} + \mathcal{H} \frac{\partial}{\partial \tau} \right) \mathbf{M} \right] = \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) (|\mathbf{1} + \mathbf{M}(\mathbf{q}, \tau)|^{-1} - \mathbf{1}), \quad (2.57)$$

which gives us the equation of motion in LPT.

**Zel'dovich approximation.** Eq. (2.49), which looks very similar to Eq. 2.18 and Eq. 2.20, has the first order solution

$$\mathbf{s}^{(1)}(\mathbf{q}, \tau) = -\frac{\nabla}{\nabla^2} \delta^{(1)}(\mathbf{q}, \tau). \quad (2.58)$$

This is known as Zel'dovich approximation, which tells us that, at first order in the displacement, the particles follow a straight line.

## 2.3 Galaxy bias

In this section, we discuss how to construct the bias expansion. A comprehensive discussion of galaxy bias models can be found in [66].

### 2.3.1 The leading local gravitational observable

The equivalence principle posits that in a free-falling reference frame, such as one that is commoving with the trajectory of the galaxy<sup>1</sup>, the leading locally observable gravitational effects are governed by the second derivatives of the metric tensor. Since galaxies are non-relativistic, the only relevant component of the metric tensor is the time-time component  $g_{00} \sim R_{0i0j}$ , since the motion of such tracers can be approximated by geodesics where the relativistic effects are negligible. Furthermore, on sub-horizon scales, which refers to scales smaller than the cosmological horizon where gravitational effects are felt over short distances, we have in conformal-Newtonian gauge that

$$ds^2 = -(1 + 2\Phi)d\tau^2 + (1 - 2\Psi)\delta_{ij}dx^i dx^j, \quad (2.59)$$

where  $\Phi$  and  $\Psi$  are the gravitational potentials. In the absence of anisotropic stress, the Einstein equations dictate  $\Phi = \Psi$ . We conclude that the leading observable that can

---

<sup>1</sup>we will be using the term galaxy as a general term for any cosmological tracer, as the galaxy bias expansion is valid for any tracer.

influence the formation of galaxies are the second derivatives of the gravitational potential along the fluid trajectory of the galaxy, denoted as  $\partial_i \partial_j \Phi(\mathbf{x}_{fl}(\tau), \tau)$  at fixed time  $\tau$ , and its time derivatives. This object can be decomposed into its trace and trace-free components,

$$\partial_i \partial_j \Phi = \frac{1}{3} \delta_{ij} \delta + K_{ij}, \quad (2.60)$$

where  $\delta$  is the gravitational potential, which satisfies the Poisson equation in Newtonian gravity,  $K_{ij}$  can be expressed in terms of  $\delta$  as

$$K_{ij} = \left( \frac{\partial_i \partial_j}{\nabla^2} - \frac{1}{3} \delta_{ij}^K \right) \delta. \quad (2.61)$$

### 2.3.2 EFT description

Early attempts to model galaxy bias focused on relating the galaxy distribution directly to the initial conditions, specifically by modeling the galaxy overdensity field  $\delta_g(\mathbf{x}, \tau)$  in terms of the initial matter density field [28, 110]. These approaches primarily considered correlations at early times, neglecting the complex processes involved in the subsequent gravitational evolution of structures.

A more general and physically motivated approach is to track the formation and evolution of galaxies over time, recognizing that we lack detailed knowledge of the small-scale physics governing galaxy formation. Instead of modeling these complicated microphysical processes explicitly, we parametrize their effects through a set of bias parameters. This effective description treats galaxies statistically as a fluid in phase space, following the same geodesic equations as dark matter.

Because typically  $k_{\text{NL}} < \mathcal{H}^{-1}$ , galaxy formation occurs over long time scales but small spatial scales, resulting in a “spaghetti”-like structure (see Fig. 2.1). The characteristic spatial scale is usually set by  $R_*$ . For halos and galaxies,  $R_*$  corresponds approximately to the Lagrangian radius of the (host) halo. However, for other tracers, such as line emitters, this characteristic scale can differ significantly due to the influence of long-range electromagnetic effects.

In the framework of effective field theory of LSS (EFTofLSS), we construct the theory by writing down all operators that are consistent with the symmetries of the system. In the case of LSS, the fundamental symmetry is general covariance, reflecting the diffeomorphism invariance of gravity, where the galaxy number density corresponds to the time component of the four-vector momentum density.

As we saw in Sec. 2.1, equivalence principle guarantees that the leading observable for non-relativistic observers is  $\partial_i \partial_j \Phi(\mathbf{x}_g(\tau), \tau)$  and its time derivatives along the fluid trajectory. In the EFTofLSS context [11, 31, 51, 175, 199], we aim to describe the number density of galaxies by including all possible dependencies on this quantity at a given fixed order in perturbation theory. By doing so, we refrain ourselves from modeling all unknown small-scale physics, as they can be absorbed into EFT parameters to be determined from data. The ultraviolet (or small-scale) cutoff  $\Lambda$  should be seen as a computational tool

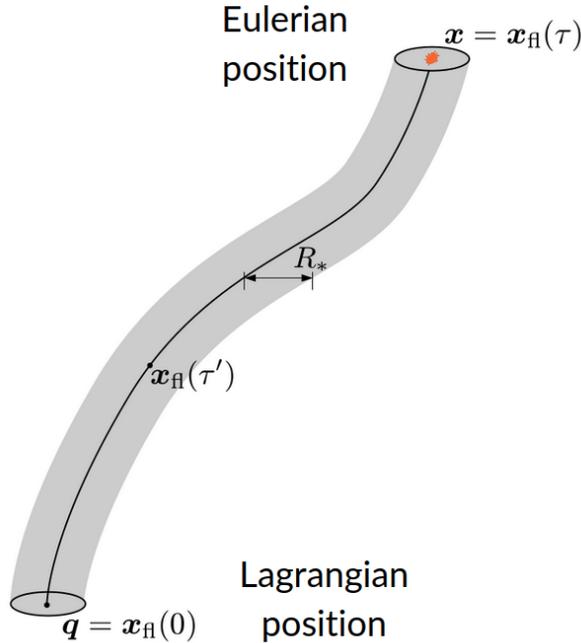


Figure 2.1: Galaxy formation picture, happening on large time scales and small spatial scales. Adapted from [66].

that guarantees loop integrals remain finite, as any observable has to be independent of its choice.

The contributions to the bias expansion are organized according to their perturbative order and the number of spatial derivatives, following a gradient expansion. Operators with higher numbers of derivatives are suppressed at large scales. Importantly, fields such as the gravitational potential  $\phi$ , its gradient  $\nabla\phi$ , and the peculiar velocity  $v$  cannot appear directly in the bias expansion. This is because it is always possible to choose a local coordinate frame in which these quantities vanish (by a suitable choice of gauge or frame), making them unphysical for the purpose of constructing a covariant expansion.

Furthermore, since we are considering operators that can appear during the entire time evolution of a galaxy along its fluid trajectory, certain quantities are redundant. For example, the velocity divergence  $\theta = \nabla \cdot \mathbf{v}$  and the shear tensor  $\partial_i u^j$  can be eliminated using the equations of motion. Thus, such quantities do not need to be included as independent operators in the bias basis.

### 2.3.3 Time factorization and finite basis

All that we are going to assume is that galaxy formation is local, and therefore should only depend on the local observables along the fluid flow, which is the leading gravitational observable  $\partial_i \partial_j \phi$ . Ignoring time evolution for now, we can write a local bias relation

$$\delta_g(\mathbf{x}, \tau) = F_g(\partial_i \partial_j \phi(\mathbf{x}, \tau), \tau), \quad (2.62)$$

so it is easy to expand in  $\partial_i \partial_j \phi(\mathbf{x}, \tau)$  and write all the possible scalar contractions of this functional as

$$\delta_g = \sum_O b_O O, \quad (2.63)$$

where  $b_O$  are the functional terms which are left to be constrained to data. We know however that we need to integrate over the fluid trajectory, so we need to promote the function to a nonlocal-in-time functional

$$\delta_g(\mathbf{x}, \tau) = F_g(\partial_i \partial_j \phi(\mathbf{x}_\#(\tau'), \tau')). \quad (2.64)$$

It does not seem possible to Taylor expand this function, since  $\phi$  will evolve significantly over the time evolution of the galaxy. For example, let us say we have an operator constructed out of  $\partial_i \partial_j \phi(\mathbf{x}_\#(\tau'), \tau')$ . If we try to do the expansion

$$\delta_g(\mathbf{x}, \tau) \supset \int^\tau d\tau' f_O(\tau, \tau') O(\mathbf{x}_\#(\tau'), \tau') \quad (2.65)$$

$$= \left[ \int^\tau d\tau' f_O(\tau, \tau') \right] O(\mathbf{x}, \tau) + \left[ \int^\tau d\tau' (\tau' - \tau) f_O(\tau, \tau') \right] \frac{D}{D\tau} O(\mathbf{x}, \tau) + \dots, \quad (2.66)$$

the higher-order terms are not necessarily small. However, we can invoke the scale-free nature of gravity. Suppose we have a linear functional, the functional can be written as

$$\int_0^\tau d\tau' f_{g,\delta}(\tau, \tau') \delta(\mathbf{x}_\#(\tau'), \tau'). \quad (2.67)$$

In perturbation theory, this can be written as a scale-invariant function where we separate time and space dependencies,

$$\delta(\mathbf{x}_\#(\tau'), \tau') = D(\tau') \delta^{(1)}(\mathbf{x}, \tau_0). \quad (2.68)$$

Now, we can clearly see that  $\delta^{(1)}(\mathbf{x}, \tau_0)$  is independent of time, so it comes out of the integral and we parametrize the time dependent integral as a linear bias term which we just have to fit to data and we don't care about its specific form. In fact,  $b_1(\tau) \delta^{(1)}(\mathbf{x}, \tau)$  is the linear term appearing in the bias expansion. We can continue at higher orders and see we will always obtain a similar structure, with the same operators appearing or additional ones with new free bias coefficients.

### 2.3.4 Lagrangian Bias

Let us now explicitly construct a Lagrangian basis of bias operators. Since at leading order in spatial derivatives all the allowed terms by EFTofLSS are contained in  $\mathbf{M}(\mathbf{q}, \tau)$  and its time derivatives [148], the deterministic part of the galaxy overdensity field in the Lagrangian picture can be written as

$$\delta_{g,\text{det}}^L(\mathbf{q}, \tau) = \int_0^\tau d\tau' F_g[\mathbf{M}(\mathbf{q}, \tau'), \tau', \tau], \quad (2.69)$$

where  $F_g$  is a nonlocal in time functional. By expanding  $F_g$  in  $\mathbf{M}$ , we obtain all the rotational invariants of  $\mathbf{M}$ , and the time integral can be performed thanks to the aforementioned time factorization. We can construct the Lagrangian bias expansion as

$$\delta_{g,\text{det}}^L(\mathbf{q}, \tau) = \sum_{\mathcal{O}^L} b_{\mathcal{O}^L}(\tau) \mathcal{O}^L(\mathbf{q}, \tau), \quad (2.70)$$

where  $b_{\mathcal{O}^L}$  denote the Lagrangian bias parameters that encode all complex and non-linear physics of galaxy formation from  $F_g$ . The Lagrangian operators  $\mathcal{O}^L$  are determined by taking all scalar invariants of the symmetric part of  $\mathbf{M}^{(n)}$  and their independent products. It is not necessary to include  $\text{tr}[\mathbf{M}^{(n)}]$  for  $n > 1$  nor the antisymmetric contributions, since they can be expressed in terms of lower-order symmetric ones via the equations of motion [143, 231].

At linear order,  $M_{ij}^{(1)}$  is directly proportional to  $\partial_{q,i} \partial_{q,j} \Phi^{(1)}$ . Since time derivatives can always be rewritten as a sum of higher-order operators, we only need to consider all possible scalar contractions of the contributions  $M_{ij}^{(n)}$  at each perturbative order.

The local basis of the deterministic Lagrangian operators up to fourth order is then given by [148]

$$\begin{aligned} 1^{\text{st}} & \quad \text{tr}[\mathbf{M}^{(1)}] \\ 2^{\text{nd}} & \quad \text{tr}[(\mathbf{M}^{(1)})^2], (\text{tr}[\mathbf{M}^{(1)}])^2 \\ 3^{\text{rd}} & \quad \text{tr}[(\mathbf{M}^{(1)})^3], \text{tr}[(\mathbf{M}^{(1)})^2] \text{tr}[\mathbf{M}^{(1)}], (\text{tr}[\mathbf{M}^{(1)}])^3, \text{tr}[\mathbf{M}^{(1)} \mathbf{M}^{(2)}] \\ 4^{\text{th}} & \quad \text{tr}[(\mathbf{M}^{(1)})^4], \text{tr}[(\mathbf{M}^{(1)})^3] \text{tr}[\mathbf{M}^{(1)}], (\text{tr}[(\mathbf{M}^{(1)})^2])^2, (\text{tr}[\mathbf{M}^{(1)}])^4, \\ & \quad \text{tr}[\mathbf{M}^{(1)}] \text{tr}[\mathbf{M}^{(1)} \mathbf{M}^{(2)}], \text{tr}[\mathbf{M}^{(1)} \mathbf{M}^{(1)} \mathbf{M}^{(2)}], \text{tr}[\mathbf{M}^{(1)} \mathbf{M}^{(3)}], \text{tr}[\mathbf{M}^{(2)} \mathbf{M}^{(2)}] \end{aligned} \quad (2.71)$$

At fourth order, we utilize the fact that  $M_{ij}^{(1)}$  is a symmetric  $3 \times 3$  matrix and is characterized by three independent rotational invariants. This allows us to eliminate  $\text{tr}[(\mathbf{M}^{(1)})^2] (\text{tr}[\mathbf{M}^{(1)}])^2$ . From third order onwards,  $M_{ij}$  is no longer necessarily symmetric, meaning that the displacement vector can also include a curl component. However, at each perturbative order, the antisymmetric part of  $M_{ij}$  can be rewritten in terms of lower-order contributions to the symmetric part, defined as

$$M_{(ij)} \equiv \frac{M_{ij} + M_{ji}}{2}, \quad (2.72)$$

through the equations of motion. Consequently, the bias expansion can be written purely in terms of contributions to the symmetric part  $M_{(ij)}^{(n)}$  [66].

### 2.3.5 Eulerian Bias

Following [66, 148], we define

$$\Pi_{ij}^{[1]}(\mathbf{x}, \tau) = \frac{2}{3\Omega_m a H^2} \partial_{x_i} \partial_{x_j} \Phi(\mathbf{x}, \tau) = K_{ij}(\mathbf{x}, \tau) + \frac{1}{3} \delta_{ij}^K \delta(\mathbf{x}, \tau), \quad (2.73)$$

where  $\mathbf{k} \equiv \nabla\delta$  is the tidal field tensor, and  $\delta^K$  denotes the Kronecker delta. The superscript [1] indicates that the lowest-order contribution appears at  $n = 1$  in perturbation theory. The operator  $\mathbf{\Pi}^{[n]}$  follows the recursion relation

$$\Pi_{ij}^{[n]} = \frac{1}{(n-1)!} \left[ (aHf)^{-1} \frac{D}{D\tau} \Pi_{ij}^{[n-1]} - (n-1) \Pi_{ij}^{[n-1]} \right], \quad (2.74)$$

where  $D/D\tau \equiv \partial_\tau + v^i \partial_x$  is the convective derivative with respect to  $\tau$ . The complete set of operators in Eulerian space up to fourth order is

$$\begin{aligned} 1^{\text{st}} & \quad \text{tr}[\mathbf{\Pi}^{[1]}] \\ 2^{\text{nd}} & \quad \text{tr}[(\mathbf{\Pi}^{[1]})^2], (\text{tr}[\mathbf{\Pi}^{[1]}])^2 \\ 3^{\text{rd}} & \quad \text{tr}[(\mathbf{\Pi}^{[1]})^3], \text{tr}[(\mathbf{\Pi}^{[1]})^2] \text{tr}[\mathbf{\Pi}^{[1]}], (\text{tr}[\mathbf{\Pi}^{[1]}])^3, \text{tr}[\mathbf{\Pi}^{[1]}\mathbf{\Pi}^{[2]}] \\ 4^{\text{th}} & \quad \text{tr}[(\mathbf{\Pi}^{[1]})^4], \text{tr}[(\mathbf{\Pi}^{[1]})^3] \text{tr}[\mathbf{\Pi}^{[1]}], (\text{tr}[(\mathbf{\Pi}^{[1]})^2])^2, (\text{tr}[\mathbf{\Pi}^{[1]}])^4, \\ & \quad \text{tr}[\mathbf{\Pi}^{[1]}] \text{tr}[\mathbf{\Pi}^{[1]}\mathbf{\Pi}^{[2]}], \text{tr}[\mathbf{\Pi}^{[1]}\mathbf{\Pi}^{[1]}\mathbf{\Pi}^{[2]}], \text{tr}[\mathbf{\Pi}^{[1]}\mathbf{\Pi}^{[3]}], \text{tr}[\mathbf{\Pi}^{[2]}\mathbf{\Pi}^{[2]}] \end{aligned} \quad (2.75)$$

Different equivalent linear combinations exist in the literature, as discussed in [22, 144]; further details can be found in Appendix C.2 of [66]. The complete set of EFT bias operators up to third order, following the conventions of [66, 148], is

$$O \in [\delta, \delta^2, K^2, \delta^3, K^3, \delta K^2, O_{\text{td}}], \quad (2.76)$$

where  $K^m$  denote scalar contractions of  $m$  instances of  $K_{ij}$  and  $O_{\text{td}}$  is written as

$$O_{\text{td}} \equiv \frac{8}{21} K_{ij}^{(1)} \mathcal{D}^{ij} \left[ (\delta^{(1)})^2 - \frac{3}{2} (K_{ij}^{(1)})^2 \right], \quad (2.77)$$

and we defined

$$\mathcal{D}^{ij} \equiv \frac{\partial_i \partial_j}{\nabla^2} - \frac{1}{3} \delta_{ij}^K. \quad (2.78)$$

### 2.3.6 Gravity and non-locality

As we have just seen, the operator  $O_{\text{td}}$  has a non-local structure  $\partial_i \partial_j / \nabla^2$  appearing in  $\mathcal{D}_{ij}$ . Moreover, note that this has order zero in spatial derivatives, so it does not count as a higher-derivative operator (see below). This is because we count the order in the derivative expansion as powers of  $\partial_i \partial_j$  acting on  $\partial_i \partial_j \phi$ ;  $O_{\text{td}}$  has two powers of derivatives in the numerator and denominator, so the net order in spatial derivatives is zero. At first glance, this appears to be a contradiction with our supposition that galaxy formation is a local process. However, it turns out that it is a local observable. It can be shown [66] that this operator can be rewritten in terms of convective time derivatives of  $K_{ij}$ . Since time derivatives of observable quantities should be also observables, it follows that  $O_{\text{td}}$  is also a local observable.

This can be understood by thinking of non-locality of gravity due to its long-range nature, which produces local observables that have to be taken into account in the bias

expansion. For instance,  $K_{ij}$  contains non-local information about  $\delta$ , similarly to how gravitational waves propagate through spacetime carrying information on distant sources, due to the long-range gravity. In particular, it can be shown that  $K_{ij}$  is a projection of the Weyl tensor [100], which is non-locally related to  $T_{\mu\nu}$ , or the density, as is clear from  $\nabla^{-2}$  appearing in Eq. (2.61). Since the Weyl tensor is not directly determined by local sources, this means that tidal effects at a given point can be influenced by distant masses.  $O_{\text{td}}$  in turn is non-local in  $\partial_i\partial_j\phi$ , arising due to the gravitational evolution of  $K_{ij}$ . Assuming a local bias expansion means that galaxy formation should only depend on local gravitational observables and its convective time derivatives, including the ones non-local in  $\delta$  or  $\partial_i\partial_j\phi$  generated by gravity.

### 2.3.7 Advection of operators

The observed galaxy density field is in Eulerian space, while the Lagrangian operators listed in Eq. (2.71) are expressed in Lagrangian space. We use Eq. (2.43) to *advect* the Lagrangian operators to Eulerian space. The continuity equation for galaxies gives the relation between the galaxy overdensity in Eulerian and Lagrangian coordinates, i.e.,

$$\delta_{g,\text{det}}(\mathbf{k}, \tau) = \int \frac{d\mathbf{q}}{(2\pi)^3} e^{i\mathbf{k}\cdot[\mathbf{q}+s(\mathbf{q},\tau)]} [1 + \delta_g^L(\mathbf{q}, \tau)]. \quad (2.79)$$

By expanding the displacement as in Eq. (2.46) and the exponential in Eq. (2.79), it is possible to obtain the relation between the Lagrangian and Eulerian basis order by order [3]. It is worth emphasizing that Lagrangian and Eulerian descriptions correspond merely to different coordinate choices, that is, one does not assume a conserved galaxy number.

### 2.3.8 Higher-order derivative terms

So far, we have assumed that the galaxy formation process is perfectly local in space. Going beyond this approximation requires local operators  $O(\mathbf{x}, \tau)$  with functionals. For example, the linear-order operator in the Eulerian basis,  $O = \delta$ , becomes

$$b_\delta(\tau)\delta(\mathbf{x}, \tau) \rightarrow \int d^3y F_\delta(\mathbf{y}, \tau)\delta(\mathbf{x} + \mathbf{y}, \tau), \quad (2.80)$$

where  $F_\delta(\mathbf{y}, \tau)$  is a time dependent independent of  $\mathbf{x}$  due to homogeneity of the Universe. We can now perform a formal series expansion of  $\delta$  around  $\mathbf{x}$ , leading to

$$b_\delta(\tau)\delta(\mathbf{x}, \tau) \rightarrow \left[ \int d^3y F_\delta(\mathbf{y}, \tau) \right] \delta(\mathbf{x}, \tau) + \left[ \frac{1}{6} \int d^3y |\mathbf{y}|^2 F_\delta(\mathbf{y}, \tau) \right] \nabla_x^2 \delta(\mathbf{x}, \tau) + \dots \quad (2.81)$$

$$= b_\delta(\tau)\delta(\mathbf{x}, \tau) + b_{\nabla^2\delta}(\tau)\nabla_x^2\delta(\mathbf{x}, \tau) + \dots, \quad (2.82)$$

where statistical isotropy demands the absence of any preferred directions with which the derivative operators could be contracted.

This introduces higher-(spatial-)derivative contributions, the leading of which is  $\nabla^2\delta$ . By dimensional analysis, these contributions are associated with a spatial length scale  $R_*$ . We therefore need to introduce a new scale in the problem, in analogy to the speed of sound arising from pressure terms in the EFTofLSS for dark matter [32]. Supposing that  $R_*$  is of the same order as the nonlinearity scale, then within the perturbative regime it is sufficient to keep the leading order high-derivative term  $b_{\nabla^2\delta}\nabla^2\delta(\mathbf{x}, \tau)$ , as the higher-order ones are suppressed by more powers of  $(R_*k)^2$ . The higher-derivative bias are naturally introduced when imposing that the observables should not depend on the EFT cutoff [66], where it is important to stress that the nonlocality scale  $R_*$  differs from the cutoff scale  $\Lambda^{-1}$ . In particular, all the counterterms needed to renormalize the operators of Eq. (2.71) are the ones already listed there, together with their associated higher-order derivatives terms [148].

### 2.3.9 Stochasticity

So far we have discussed the deterministic operators, as in the form of Eq. (2.70), but the complete basis in the bias expansion should also account for the stochastic effect of small scales on galaxy formation. Although we are working on the context of the EFT and essentially integrating out the small-scale physics, it can happen that UV modes couple to produce a large-scale wavelength fluctuations. While we can't predict its value, it is completely described by their  $n$ -point functions, so its statistical effect can be captured by the stochastic amplitudes as free coefficients. For the two- and three- point functions, we have that, at leading order and at a fixed time slice,

$$\begin{aligned} \left\langle \delta_g(\mathbf{k}_1)\delta_g(\mathbf{k}_2) \right\rangle' \Big|_{\text{stoch.}}^{\text{LO}} &= P_\varepsilon, \\ \left\langle \delta_g(\mathbf{k}_1)\delta_g(\mathbf{k}_2)\delta_g(\mathbf{k}_3) \right\rangle' \Big|_{\text{stoch.}}^{\text{LO}} &= B_\varepsilon + 2b_1P_{\varepsilon\varepsilon\delta} \left( P_m(\mathbf{k}_1) + 2 \text{ perm.} \right), \end{aligned} \quad (2.83)$$

where the prime denotes that the momentum-conserving Dirac delta function is dropped,  $b_1$  is the Eulerian bias associated to  $\delta(\mathbf{x}, \tau)$  and  $P_m$  is the non-linear matter power spectrum. The noise parameters  $P_\varepsilon$ ,  $B_\varepsilon$  and  $P_{\varepsilon\varepsilon\delta}$  are to be determined together with the bias parameters, where the posterior of the cosmological parameters can be obtained by marginalizing over them. From central limit theorem, we expect the stochasticity to be approximately Gaussian distributed [192].

Just like for large-scale perturbations, galaxy formation also depends nonlocally on small-scale perturbations inside the region characterized by  $R_*$ , such that we also have to take into account higher-derivative contributions for the stochastic fields,

$$\langle \varepsilon(\mathbf{k})\varepsilon(\mathbf{k}') \rangle = P_\varepsilon^{\{0\}} + R_*^2 k^2 P_\varepsilon^{\{2\}} + \dots \quad (2.84)$$

In the following, we usually refer to  $P_\varepsilon^{\{0\}}$  as simply  $P_\varepsilon$ . In the case of galaxies, there is no mass and momentum conservation, so it appears already as a constant at  $k \rightarrow 0$ .

The galaxy stochasticity can be interpreted due to the discrete nature of tracers, since there is a stochasticity in the fact of whether we find or not a galaxy in within a certain

cell, commonly referred to as the “shot-noise”. This is often assumed to be close to its Poisson expected value, where its two-point amplitude goes as  $1/n_g$ . While this is not exactly true, usually  $P_\varepsilon$  is of this order.

Another interpretation is that it arises to renormalize  $P_{22}$ , absorbing the UV dependencies that appear in

$$P_{22}(k) = \langle \delta^2(\mathbf{k})\delta^2(\mathbf{k}') \rangle' = 2 \int_{\mathbf{q}} P(\mathbf{q})P(|\mathbf{k} - \mathbf{q}|)\delta_D(\mathbf{k} + \mathbf{k}'), \quad (2.85)$$

which also goes to a constant as  $k \rightarrow 0$ .

## 2.4 Wilson-Polchinski for galaxies

### 2.4.1 Wilson-Polchinski RG

The partition function (or generating functional) plays a central role in quantum field theory (QFT). It is defined as the path integral over all field configurations  $\phi(x)$ , weighted by the exponential of the action [167],

$$Z[J] = \int \mathcal{D}\phi e^{iS[\phi] + i \int d^4x J(x)\phi(x)}, \quad (2.86)$$

where  $S[\phi]$  is the action and  $J(x)$  is an external source. The functional  $Z[J]$  encodes all correlation functions via functional derivatives with respect to  $J$ .

In statistical or Euclidean field theory, one often works with the Wick-rotated version,

$$Z[J] = \int \mathcal{D}\phi e^{-S_E[\phi] + \int d^4x J(x)\phi(x)}, \quad (2.87)$$

where  $S_E$  is the Euclidean action. In contrast to the Minkowski formulation, where the integrand involves an oscillatory phase  $e^{iS}$ , the Euclidean weight  $e^{-S_E}$  is real and positive, suggesting an interpretation as a probability density over field configurations. In the classical or semiclassical limit, where quantum fluctuations are suppressed, the path integral is dominated by configurations that extremize the action, corresponding to classical solutions  $\phi_{\text{cl}}$ . In this limit,  $e^{-S_E}$  becomes sharply peaked around the classical configuration, further supporting its interpretation as a probability density function (PDF).

Moreover, the Euclidean path integral closely resembles the partition function of a statistical mechanical system, where the role of the Boltzmann factor  $e^{-\beta H}$  is played by  $e^{-S_E}$ , with  $H$  being the Hamiltonian and  $\beta$  the inverse temperature. This analogy reinforces the idea that in Euclidean signature, the path integral defines a statistical ensemble over field configurations. It should be noted, however, that in the full quantum theory  $e^{-S_E}$  generates quantum correlations rather than probabilities in a classical sense, and thus the PDF interpretation is strictly valid only in the classical or semiclassical limit.

To regularize the theory, a momentum cutoff  $\Lambda$  is introduced, making the theory explicitly cutoff-dependent. Physical observables at energies  $E \ll \Lambda$  must, however, remain

independent of  $\Lambda$ . This requirement leads to renormalization group (RG) flow equations, encapsulated in the condition:

$$\frac{d}{d\Lambda} Z[J; \Lambda] = 0. \quad (2.88)$$

To understand the evolution of the theory with scale, we introduce the *effective action*  $S_\Lambda[\phi]$ , defined by integrating out high-momentum modes above the scale  $\Lambda$ . The operators  $\mathcal{O}_i$  appearing in the effective action are accompanied by scale-dependent *Wilson coefficients*  $g_i(\Lambda)$ , encoding the imprint of high-energy physics on the low-energy theory. For a general bare action with  $g_i(\Lambda)$ ,

$$S_\Lambda[\phi] = \int d^4x \left[ \frac{1}{2} (\partial\phi)^2 + \sum_i g_i(\Lambda) \mathcal{O}_i(\phi) \right], \quad (2.89)$$

the requirement of cutoff independence implies the couplings must run with  $\Lambda$ ,

$$\Lambda \frac{dg_i}{d\Lambda} = \beta_i(g_1, g_2, \dots), \quad (2.90)$$

where the  $\beta_i$  functions characterize how couplings evolve with energy scale.

This framework underlies Wilsonian renormalization [226]: high-energy modes are integrated out incrementally, modifying the couplings so that low-energy physics remains unchanged. Polchinski later formulated this RG procedure more precisely using a differential flow equation in the path integral formalism [172], leading to what is now called the Wilson-Polchinski RG. This formalism integrates out momentum shells infinitesimally, and describes how the effective action  $S_\Lambda[\phi]$  evolves continuously with  $\Lambda$ , making the RG flow manifest.

Crucially, for the theory to be renormalizable, the set of operators  $\{\mathcal{O}_i\}$  must be closed under RG flow [204]. That is, the RG evolution should not generate new operators beyond those included in the original effective action (or at least, all new terms should be accounted for systematically in an effective field theory expansion). This closure ensures that the theory remains predictive at each scale.

## 2.4.2 The partition function of galaxies

Similarly to quantum field theory, we can write a partition function for galaxies. Here, the arbitrary cutoff has to be chosen such that  $\Lambda < k_{\text{NL}}$ , where perturbation theory breaks down (see Fig. (2.2)). We emphasize here that all modes  $k < \Lambda$  are kept explicitly in the calculations.

The path integral approach was first applied to LSS in the context of the matter density field [51], where the interaction action is given by the solution to the equations of motion for the matter density field, essentially the Euler and continuity equations. Instead of smoothing the equations of motion as is commonly done in the context of the EFTofLSS [32], [51] demands that the coefficients in the effective action are closed under renormalization group flow, where the counterterms with structure  $c_{\nabla^2\delta}(\Lambda)\nabla^2\delta$  absorb the cutoff dependence introduced by integrating out high-energy modes.

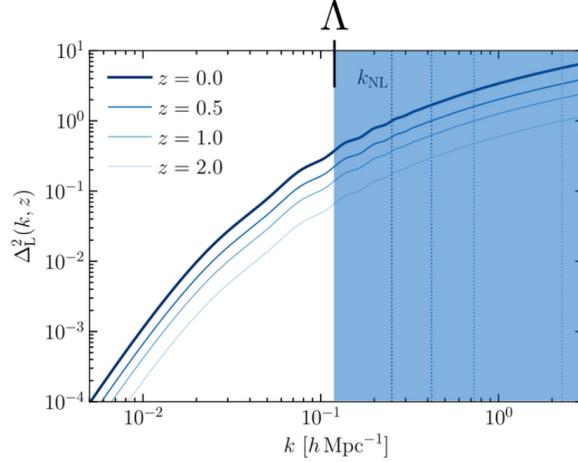


Figure 2.2: Normalized linear power spectrum at different redshifts and corresponding non-linear scale  $k_{\text{NL}}$ , which increases with redshift due to non-linear clustering. The cutoff scale  $\Lambda$  is also indicated, where the blue shaded region correspond to the small-scales that are integrated over in the EFTofLSS. Adapted from [69].

For galaxies, the bias coefficients are Wilson coefficients of our theory, where they depend on  $\Lambda$  while the cosmological parameters should independent of the cutoff. An interaction action in this case  $S_{\text{int}}[\delta_{\Lambda}^{(1)}, J_{\Lambda}]$  is “reverse-engineered” to give us the bias expansion we derived in the previous sections based on symmetry arguments, together with the free propagator for the linear density field  $P_L^{\Lambda}(k) = \langle \delta_{\Lambda}^{(1)}(\mathbf{k}) \delta_{\Lambda}^{(1)}(\mathbf{k}')' \rangle$ . This allows us to write the galaxy partition function [46, 186]

$$\begin{aligned}
\mathcal{Z}[J_{\Lambda}] &= \int \mathcal{D}\delta_{\Lambda}^{(1)} \mathcal{P}[\delta_{\Lambda}^{(1)}] \exp\left(S_{\text{int}}[\delta_{\Lambda}^{(1)}, J_{\Lambda}]\right) \\
&= \int \mathcal{D}\delta_{\Lambda}^{(1)} \mathcal{P}[\delta_{\Lambda}^{(1)}] \exp\left(\int_{\mathbf{k}} J_{\Lambda}(\mathbf{k}) \left[\sum_{\mathcal{O}} b_{\mathcal{O}}^{\Lambda} O[\delta_{\Lambda}^{(1)}](-\mathbf{k})\right]\right) \\
&\quad + \sum_{m=2} \frac{C_{\mathbf{1}}^{(m)}(\Lambda)}{m!} \int_{\mathbf{x}} [J_{\Lambda}(\mathbf{x})]^m + \sum_{m=2} \sum_{\mathcal{O}} \frac{C_{\mathcal{O}}^{(m)}(\Lambda)}{m!} \int_{\mathbf{x}} [J_{\Lambda}(\mathbf{x})]^m O[\delta_{\Lambda}^{(1)}](\mathbf{x}),
\end{aligned} \tag{2.91}$$

where the current  $J_{\Lambda}$  has support up to  $\Lambda$ . The introduction of a finite cutoff guarantees that all observables derived from  $\mathcal{Z}$  remain finite. However, their accuracy is limited to energy scales below  $\Lambda$ , which is why the source current is truncated at  $\Lambda$ . The integration is done over the initial conditions  $\delta_{\Lambda}^{(1)}$ , where we suppress all modes  $k > \Lambda$ , which in the absence of primordial non-Gaussianities follows

$$\mathcal{P}[\delta_{\Lambda}^{(1)}] = \left(\prod_k^{\Lambda} 2\pi P_L^{\Lambda}(k)\right)^{-1/2} \exp\left[-\frac{1}{2} \int_{\mathbf{k}}^{\Lambda} \frac{|\delta_{\Lambda}^{(1)}|^2}{P_L^{\Lambda}(k)}\right]. \tag{2.92}$$

The terms linear in  $J_{\Lambda}$  correspond to the deterministic part of the bias expansion, while the terms in the second line of Eq. (2.91) correspond to the stochastic contributions. On

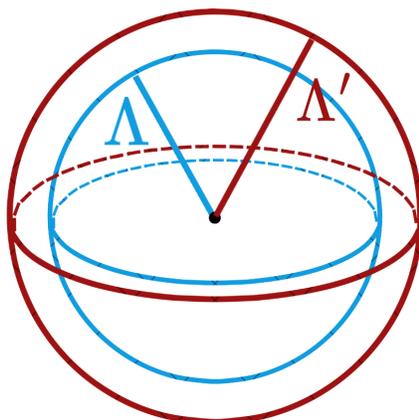


Figure 2.3: Schematic representation of  $\delta_\Lambda^{(1)}$ : the shell is given by the part in between  $\Lambda$  and  $\Lambda'$ .

the left hand side, we have the purely stochastic terms of order  $\mathcal{O}[J_\Lambda^m]$  for  $m \geq 2$ . For instance, the two-point amplitude  $P_\varepsilon = C_{\mathbb{1}}^{(2)}$  and three-point amplitude  $B_\varepsilon = C_{\mathbb{1}}^{(3)}$  (where  $\mathbb{1}$  corresponds to the “unity” operator  $O(\mathbf{x}) = \mathbb{1}$ ). On the right hand side, we have the coupling of stochastic and deterministic contributions of order  $\mathcal{O}[J_\Lambda^m \delta_\Lambda^{(1)}]$  (for  $m \geq 2$ ), encoding terms such as the density dependent noise amplitude  $P_{\varepsilon\delta} = C_\delta^{(2)}$ .

We can use this partition function to generate all  $n$ -point functions via functional derivatives of  $\mathcal{Z}$  with respect to the current. For example,

$$\begin{aligned} \langle \delta_g(\mathbf{k}) \delta_g(\mathbf{k}') \rangle' &= \left( \frac{\partial^2 \ln(\mathcal{Z}[J_\Lambda]/\mathcal{Z}[0])}{\partial J_\Lambda(\mathbf{k}) \partial J_\Lambda(\mathbf{k}')} \Big|_{J_\Lambda=0} \right)' \\ &= \sum_{O, O'} b_O b_{O'} \langle O(\mathbf{k}) O'(\mathbf{k}') \rangle' + P_\varepsilon(k) \end{aligned} \quad (2.93)$$

Remarkably, the generating function is able to predict the galaxy  $n$ -point functions structure at any order in perturbation theory. At leading order, we obtain the well-known expression

$$\langle \delta_g(\mathbf{k}) \delta_g(\mathbf{k}') \rangle' = b_1^2 P_L(k) + P_\varepsilon(k). \quad (2.94)$$

### 2.4.3 Galaxy bias RG equations

The Wilson-Polchinski approach to renormalization proceeds by integrating out modes in the free field above a momentum cutoff  $\Lambda$ . In the LSS case, this corresponds to integrating modes above the cutoff in the linear density field. This is in contrast to the  $n$ -point-function-based renormalization approach of [22], since here all linear modes below the cutoff are kept explicitly, whereas the  $n$ -point-function approach integrates over all modes, including those on large scales.

The partition function at some scale  $\Lambda'$  will predict the running of the bias parameters at a lower scale  $\Lambda$ . Since we need to integrate out the high-energy modes step by step, we separate the modes of the linear density field in a thin shell  $\delta_{\text{shell}}^{(1)}(\mathbf{k})$  for  $\mathbf{k} \in (\Lambda, \Lambda']$  (see Fig. 2.3), such that

$$\delta_{\Lambda'}^{(1)}(\mathbf{k}) = \delta_{\Lambda}^{(1)}(\mathbf{k}) + \delta_{\text{shell}}^{(1)}(\mathbf{k}). \quad (2.95)$$

Thanks to the Gaussianity of the initial conditions, the free action factorizes as

$$\mathcal{Z}_{\text{eff}}[J_{\Lambda}] = \int \mathcal{D}\delta_{\Lambda}^{(1)} \mathcal{P}[\delta_{\Lambda}^{(1)}] \int \mathcal{D}\delta_{\text{shell}}^{(1)} \mathcal{P}[\delta_{\text{shell}}^{(1)}] \exp \left( \int_{\mathbf{k}} J_{\Lambda}(\mathbf{k}) \left[ \sum_{\mathcal{O}} b_{\mathcal{O}}^{\Lambda'} O[\delta_{\Lambda}^{(1)} + \delta_{\text{shell}}^{(1)}](-\mathbf{k}) \right] \right), \quad (2.96)$$

where we discard stochastic terms, which do not contribute to the flow of deterministic bias parameters [187].

We proceed by considering an infinitesimally thin shell to expand Eq. (2.96) in  $\delta_{\text{shell}}^{(1)}$ . Keeping only the non-trivial terms linear in  $\delta_{\text{shell}}^{(1)}$ , we obtain

$$\mathcal{Z}[J_{\Lambda}] = \int \mathcal{D}\delta_{\Lambda}^{(1)} \mathcal{P}[\delta_{\Lambda}^{(1)}] \exp \left( \int_{\mathbf{k}} J_{\Lambda}(\mathbf{k}) \left[ \sum_{\mathcal{O}} b_{\mathcal{O}}^{\Lambda'} \left( O[\delta_{\Lambda}^{(1)}](-\mathbf{k}) + \mathcal{S}_{\mathcal{O}}^2[\delta_{\Lambda}^{(1)}](-\mathbf{k}) \right) \right] \right), \quad (2.97)$$

where we defined the operator expectation value over the shell

$$\mathcal{S}_{\mathcal{O}}^2[\delta_{\Lambda}^{(1)}] = \sum_{n \geq 2} \int \mathcal{D}\delta_{\text{shell}}^{(1)} \mathcal{P}[\delta_{\text{shell}}^{(1)}] O^{(n),(2),\text{shell}}[\delta_{\Lambda}^{(1)}, \delta_{\text{shell}}^{(1)}](\mathbf{k}) \quad (2.98)$$

and  $O^{(n),(2),\text{shell}}[\delta_{\Lambda}^{(1)}, \delta_{\text{shell}}^{(1)}]$  involves 2 powers of  $\delta_{\text{shell}}^{(1)}$  and  $n - 2$  powers of  $\delta_{\Lambda}^{(1)}$ . By comparing Eq. (2.97) to Eq. (2.91), imposing no dependence on the cutoff gives us

$$\sum_{\mathcal{O}} b_{\mathcal{O}}^{\Lambda} O[\delta_{\Lambda}^{(1)}] = \sum_{\mathcal{O}} b_{\mathcal{O}}^{\Lambda'} \left( O[\delta_{\Lambda}^{(1)}] + \mathcal{S}_{\mathcal{O}}^2[\delta_{\Lambda}^{(1)}] \right). \quad (2.99)$$

If (and only if) the bias expansion is closed under the renormalization group flow, the effect of integrating out the modes in  $\delta_{\text{shell}}^{(1)}(\mathbf{k})$  are absorbed into the bias coefficients, i.e.,  $\mathcal{S}_{\mathcal{O}}^2[\delta_{\Lambda}^{(1)}]$  can only depend on  $\delta_{\Lambda}^{(1)}$  through the bias operators. This allows us to derive the RG equations for galaxy bias in the form of  $db_{\mathcal{O}}/d\Lambda$  that depends on the bias coefficients, the equivalent of the  $\beta$  functions in QFT describing how the Wilson coefficients of the theory run with scale. We can also study flow in bias parameter space and draw trajectories of their running in theory space by integrating the differential equations.

Most importantly, this framework allows us to make the connection of bias parameters measured at some scale  $\Lambda$ , such as with the EFT forward model `LEFTfield` (see Sec. 2.5), to the ones measured at the large-scale limit  $\Lambda \rightarrow 0$ , such as through the separate universe technique [123]. The RG flow equations can be useful for introducing priors on bias parameters derived from the large-scale limit to bias parameters from the EFT forward model.

Although we focused our derivation on deterministic bias parameters, the path integral formalism has also been extended to predict the RG flow of stochastic parameters [187]

(those arising from the terms of order  $\mathcal{O}[J_\Lambda^m \delta_\Lambda^{(1)}, J_\Lambda^m]$  for  $m \geq 2$  in Eq. (2.91)) and for the bias parameters arising from PNGs [158], where the measure in Eq. (2.91) has to be modified to account for non-Gaussian initial conditions.

#### 2.4.4 The field-level galaxy likelihood

Unlike QFT, in the context of galaxy clustering the galaxy density field is a direct observable. Therefore, one might ask whether we can make use of the entire galaxy density field for cosmological inference instead of relying on compressions such as the galaxy  $n$ -point functions, with the hope of obtaining more cosmological information. The first ingredient for such an analysis is writing a field-level likelihood, or EFT likelihood [45–47]  $\mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}]$ , which describes the probability of the galaxy density field given the parameters of the model, such as the cosmological and bias parameters. This object can be later used to obtain the posterior of cosmological parameters given the entire galaxy density field. In the following, we describe how to make use of the galaxy partition function defined in Eq. (2.91) to obtain an expression for the field-level galaxy likelihood.

Let us suppose that we have a model for the galaxy density field  $\delta_g[\boldsymbol{\theta}, \delta_\Lambda^{(1)}]$  that depends on our parameters of interest. Since the output of this model will depend on the particular initial conditions realization, we want to average out our probability in  $\delta_\Lambda^{(1)}$  to obtain

$$\mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}] = \left\langle \delta_D^{(\infty)} [\delta_g^{\text{obs}} - \delta_g[\boldsymbol{\theta}, \delta_\Lambda^{(1)}]] \right\rangle_{\delta_\Lambda^{(1)}}, \quad (2.100)$$

where we used the notation

$$\left\langle \cdot \right\rangle_{\delta_\Lambda^{(1)}} \equiv \int \mathcal{D}\delta_\Lambda^{(1)} \mathcal{P}[\delta_\Lambda^{(1)}]. \quad (2.101)$$

Introducing an auxiliary field  $X$ , we can rewrite Eq. (2.100) as

$$\begin{aligned} \mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}] &= \int \mathcal{D}X \left\langle e^{iX(\delta_g^{\text{obs}} - \delta_g[\boldsymbol{\theta}, \delta_\Lambda^{(1)}])} \right\rangle_{\delta_\Lambda^{(1)}} \\ &= \int \mathcal{D}X e^{iX\delta_g^{\text{obs}}} \left\langle e^{-iX\delta_g[\boldsymbol{\theta}, \delta_\Lambda^{(1)}]} \right\rangle_{\delta_\Lambda^{(1)}}, \end{aligned} \quad (2.102)$$

where the last equality follows from the fact that  $\delta_g^{\text{obs}}$  is independent from  $\delta_\Lambda^{(1)}$ . Looking at Eq. (2.91), we can immediately recognize that

$$\mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}] = \int \mathcal{D}X e^{iX\delta_g^{\text{obs}}} \mathcal{Z}[-iX]. \quad (2.103)$$

Neglecting the terms in the partition function that are  $\mathcal{O}[J_\Lambda^2 \delta_\Lambda^{(1)}, J_\Lambda^3]$ , we are left only with terms up to  $\mathcal{O}(J_\Lambda^2)$ , what allows us to perform a Gaussian integral over the leading order stochastic contribution to finally obtain

$$\ln \mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}] = -\frac{1}{2} \int_{k < \Lambda} \frac{|\delta_g^{\text{obs}}(\mathbf{k}) - \delta_{g,\text{det}}^\Lambda[\boldsymbol{\theta}, \delta_\Lambda^{(1)}](\mathbf{k})|^2}{P_\varepsilon^\Lambda(k)} + \ln[2\pi P_\varepsilon^\Lambda(k)], \quad (2.104)$$

where

$$\delta_{g,\text{det}}^\Lambda[\boldsymbol{\theta}](\mathbf{k}) = \sum_{\mathcal{O}} b_{\mathcal{O}}^\Lambda O[\boldsymbol{\theta}](\mathbf{k}). \quad (2.105)$$

In practice, it is unfeasible to analytically perform the integral over initial conditions (Eq. (2.101)). What one can do instead is develop a field-level forward model and make use of Eq. (2.104) to numerically sample the posterior  $\mathcal{P}[\boldsymbol{\theta}|\delta_g^{\text{obs}}]$ . This is a very complicated task, since the dimensionality of the problem scales as  $N_g^3 = \Lambda^{-3}$ , where  $N_g$  is the (cubic) grid size. Moreover, the model  $\delta_g[\boldsymbol{\theta}, \delta_\Lambda^{(1)}]$  has to be extremely precise at the field-level to avoid any complications with model mismatch; while at the level of the  $n$ -point functions one would expect differences in the model to the observed data to be attenuated by the averages that are taken over the field, the likelihood in Eq. (2.104) is comparing the model and the observed data at each grid cell. In the following section, we describe how to construct a forward model based on the galaxy bias expansion, while in Sec. 3.6, we describe how to use it to perform field-level inference in practice.

## 2.5 LEFTfield

LEFTfield is a *Lagrangian* forward model based on the *EFTofLSS*, where the bias expansion is implemented directly at the *field* level. Originally introduced in detail by [193], its physical and numerical convergence was further investigated by [206] in restframe and [207] in redshift space. The model serves as a robust and efficient simulator for simulation-based inference, demonstrated in [217] using mock data (Chapter 4) and dark-matter halos (Chapter 5) to infer the cosmological parameter  $\sigma_8$  in the restframe. Additionally, it supports field-level inference by enabling marginalization over initial conditions, as discussed in the previous section. Its effectiveness in this context has been shown by [118] with mock data, by [157] and [206] for dark-matter halos (Chapter 5), by [38] for HOD galaxies in a blind challenge, and by [205] in redshift space with dark-matter halos. Furthermore, LEFTfield facilitates field-level inference of the BAO scale [23, 24] (Chapter E), and provides an efficient method to measure bias parameters for both dark-matter halos [121] and galaxies [30] in simulations.

**Lagrangian-EFT forward model.** In essence, the cell size in LEFTfield is given by  $\Lambda^{-1}$ , where we evolve all modes up to the cutoff at the field-level. The fact that LEFTfield is a Lagrangian forward model means that it uses LPT for the gravity evolution (instead of an N-body solver, for example), although it allows for both Lagrangian and Eulerian bias expansions via the mapping given in Eq. (2.43) (see below). In practice, this also means that redshift space modeling is more straightforward, essentially by modifying Eq. (2.43) and the bias operators accordingly [207]. However, in this thesis we will only be working in restframe.

**Bias running.** The inferred bias and stochastic parameters will depend on (“run with”)  $\Lambda$  [185], while inferred cosmological parameters are independent of  $\Lambda$ . We stress the dif-

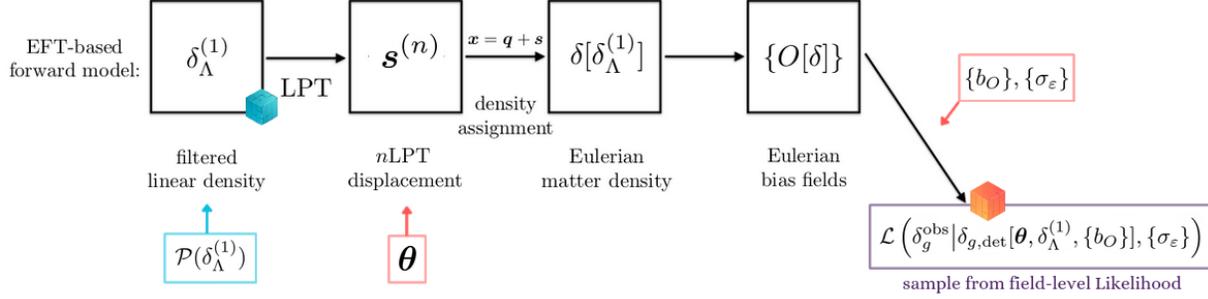


Figure 2.4: Diagram for the construction of Eulerian bias operators in LEFTfield.

ference between our finite, explicit cutoff  $\Lambda$  and the formal cutoff used in semi-analytical loop calculations, which is usually sent to  $\infty$  [22, 51, 185]. However, the running of the bias parameters can in principle be predicted and connected to the large-scale limit by means of the galaxy bias RG equations (see Sec. 2.4.3).

**Initial conditions.** LEFTfield, which is based on [190], works as follows. We start from an initial power spectrum  $P_L$  which is scaled by  $\alpha^2$ , where we define

$$\alpha \equiv \sigma_8 / \sigma_8^{\text{fid}} \quad (2.106)$$

and set  $\sigma_8^{\text{fid}}$  to a fiducial value, while all other cosmological parameters are fixed. For now, only changes on  $\alpha$  are implemented in the forward model (and the BAO scale, although it is kept fixed throughout this thesis until Chapter E). A Gaussian random field  $\delta_\Lambda^{(1)}$  of size  $N_g^3$  is generated from this scaled power spectrum as

$$\delta_\Lambda^{(1)}(\mathbf{k}, z) = W_\Lambda(k) \sqrt{\alpha^2 P_L(k, z)} \hat{s}(\mathbf{k}), \quad (2.107)$$

where  $\hat{s}(\mathbf{k})$  is a Gaussian random field of zero mean and unit variance,  $z$  is the redshift and the grid is smoothed at scale  $\Lambda$  with a sharp- $k$  filter  $W_\Lambda(k)$ , guaranteeing that we are only treating perturbative modes in the forward model [192]. This is related to the expansion of  $\mathbf{M}$  by expanding Eq. (2.45) to first order, yielding

$$\sigma_\Lambda^{(1)} \equiv \text{tr}[\mathbf{M}_\Lambda^{(1)}] = -\delta_\Lambda^{(1)}. \quad (2.108)$$

We then use this relation to set our initial conditions for  $\sigma_\Lambda^{(1)}$ , while  $\mathbf{t}$  is constructed in the same grid using the evolution equations. The displacement  $\mathbf{s}(\mathbf{k})$  is then evaluated from Eq. (2.47) up to the desired order in LPT.

**Lagrangian bias operators.** Eq. (2.108) gives us the initial conditions for  $\sigma_\Lambda^{(1)}$ , while  $\mathbf{M}$  can also be constructed in the same grid using the evolution equations. We then use the displacement to advect the Lagrangian operators to Eulerian space. As previously discussed, in perturbation theory this is done by both expanding the exponential and the

displacement in Eq. (2.79). With the aim of preventing noise generation on large-scales, instead of expanding the exponential we use a density assignment scheme, where the “mass of the particles” corresponds to the value of the Lagrangian operators at each grid cell, i.e.,  $M_{\text{assigner}}^O(\mathbf{q}) = O^L(\mathbf{q})$ . Note that this introduces an extra term from the Jacobian,

$$\left| \frac{\partial \mathbf{x}}{\partial \mathbf{q}} \right|^{-1} O^L(\mathbf{x}, \tau) = (1 + \delta(\mathbf{x}, \tau)) O^L(\mathbf{x}, \tau), \quad (2.109)$$

which results in the operators of interest at leading order, while  $\delta(\mathbf{x}, \tau)O^L(\mathbf{x}, \tau)$  is degenerated with higher-order terms. This approach is similar to what has been previously done in the literature; however, while [194] uses the Zel’dovich approximation (1LPT), we expand the displacement up to  $n$ -th order and [116, 228] estimate the full displacement from  $N$ -body simulations.

After the displacement  $\mathbf{s}(\mathbf{k})$  is inversely Fourier transformed to configuration space and copied to a larger grid of dimension  $N_{\text{assigner}}^3$ , we use Eq. (2.43) to advect the deterministic Lagrangian operators  $O^L$  listed in Eq. (2.71) up to the desired order in the bias expansion, with the exception of  $\text{tr}[\mathbf{M}^{(1)}]$ . Since  $\delta$  and  $\mathbf{M}$  are nonlinearly related through Eq. (2.45), the usual bias term  $b_1\delta(\mathbf{q})$  contributes to several terms of Eq. (2.71), so we instead displace a field with mass  $M_{\text{assigner}}^\delta = 1$  to obtain the Eulerian density directly. Although displacing  $\text{tr}[\mathbf{M}^{(1)}]$  would equivalently lead to a complete basis of operators, this approach allows us to work with the linear bias  $b_1$ , facilitating the interpretation of our results. The deterministic Lagrangian bias expansion in Eulerian space then reads

$$\delta_{g,\text{det}}^L(\mathbf{x}, \tau) = b_1(\tau)\delta(\mathbf{x}, \tau) + b_{\nabla^2\delta}(\tau)\nabla^2\delta(\mathbf{x}, \tau) + \sum_{O^L \neq \text{tr}[\mathbf{M}^{(1)}]} b_{O^L}(\tau)O^L(\mathbf{x}, \tau). \quad (2.110)$$

For example, at second order, we have the operators

$$O^L = \left\{ \delta, \nabla^2\delta, \left( \text{tr}[\mathbf{M}^{(1)}] \right)^2, \text{tr}[\mathbf{M}^{(1)}\mathbf{M}^{(1)}] \right\}. \quad (2.111)$$

**Eulerian bias operators.** For constructing the Eulerian bias operators (Fig. 2.4), we again displace a field with mass  $M_{\text{assigner}}^\delta = 1$  to obtain the Eulerian density  $\delta[\delta_\Lambda^{(1)}]$ . Since here the displacement field contains only large-scale modes and is protected by the equivalence principle from bias corrections, the Lagrangian basis construction is not sensitive to small-scale effects. However, this is not the case for the evolved density field, where an extra filtering is necessary to avoid uncontrolled aliasing from higher-order modes which have been populated by the nonlinear displacement. Therefore, we need to apply an extra cut before constructing the Eulerian bias operators,

$$\delta_{\Lambda_{\text{bias}}}(\mathbf{k}) = W_{\Lambda_{\text{bias}}}(k)\delta[\delta_\Lambda^{(1)}](\mathbf{k}), \quad (2.112)$$

where  $\Lambda_{\text{bias}} \geq \Lambda$  [206]. We proceed by constructing the bias fields  $O$  from the filtered Eulerian density grid,  $O[\delta_{\Lambda_{\text{bias}}}]$ , such that the Eulerian bias expansion reads

$$\delta_{g,\text{det}}(\mathbf{x}, \tau) = \sum_O b_O(\tau)O(\mathbf{x}, \tau). \quad (2.113)$$

Our choice of the Eulerian bias expansion leads to the basis, up third order,

$$O \in [\delta, \delta^2, K^2, \delta^3, K^3, \delta K^2, O_{\text{td}}, \nabla^2 \delta]. \quad (2.114)$$

**Stochasticity.** The stochastic contributions are added directly in Eulerian space, such that the final galaxy overdensity field reads

$$\delta_g(\mathbf{x}, \tau) = \delta_{g,\text{det}}(\mathbf{x}, \tau) + \varepsilon(\mathbf{x}, \tau) + c_{\varepsilon\delta}(\tau)\varepsilon(\mathbf{x}, \tau)\delta(\mathbf{x}, \tau) + c_{\varepsilon^2}(\tau)\varepsilon^2(\mathbf{x}, \tau), \quad (2.115)$$

where we also omitted the mean subtraction for  $\varepsilon^2$ . This expression exhibits some differences from the usual stochastic contributions in the bias expansion [66]. Since the simulated stochastic field  $\varepsilon$  is sampled from a Gaussian distribution with zero mean and variance  $P_\varepsilon$ , non-Gaussianity of the noise, specifically the noise bispectrum  $B_\varepsilon$  is generated via the  $\varepsilon^2$  term in the model. Precisely, we obtain

$$\left\langle \delta_g(\mathbf{k}_1)\delta_g(\mathbf{k}_2)\delta_g(\mathbf{k}_3) \right\rangle'_{\text{stoch.}}^{\text{LO}} = 6 c_{\varepsilon^2} P_\varepsilon^2 + 2 b_1 c_{\varepsilon\delta} P_\varepsilon \left( P_m(\mathbf{k}_1) + 2 \text{ perm.} \right), \quad (2.116)$$

which thus captures both the  $B_\varepsilon$  and  $P_{\varepsilon\varepsilon\delta}$  terms in Eq. (2.83) via  $c_\varepsilon^2$  and  $c_{\varepsilon\delta}$ , as derived in Appendix A. In the Poisson limit, i.e., when the stochasticity of galaxies perfectly follows Poisson statistics, determined by their mean comoving density  $\bar{n}_g$ , we expect that  $c_{\varepsilon^2} = 1/6$  and  $c_{\varepsilon\delta} = b_1/2$  (Appendix A). In general, one can go to the desired perturbative order in the stochastic field to reproduce the desired order of the stochastic  $n$ -point amplitudes.

**Assignment schemes.** There are a few options for the assignment scheme. In ascending order of complexity and precision, we have nearest-grid-point (NGP), cloud-in-cell (CIC) and non-uniform Fast Fourier Transform (NUFFT) [29]. NUFFT performs a nonuniform-to-uniform discrete Fourier transform  $f(\vec{x}) \rightarrow \tilde{f}(\vec{k})$  by mapping pseudoparticles at positions  $\vec{x}_i$  with weights  $f(\vec{x}_i)$  to a supersampled grid of size  $(N_{\text{grid}}^{\text{NUFFT}})^3$  (typically  $N_{\text{grid}}^{\text{NUFFT}} = 1.2 - 2 N_{\text{grid}}^{\text{Eul}}$ ), using an assignment kernel with compact support (typically spanning 4-16 grid cells). The algorithm then executes an FFT on the supersampled grid, deconvolves the assignment kernel, and resizes the Fourier space grid to yield the final uniform discrete Fourier transform  $\tilde{f}(\vec{k})$ . Although approximate, this method achieves accuracy close to machine precision with reasonable computational cost [29].

# Chapter 3

## Cosmological inference

### 3.1 Bayesian inference

Bayesian inference provides a powerful and coherent framework for reasoning under uncertainty. It is widely used across fields such as cosmology, statistics, machine learning, and particle physics. At its core lies Bayes' theorem, which updates prior knowledge about unknown parameters based on observed data.

**Bayes' Theorem** Let  $\boldsymbol{\theta}$  be a set of parameters describing a model, and let  $\mathbf{x}_o$  denote observed data. Bayes' theorem relates the *posterior* probability of the parameters given the data to the *prior* belief and the *likelihood* of the data,

$$p(\boldsymbol{\theta}|\mathbf{x}_o) = \frac{p(\mathbf{x}_o|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{x}_o)}. \quad (3.1)$$

Here,  $p(\boldsymbol{\theta}|\mathbf{x}_o)$  is the posterior, our updated belief about  $\boldsymbol{\theta}$  after observing  $\mathbf{x}_o$ .  $p(\mathbf{x}_o|\boldsymbol{\theta})$  is the likelihood, the probability of observing the data  $\mathbf{x}_o$  assuming parameters  $\boldsymbol{\theta}$ , and  $p(\boldsymbol{\theta})$  is the prior, our knowledge or assumptions about  $\boldsymbol{\theta}$  before seeing the data. The prior reflects our initial beliefs or theoretical assumptions about the parameters. It can encode previous measurements, theoretical constraints, or domain knowledge. Lastly,  $p(\mathbf{x}_o)$  is the evidence or marginal likelihood,

$$p(\mathbf{x}_o) = \int p(\mathbf{x}_o|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3.2)$$

ensuring normalization of the posterior.

In practice, posteriors are often intractable to compute analytically. Common sampling-based methods include Markov Chain Monte Carlo (MCMC), e.g., Metropolis-Hastings and Hamiltonian Monte Carlo.

**Marginalization and Nuisance Parameters.** In Bayesian inference, we marginalize (i.e., integrate out) nuisance parameters to focus on the variables of interest. Suppose

$\theta = (\phi, \psi)$ , where  $\psi$  is a nuisance parameter. Then the marginal posterior for  $\phi$  is

$$p(\phi|\mathbf{x}_o) = \int p(\phi, \psi|\mathbf{x}_o) d\psi. \quad (3.3)$$

## 3.2 Standard cosmological inference

Cosmological analyses have traditionally focused on extracting information primarily from the Baryon Acoustic Oscillation (BAO) “wiggles” in the galaxy two-point function. Recently, it has become possible to perform “full-shape” analyses of the galaxy power spectrum, allowing for the inclusion of information beyond the BAO scale.

The standard approach to cosmological inference from spectroscopic surveys involves a likelihood-based analysis, typically assuming a Gaussian likelihood for the galaxy  $n$ -point functions,

$$-2 \ln \mathcal{L}(\mathbf{D} | \theta) = (\mathbf{D} - \mathbf{T}(\theta))^\top \mathbf{C}^{-1} (\mathbf{D} - \mathbf{T}(\theta)), \quad (3.4)$$

where  $\mathbf{D}$  is the observed data vector,  $\mathbf{T}(\theta)$  is the theoretical prediction depending on cosmological and bias parameters, and  $\mathbf{C}$  is the covariance matrix. Note that only the mean of the data vector is modeled as a function of parameters, while the covariance is typically computed analytically or estimated from mocks at fixed fiducial values.

Using BOSS data, this methodology has been applied to a series of increasingly sophisticated analyses in the context of the EFTofLSS: first to the galaxy power spectrum [60, 104], then combining pre- and post-reconstructed power spectra [170], including the bispectrum monopole [168], bispectrum multipoles [102], the one-loop bispectrum [59, 169], and finally to the correlation function [229].

While this program has been highly successful, it also faces significant conceptual and practical challenges. Below, we highlight some of the key advantages of adopting Simulation-Based Inference (SBI) techniques for galaxy clustering analysis as a promising next step.

## 3.3 Why SBI for galaxy clustering?

Simulation-based inference (SBI) [57, 137], also known as likelihood-free inference (LFI) or implicit-likelihood inference, arises as an alternative to this approach. SBI makes use of a *simulator model* that specifies a process to generate “mock” data; that is, for a given set of parameters  $\theta$ , the model generates (or forward-models) independent samples of  $p(\mathbf{x}|\theta)$ , where  $\mathbf{x}$  is the data vector (of summary statistics, in our case). The inference makes use of the simulated dataset to obtain the posterior and, since it does not require a direct evaluation of the conditional density, is therefore in contrast to *density models*, or explicit-likelihood models, where  $p(\mathbf{x}|\theta)$  has to be specified to obtain the posterior via Bayes’ rule. This latter class of models is currently the standard in cosmology inference.

Simulator models, by virtue of their construction, operate as dynamic representations of the actual mechanisms at work based on our well-established scientific knowledge, providing

a direct and intuitive link between theoretical models and observable phenomena. They are therefore more natural and interpretable than density models, where the density function specification depends on a more abstract understanding of real-world phenomena and often requires analytical approximations [161].

Although we work with  $n$ -point functions throughout this thesis, it is important to stress that the SBI approach generalizes straightforwardly to any general data compression scheme, as long as the simulator still provides accurate predictions and the data vector is not extremely large. One can therefore study different summary statistics and their combinations with no concerns on finding an analytical expression for their distribution. This facilitates the use of more informative summary statistics, in particular those learned by neural networks [75, 109], which can be flexibly optimized according to each particular inference problem. In the context of galaxy clustering, recent progress has been made by using graphs networks to capture the map information from the galaxy distribution [141, 223].

SBI has been widely used in the cosmological inference context, including weak-lensing [15, 16, 106, 131], type IA supernovae [15–17, 36, 52, 111, 124], standard sirens [79, 80], CMB [55, 128], galaxy cluster abundance [183], Gaussian and lognormal fields [13, 58, 125, 140], dark-matter overdensity fields [58, 150], voids [212], dark-matter halos [141, 200] and galaxies [64, 90, 91, 149].

Nonetheless, only the results presented in this thesis uses SBI in the specific context of the EFTofLSS and the bias expansion [66], which together provide a rigorous framework for modeling galaxy clustering as described in Section 2.2.6. Below, we motivate some of the main advantages of using SBI for galaxy clustering analysis.

**Likelihood approximation.** Focusing on the cosmological inference procedure for galaxy clustering, the use of an analytical approximation for the likelihood becomes already inadequate for the galaxy power spectrum, the “simplest” summary statistic. It is well known that the Gaussian approximation for cosmological two-point functions fails at low wavenumbers  $k$ , since they are estimated as a sum of field amplitudes squared, and cosmic variance breaks the central limit theorem due to the small number of modes [197, 198] (see Appendix D for an illustration). This issue is investigated in [198] for weak lensing power spectra, in [89] for the 3D galaxy power spectrum, and in [126] for a log-normal density field. The Gaussian likelihood approximation particularly affects the posterior of parameters that are sensitive to low- $k$  modes, such as in the case of  $f_{\text{NL}}$  [117], which expresses the effect of primordial non-Gaussianities (PNGs), i.e., deviations from Gaussianity in the perturbations generated during the inflation epoch. In addition, by coupling different Fourier modes, nonlinear evolution also breaks the Gaussian assumption for the likelihood on small scales, not only for the power spectrum but also for many other summary statistics.

**Covariance estimation.** Even if the Gaussian likelihood assumption is accurate, we need to determine the covariance of the data vector. The usual procedure is to either assume an analytical approximation, or to estimate the covariance from simulated mock

catalogs with fixed cosmology. Analytical approximations are not always sufficient, and estimation from mocks might be cumbersome and unpractical depending on the summary statistics chosen. It is worth noting in this context that existing EFT-based analyses of galaxy clustering make use of covariances estimated from mocks, while the mean of the data vector is based on the EFT, which is strictly speaking inconsistent. In addition, there is the issue of at which (fixed or varying) point in parameter space the covariance of the data vector should be computed. SBI circumvents these issues and provides a means to use a consistent prediction of the entire distribution of the data vector and its dependence on the parameters  $\theta$ , including correlations between different elements of the data vector and higher-order moments. All of this is of great importance for scientific reasoning and the assurance that the errors are not underestimated via a poor covariance estimation; however, specific testing methods have to be applied to the SBI posterior for an evaluation of whether it itself is underestimating the errors [96].

**Binning effects.** In standard explicit-likelihood approaches, it is necessary to bin-average the theory prediction for the data vector in accordance with the binning used for the data. This is known to be important in the bispectrum case, see for example [35]. For the bispectrum, there is also the distinction between “open” and “closed” triangles, which arises from the fact that “open” triangle bins (i.e., those which do not satisfy the closed triangle condition) contain individual “closed” triangle modes [103]. These concerns are not present in the SBI approach, as the data vector can be freely chosen, as long as the same procedure is applied to observed and simulated data.

**Forward modeling.** As already mentioned, SBI makes use of simulator models to obtain the simulated dataset and therefore is an application of forward modeling, where a prescription of the generative process for the data vector as a function of the model parameters is given. Although not particular to SBI, forward modeling presents several advantages over other methods, such as the straightforward inclusion of observational effects, including window functions, redshift-space distortions and systematic effects [91, 174, 197, 205, 216, 227]. In the context of the EFT forward model employed here, `LEFTfield`, forward modeling at the field-level allows us to reach essentially arbitrary perturbative (loop) order [190].

To sum up, galaxy clustering for cosmological precision analysis is inevitably moving towards the need of higher  $n$ -point functions, combining different summary statistics and expanding the range of scales probed. It is evident how the covariance computation complexity scales with higher  $n$ -point functions, and how the need of an analytical approximation for the likelihood limits the summary statistics which can be used for analysis. Given a trustable simulator model, SBI therefore provides a flexible and rigorous framework for statistical inference in the context of galaxy clustering, provided that the following points are addressed: *model misspecification*, or how accurate is the simulator, *convergence*, or how many simulations are needed for a reliable posterior estimation, and *diagnostics*, or how reliable is the estimated posterior. In particular, we are interested in determining the

*calibration* of the posterior [96, 210] to check whether the uncertainties of the posterior are at least not underestimated.

## 3.4 Simulation-based inference

We describe in detail how simulation-based algorithms work in this section. Its fundamental requirement is to have a *simulator model*, such as `LEFTfield`. Let  $\boldsymbol{\theta}$  be the vector of parameters of interest of dimension  $N_\theta$  (e.g., cosmological, bias and noise parameters) and  $\mathbf{x}$  be the data vector of dimension  $D$  (e.g., power-spectrum bins). Assuming that we have a proposal distribution over the parameters  $\tilde{p}(\boldsymbol{\theta})$  (not necessarily the prior) and a simulator that generates  $\mathbf{x}$  given  $\boldsymbol{\theta}$ , it is possible to generate a simulated dataset of  $N_{\text{sim}}$  samples  $\{(\boldsymbol{\theta}_n, \mathbf{x}_n)\}_{n=1}^{N_{\text{sim}}}$ , where  $(\boldsymbol{\theta}_n, \mathbf{x}_n)$  is a joint sample from  $p(\boldsymbol{\theta}, \mathbf{x}) = p(\mathbf{x}|\boldsymbol{\theta})\tilde{p}(\boldsymbol{\theta})$ . We describe below how the methods which will be used throughout this paper estimate the parameters posterior from this dataset given the observed data  $\mathbf{x}_o$ .

### 3.4.1 Approximate Bayesian Computation

The first approach to perform SBI goes back to the idea of Approximate Bayesian Computation (ABC). In its simplest form, rejection ABC [179], one defines a distance metric  $\rho(\mathbf{x}, \mathbf{x}_o)$  between the simulated samples  $\mathbf{x}$  and the observed data  $\mathbf{x}_o$ , and it is possible to obtain samples from an approximate posterior  $p(\boldsymbol{\theta}|\rho(\mathbf{x}, \mathbf{x}_o) < \epsilon)$  by accepting the parameters  $\boldsymbol{\theta}_n$  for which the corresponding metric  $\rho(\mathbf{x}_n, \mathbf{x}_o)$  lies below a given threshold  $\epsilon$ . In other words, the posterior is obtained by the histogram of parameters for which the simulated samples most closely resembles the observed data. In the limit  $\epsilon \rightarrow 0$  and under suitable regularity conditions, the approximated posterior recovers the true posterior [177].

In practice, however, the value of  $\epsilon$  controls the sample acceptance rate and the required simulation budget  $N_{\text{sim}}$ , so due to computational efficiency reasons it can't be chosen to be arbitrarily small. This is in tension with the aim for the smallest possible threshold to obtain a better approximation of the posterior. Extensions of this method were proposed to improve its efficiency such as Population Monte Carlo (PMC) [12], where a sequential approach is used and the approximated posterior is used as proposal distribution  $\tilde{p}(\boldsymbol{\theta})$  for the next round until it has converged. These methods however still discard all the simulations which were rejected. Below, we discuss more efficient algorithms in which the entire dataset is used for the posterior estimation.

### 3.4.2 Neural density estimators

One of the biggest problems related to rejection-based algorithms such as ABC and PMC concerns the fact that these inherently have to throw away all the simulations which lie below a certain threshold. Neural Density Estimators (NDEs) come as a rescue, since they use all available simulations to estimate the target distribution. The dataset of parameters and simulated summaries  $\{(\boldsymbol{\theta}_n, \mathbf{x}_n)\}_{n=1}^{N_{\text{sim}}}$  is used to find a certain probability distribution.

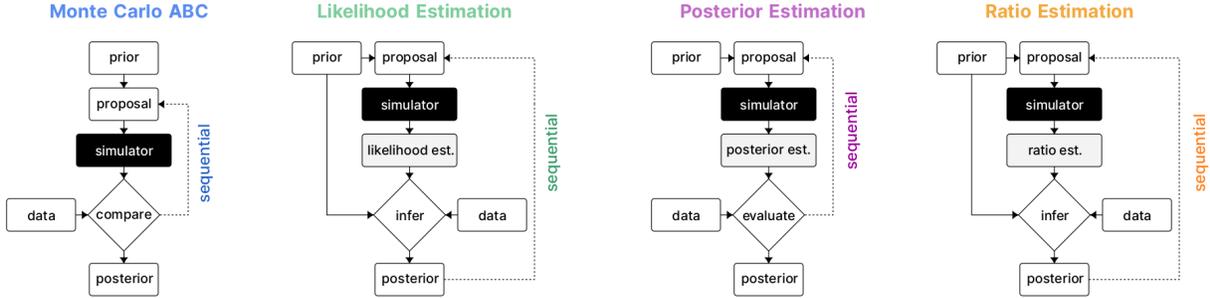


Figure 3.1: Different neural density estimation algorithms and their sequential versions. From [137].

Neural Posterior Estimation (NPE) [85] and Neural Likelihood Estimation (NLE) [165] can use *normalizing flows* [163] to fit the posterior or the likelihood, respectively. We give a detailed explanation of normalizing flows in Sec. 3.4.3. There are also methods to learn the likelihood-to-evidence ratio which use classifiers instead, such as Neural Ratio Estimation (NRE) [95] (see Fig. 3.1).

**Neural Likelihood Estimator.** Suppose that we aim to approximate the unknown target distribution  $p(\mathbf{x}|\boldsymbol{\theta})$  as the conditional of the data  $\mathbf{x}$  on the parameters  $\boldsymbol{\theta}$ . The idea is to fit a flow-based model  $q_\phi(\mathbf{x}|\boldsymbol{\theta})$  parametrized by  $\phi$  to the dataset  $\{(\boldsymbol{\theta}_n, \mathbf{x}_n)\}_{n=1}^{N_{\text{sim}}}$  by imposing that the model approximates the target distribution, where  $\tilde{p}(\boldsymbol{\theta}, \mathbf{x}) = p(\mathbf{x}|\boldsymbol{\theta})\tilde{p}(\boldsymbol{\theta})$ . Evidently we do not know the target, but we can proceed by using the maximum likelihood estimation method which minimizes the forward KL divergence in the support of the proposals,

$$\begin{aligned}
 \mathbb{E}_{\tilde{p}(\boldsymbol{\theta})} [D_{\text{KL}} [p(\mathbf{x}|\boldsymbol{\theta}) || q_\phi(\mathbf{x}|\boldsymbol{\theta})]] &= \int d\boldsymbol{\theta} \tilde{p}(\boldsymbol{\theta}) \int d\mathbf{x} p(\mathbf{x}|\boldsymbol{\theta}) \log \left( \frac{p(\mathbf{x}|\boldsymbol{\theta})}{q_\phi(\mathbf{x}|\boldsymbol{\theta})} \right) \\
 &= \int d\boldsymbol{\theta} d\mathbf{x} \tilde{p}(\boldsymbol{\theta}, \mathbf{x}) \log \left( \frac{p(\mathbf{x}|\boldsymbol{\theta})}{q_\phi(\mathbf{x}|\boldsymbol{\theta})} \right) \\
 &= -\mathbb{E}_{\tilde{p}(\boldsymbol{\theta}, \mathbf{x})} [\log q_\phi(\mathbf{x}|\boldsymbol{\theta})] + \text{const.} \\
 &\approx -\frac{1}{N_{\text{sim}}} \sum_{n=1}^{N_{\text{sim}}} \log q_\phi(\mathbf{x}_n|\boldsymbol{\theta}_n) + \text{const.},
 \end{aligned} \tag{3.5}$$

where on the third line we identified the part which is independent of  $\phi$  as a constant and on the last line we approximated the expectation over  $\tilde{p}(\boldsymbol{\theta}, \mathbf{x})$  with Monte Carlo. The resulting Monte Carlo estimate of the KL divergence we wish to minimize is therefore independent of the explicit form of the target distribution and equivalent to the sum over the negative log-likelihood of the simulated datasets batches under the flow-based model.

The estimation of conditionals is a natural extension of Masked Autoregressive Flows (MAF) [164], where only the conditionals corresponding to  $\mathbf{x}$  are modelled by augmenting the set of input variables with  $\boldsymbol{\theta}$ , i.e.,  $\mathbf{z}_t = f_t(\mathbf{z}_{t-1}, \boldsymbol{\theta})$  (see Sec. 3.4.3). The only requirement

is that  $\mathbf{x}$  appears before  $\boldsymbol{\theta}$  for any order used, and no connections have to be masked out from  $\boldsymbol{\theta}$  to the rest of the network. Therefore, the Neural Likelihood Estimator (NLE) builds the flow-based model with MAF as [165]

$$q_\phi(\mathbf{x}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{z}_0|\mathbf{0}, \mathbf{I}) \prod_{t=1}^T \left| \det \left( \frac{\partial f_t}{\partial \mathbf{z}_{t-1}} \right) \right|^{-1}, \quad (3.6)$$

where the network parameters are trained with the loss given by Eq. (3.5).

Note that here the model is trained for *any given data*  $\mathbf{x}$ , where for inference we are interested in a specific observed data  $\mathbf{x}_o$ . The procedure is then to use the likelihood as the estimated conditional evaluated at the observed data, and then use MCMC (which is very cheap since only the learned model has to be evaluated) together with a prior  $p(\boldsymbol{\theta})$  to obtain the estimated posterior  $\hat{p}(\boldsymbol{\theta}|\mathbf{x}_o) \propto q_\phi(\mathbf{x}_o|\boldsymbol{\theta})p(\boldsymbol{\theta})$ .

When simulating the data is expensive, it is often desirable to use *sequential* approaches such as in Sequential Neural Likelihood Estimator (SNLE) [165], where the density estimation is done over multiple *rounds* focusing on a given observation  $\bar{\mathbf{x}}$ . The first proposal  $\tilde{p}(\boldsymbol{\theta})$  is often chosen to be the prior, while the subsequent proposals are set as the estimated posterior  $\hat{p}(\boldsymbol{\theta}|\bar{\mathbf{x}})$  from the previous round. This of course limits the density estimation to a single observation and is therefore *non-amortized*, but has the advantage of requiring less simulations to obtain an accurate posterior estimation focused on  $\mathbf{x}_o$ , as the regions where the proposal density is high tend to be better approximated by the model. However, the cost of calibrating non-amortized methods is huge, as discussed below.

**Neural Posterior Estimator.** An advantage of the SNLE method is that the prior can be changed and tested, but it has the MCMC as an extra computational step. This can be avoided by estimating the posterior directly, as in the case of Sequential Neural Posterior Estimator (SNPE) [85, 138, 162]. The estimated posterior has the form

$$\hat{p}(\boldsymbol{\theta}|\mathbf{x}) = p(\boldsymbol{\theta}|\mathbf{x}) \frac{\tilde{p}(\boldsymbol{\theta})p(\mathbf{x})}{p(\boldsymbol{\theta})\tilde{p}(\mathbf{x})}, \quad (3.7)$$

where  $\tilde{p}(\mathbf{x}) = \int_{\boldsymbol{\theta}} \tilde{p}(\boldsymbol{\theta})\tilde{p}(\mathbf{x}|\boldsymbol{\theta})$ . Note that the *proposal posterior*  $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$  is equal to the target posterior if the proposal  $\tilde{p}(\boldsymbol{\theta})$  equals the prior  $p(\boldsymbol{\theta})$ .

In its first implementation [162], the undesired dependence on the proposal  $\tilde{p}(\boldsymbol{\theta})$  was adjusted by multiplying the estimated posterior  $\hat{p}(\boldsymbol{\theta}|\mathbf{x})$  by  $p(\boldsymbol{\theta})/\tilde{p}(\boldsymbol{\theta})$ . This restricts  $q_\phi(\boldsymbol{\theta}|\mathbf{x})$  to be a Mixture Gaussian Density [39] and  $\tilde{p}(\boldsymbol{\theta})$  to be a Gaussian, so that the division can be done analytically; additionally,  $p(\boldsymbol{\theta})$  can only be Gaussian or uniform. Besides these strong requirements, a drawback of this model is the fact that the division can lead to non-positive covariance matrices if  $\tilde{p}(\boldsymbol{\theta})$  has smaller variance than any of the components of  $q_\phi(\boldsymbol{\theta}|\mathbf{x})$ .

A further extension to this method was presented in [138], where the aforementioned requirements are no longer needed and the method no longer yields negative covariances. This is achieved by weighting the samples with  $w_n = p(\boldsymbol{\theta}_n)/\tilde{p}(\boldsymbol{\theta}_n)$ , such that the loss is

modified to  $-\sum_n w_n \log q_\phi(\mathbf{x}_n|\boldsymbol{\theta}_n)$ . However, the weights can have high variance, leading to instability during the training process and inaccurate predictions in some cases [85]. The algorithm developed in [85] circumvents all these issues with the idea of ‘‘atomic’’ proposals, that essentially replaces integrals by sums and allows for the usage of arbitrary density estimators, especially those based on normalizing flows such as MAFs.

**Neural Ratio Estimator.** In Neural Ratio Estimator (NRE) [65, 70, 95, 145, 146], a classifier is used instead of a normalizing flow. In principle, classifier is easier to train than a flow. To extract a likelihood-to-evidence ratio, we train a classifier to distinguish between samples drawn jointly vs marginally,

$$p(\mathbf{x}, \boldsymbol{\theta} | y) = \begin{cases} p(\mathbf{x}, \boldsymbol{\theta}) & \text{if } y = 1 \\ p(\mathbf{x})p(\boldsymbol{\theta}) & \text{if } y = 0 \end{cases} \quad (3.8)$$

The posterior for the ‘‘switching’’ variable  $y$  is

$$p(y = 1 | \mathbf{x}, \boldsymbol{\theta}) = \frac{p(\mathbf{x}, \boldsymbol{\theta} | y = 1)}{p(\mathbf{x}, \boldsymbol{\theta} | y = 0) + p(\mathbf{x}, \boldsymbol{\theta} | y = 1)} = \frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta}) + p(\mathbf{x}, \boldsymbol{\theta})}. \quad (3.9)$$

If the likelihood to evidence ratio is

$$r(\mathbf{x}, \boldsymbol{\theta}) = \frac{p(\mathbf{x}, \boldsymbol{\theta} | y = 1)}{p(\mathbf{x}, \boldsymbol{\theta} | y = 0)} = \frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\mathbf{x} | \boldsymbol{\theta})}{p(\mathbf{x})}, \quad (3.10)$$

we have that

$$p(y = 1 | \mathbf{x}, \boldsymbol{\theta}) = \frac{r(\mathbf{x}, \boldsymbol{\theta})}{r(\mathbf{x}, \boldsymbol{\theta}) + 1} = \sigma(\log r(\mathbf{x}, \boldsymbol{\theta})). \quad (3.11)$$

The training of the classifier  $d_\psi(\mathbf{x}, \boldsymbol{\theta})$  proceeds by minimizing the binary cross-entropy loss (BCE) and it can be used in MCMC to obtain samples from the posterior by using the density ratio [95].

### 3.4.3 Normalizing flows

Statistical inference has gained power with the advent of normalizing flows, a class of generative models that allows one to model complex probabilistic distributions. The key idea behind this method is to learn a map that transforms a simple, well-understood density such as a Gaussian to the target probability distribution through successive transformations. Let  $\mathbf{x}$  be a  $D$ -dimensional real vector and  $p(\mathbf{x})$  its joint distribution we wish to estimate. We proceed by defining a transformation  $f$  from a base distribution  $\pi(\mathbf{u})$  such that

$$\mathbf{x} = f(\mathbf{u}), \quad \mathbf{u} \sim \pi(\mathbf{u}). \quad (3.12)$$

We further require that  $f$  is a *diffeomorphism*, i.e., that  $f$  invertible and that  $f$  and  $f^{-1}$  are differentiable, which restricts  $\mathbf{u}$  to be  $D$ -dimensional as well [147]. The distribution

over  $\mathbf{x}$  can therefore be obtained by the change of variables [42, 188]

$$p(\mathbf{x}) = \pi\left(f^{-1}(\mathbf{x})\right) \left| \det \left( \frac{\partial f^{-1}}{\partial \mathbf{x}} \right) \right|. \quad (3.13)$$

The composable property of diffeomorphisms, i.e., the fact that for any two diffeomorphisms  $f_1$  and  $f_2$  the composition  $f_1 \circ f_2$  is also diffeomorphic, guarantees that Eq. (3.13) is still valid if we build a complex transformation from a series of simple transformations  $f = f_T \circ \dots \circ f_1$ , where  $f_t$  transforms  $\mathbf{z}_{t-1}$  into  $\mathbf{z}_t$ , and setting  $\mathbf{z}_0 = \mathbf{u}$  and  $\mathbf{z}_T = \mathbf{x}$ . *Flow* therefore refers to these successive transformations applied on the samples from  $\pi(\mathbf{u})$  to progressively deform them into the ones of  $p(\mathbf{x})$ , while *normalizing* refers to the inverse flow  $f_T^{-1} \circ \dots \circ f_1^{-1}$  that maps the samples from the distribution  $p(\mathbf{x})$  into the ones of  $\pi(\mathbf{u})$ , which is often chosen to be a multivariate *normal* [163].

The fact that normalizing flows provide explicit densities (via Eq. 3.13) in addition to sampling (via Eq. 3.12) is in contrast to other generative models like Variational Autoencoders (VAEs) [114] and Generative Adversarial Networks (GANs) [84], where usually only sampling is possible. In practice,  $f$  (or  $f^{-1}$ ) is implemented as a neural network parametrized by  $\phi$  (e.g., its weights and biases), which is optimized to learn the mapping between the distribution parameters and a given simulated dataset. The model should have tractable inverse and Jacobian determinant; in other words, it is often required that the inverse is efficient to calculate and that the Jacobian determinant time cost should be at most  $\mathcal{O}(D)$ .

**Masked Autoregressive Flows.** A simple way to achieve the desirable properties of the model is to use *autoregressive flows*, which restrict  $f_t$  to be of the form [99]

$$\mathbf{z}'_i = \tau(z_i; \mathbf{h}_i), \quad \mathbf{h}_i = c_i(\mathbf{z}_{<i}). \quad (3.14)$$

Here, the output  $\mathbf{z}'_i$  is the  $i$ -th component of a vector  $\mathbf{z}_t \in \mathbb{R}^D$  for which  $\mathbf{z}_t = f_t(\mathbf{z}_{t-1})$ , while  $z_i$  is the  $i$ -th component of  $\mathbf{z}_{t-1} \in \mathbb{R}^D$ . The transformer  $\tau$  is required to be a strictly monotonic function of  $z_i$ , so that the model is invertible. Since  $\tau$  is parametrized by  $\mathbf{h}_i$ , the output  $\mathbf{z}'_i$  does not depend on  $z_{\geq i}$  due to the *autoregressive* structure of the  $i$ -th conditioner  $c_i$ . As a consequence, the Jacobian is triangular and its determinant can be calculated as the product of its diagonal elements, leading to an  $\mathcal{O}(D)$  evaluation of the logarithm of the determinant [163]:

$$\log \left| \det \left( \frac{\partial f}{\partial \mathbf{x}} \right) \right| = \log \left| \prod_{i=1}^D \frac{\partial \tau}{\partial z_i}(z_i; \mathbf{h}_i) \right| = \sum_{i=1}^D \log \left| \frac{\partial \tau}{\partial z_i}(z_i; \mathbf{h}_i) \right|. \quad (3.15)$$

The implementation of an autoregressive flow then follows by choosing an appropriate transformer  $\tau$  and conditioner  $c_i$ . A common choice for the transformer is to use affine functions, i.e.,

$$\tau(z_i; \mathbf{h}_i) = z_i \exp \alpha_i + \beta_i, \quad \mathbf{h}_i = \{\alpha_i, \beta_i\}, \quad (3.16)$$

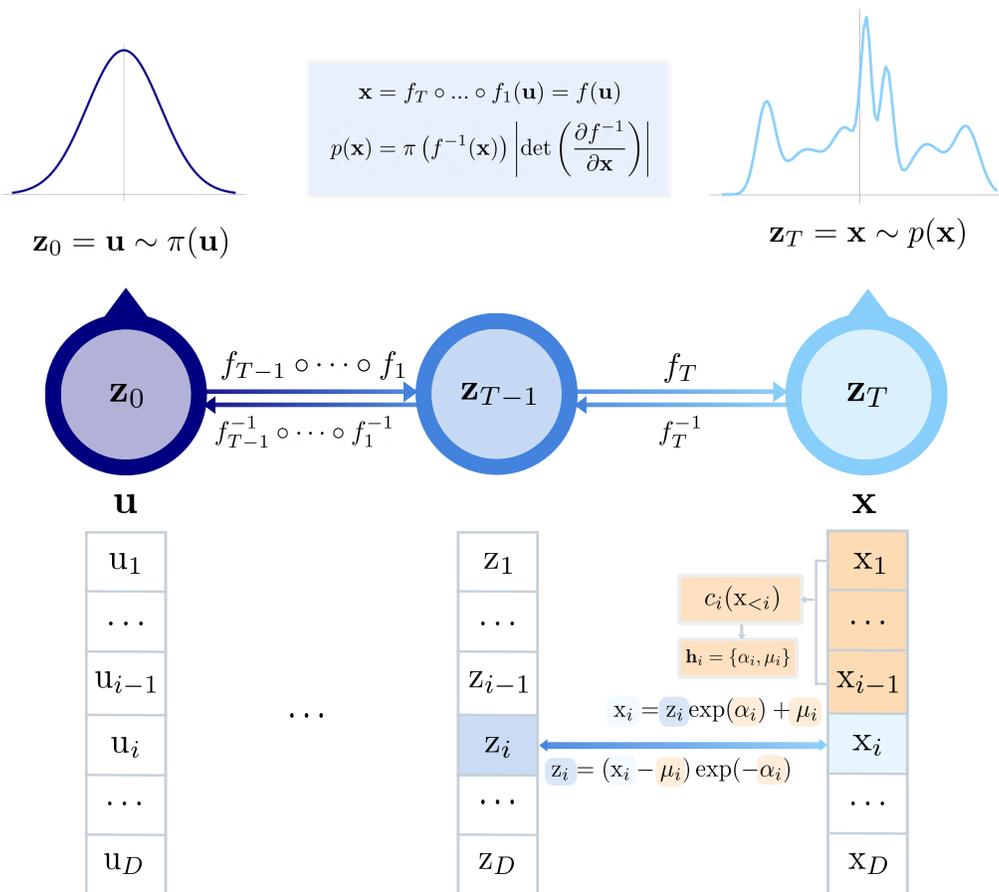


Figure 3.2: Diagram of how normalizing flows work, with the specific example of Masked Autoregressive Flows. The samples from the vector  $\mathbf{z}_0 = \mathbf{u}$ , sampled from the simple distribution  $\pi(\mathbf{u})$ , are deformed through the sequence of transformations  $f = f_T \circ \dots \circ f_1$  into those of  $\mathbf{z}_T = \mathbf{x}$ , which follow a more complex distribution  $p(\mathbf{x})$ . In the lower panel, we illustrate the conditioner that “masks out” the connections between  $z_i$  and  $\mathbf{h}_{<i}$ , as well as the affine functions applied to the vector components.

where we choose to parametrize the scale parameter of this location-scale transformation as  $\exp \alpha_i$  to guarantee its invertibility. The determinant of the Jacobian is then simply

$$\left| \det \left( \frac{\partial f^{-1}}{\partial \mathbf{x}} \right) \right| = \exp \left( - \sum_{i=1}^D \alpha_i \right). \quad (3.17)$$

We focus on Masked Autoregressive Flows (MAFs) [164] (see the flowchart Fig. 3.2), a particular case of autoregressive flows which will be used throughout this paper for density estimation, but other alternatives exist such as Inverse Autoregressive Flow (IAF) [113] and Real NVP [67]. MAFs are a special case of *autoregressive models* [218], which use the product rule of probability to decompose densities into 1-dimensional ones as

$$p(\mathbf{x}) = \prod_{i=1}^D p(x_i | \mathbf{x}_{<i}). \quad (3.18)$$

In MAFs, the conditionals are parametrized by Gaussians,

$$p(x_i | \mathbf{x}_{<i}) = \mathcal{N}(x_i; \mu_i, (\exp \alpha_i)^2), \quad \mathbf{h}_i = \{\alpha_i, \mu_i\}, \quad (3.19)$$

which can be equivalently written as an affine function (Eq. 3.16)

$$x_i = u_i \exp \alpha_i + \mu_i, \quad u_i = \mathcal{N}(0, 1), \quad (3.20)$$

where the base distribution is normal. Given that the model is easily invertible, with  $u_i = (x_i - \mu_i) \exp(-\alpha_i)$ , and it has a triangular Jacobian of the form of Eq. (3.17), autoregressive models can also be interpreted as a normalizing flow [113], as Eqs. (3.12) and (3.13) can be evaluated.

The MAF conditioner  $c_i$  in turn is implemented with *masked conditioners* as in Masked Autoencoder for Distribution Estimation (MADE) [81], where a single feedforward network outputs  $\mathbf{h}$  in one pass given the input  $\mathbf{z}$ . To preserve the autoregressive property of the conditioner, each weight matrix element is multiplied by a binary mask, where the connections between  $z_i$  and  $\mathbf{h}_{<i}$  are “masked out” by essentially multiplying their corresponding weights by zero.

The interpretation of autoregressive models as normalizing flows enables MAF to stack multiple layers of MADEs with Gaussian conditionals into a deeper flow, where the output of each layer,  $\alpha_i$  and  $\mu_i$  for all  $i$ , are used as input for the next one. In the language of the previous section, we will have that a MAF transforms the basis function  $\pi(\mathbf{u}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$  to the target density  $p(\mathbf{x})$  through the composition of transformations  $f = f_T \circ \dots \circ f_1$ , where each  $f_t$  is implemented by a MADE for  $\mathbf{z}_0 = \mathbf{u}$ ,  $\mathbf{z}_t = f_t(\mathbf{z}_{t-1})$  and  $\mathbf{z}_T = \mathbf{x}$ .

As a result of the stacking, the model is still tractable, but more expressive and flexible. For example, it has been shown that a MAF with 5 autoregressive unimodal conditional layers is able to approximate multimodal posteriors [164]. Another subtlety is that the estimated densities can depend on the order of the inputs, so MAF avoids this issue by using different input orders for each layer.

### 3.4.4 Simulation-based calibration

When using approximated methods for Bayesian inference, a lot of effort has to be made to analyse whether the obtained posterior is correct. For example, for MCMC one has to check whether the chains have converged, and to this date there exists no method to assure convergence with complete confidence [96]. Regarding SBI algorithms, an important diagnostic is to check whether the posterior is calibrated, i.e., not over- or under-estimating the parameter uncertainties. For scientific reasoning, we are mostly concerned with assuring that the posterior is not superconfident. Simulation-based calibration (SBC) [210] arises as a useful tool for this analysis, where it is important to stress that passing this test is only a sufficient condition; that is, if it fails, it is an indication that the training was not successful (for example, if not enough simulations were used), while if it passes there is no guarantee that the posterior is correct. Providing reliable tests for SBI algorithms is an active area of research [96, 127, 136, 230].

Here, we use SBC to analyse the estimated posteriors, and its results can tell us if the posterior is not well-calibrated and indicate possible systematic biases on the inference. The idea behind SBC is that, if we draw samples  $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$  from the prior and then generate a data vector  $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\theta})$  with the simulator, by drawing samples  $\hat{\boldsymbol{\theta}} \sim \hat{p}(\hat{\boldsymbol{\theta}}|\mathbf{x})$  from the estimated posterior approximated by the model  $q_\phi(\hat{\boldsymbol{\theta}}|\mathbf{x})$ , we obtain joint distribution  $\mu(\boldsymbol{\theta}, \mathbf{x}, \hat{\boldsymbol{\theta}}) = p(\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})\hat{p}(\hat{\boldsymbol{\theta}}|\mathbf{x})$ . The marginal distribution of  $\hat{\boldsymbol{\theta}}$ ,

$$\begin{aligned} \mu(\hat{\boldsymbol{\theta}}) &= \int d\mathbf{x} \int d\boldsymbol{\theta} p(\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})\hat{p}(\hat{\boldsymbol{\theta}}|\mathbf{x}) \\ &= \int d\mathbf{x} \int d\boldsymbol{\theta} p(\mathbf{x}, \boldsymbol{\theta})\hat{p}(\hat{\boldsymbol{\theta}}|\mathbf{x}) \\ &= \int d\mathbf{x} p(\mathbf{x})\hat{p}(\hat{\boldsymbol{\theta}}|\mathbf{x}), \end{aligned} \tag{3.21}$$

should be distributed according to the prior if the estimated posterior equals the target distribution; that is, if  $\hat{p}(\hat{\boldsymbol{\theta}}|\mathbf{x}) = p(\hat{\boldsymbol{\theta}}|\mathbf{x})$ , then

$$\mu(\hat{\boldsymbol{\theta}}) = \int d\mathbf{x} p(\mathbf{x}, \hat{\boldsymbol{\theta}}) = p(\hat{\boldsymbol{\theta}}). \tag{3.22}$$

Analysing the distribution of  $\hat{\boldsymbol{\theta}}$  can therefore be used as a posterior calibration test. In practice, we generate a set of “observed” data  $\mathbf{x}_o^i$ , where  $\mathbf{x}_o^i \sim p(\mathbf{x}|\boldsymbol{\theta}_i)$  is simulated given a particular  $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$  drawn from the prior, and then estimate the posterior  $\hat{p}(\boldsymbol{\theta}|\mathbf{x}_o^i)$  for each of these observations. We then draw a set of posterior samples  $\{\hat{\boldsymbol{\theta}}\}_i$  for each of the estimated posteriors and compute the *rank* of the observed data under this set by counting how many of the posterior samples  $\{\hat{\boldsymbol{\theta}}\}_i$  fall below the corresponding observed data  $\mathbf{x}_o^i$ . If the rank statistics is not uniformly distributed, then the estimated posterior is not well calibrated.

## 3.5 Data compression techniques

Since normalizing flows suffer from the curse of dimensionality, it is often desirable to compress the data  $\mathbf{x}$  to a lower-dimensional summary statistic  $\mathbf{z}$ . Then the density estimation is performed in the joint space  $\{\mathbf{z}, \boldsymbol{\theta}\}$ , rather than in the high-dimensional space  $\{\mathbf{x}, \boldsymbol{\theta}\}$ . While sub-optimal compression may increase the scatter in the  $\{\mathbf{z}, \boldsymbol{\theta}\}$  space, leading to broader posterior constraints, it does not result in biased inference. As a rule of thumb, it is usually required that  $\mathbf{z}$  has at least the same dimensionality as  $\boldsymbol{\theta}$ . We will review below a few techniques for performing data compression, following [106]. In App. C, we show how these compression techniques can be used for cosmological inference.

### 3.5.1 MOPED

Let us denote our parameters of interest as  $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^{N_\theta}$  and the data vector we want to compress as  $\mathbf{x} = \{\mathbf{x}_j\}_{j=1}^D$ . The MOPED algorithm [93, 182] is a data compression technique that reduces the dimensionality of a dataset from  $D$  to the number of parameters being constrained,  $N_\theta$ . This is achieved by projecting the data vector onto a set of carefully constructed basis vectors, known as MOPED basis vectors  $\mathbf{M}_i$ . These vectors are optimized to maximize the sensitivity of the likelihood function to each of the parameters.

To construct the compression vectors, the derivative of the data vector with respect to the model parameters is weighted by the inverse of the data covariance matrix. The resulting vectors are then orthonormalized using the Gram-Schmidt procedure. The compression matrix  $\mathbf{M}_{(N_\theta \times D)}$  has  $N_\theta$  rows and  $D$  columns, with elements given by  $M_{ij}$ . We can define its rows (the MOPED basis vectors) as

$$\mathbf{M}_i = \begin{cases} C^{-1}\mathbf{x}_{,i} & \text{if } i = 1, \\ \frac{C^{-1}\mathbf{x}_{,i} - \sum_{k=1}^{i-1}(\mathbf{x}_{,i}^T\mathbf{M}_k)\mathbf{M}_k}{\sqrt{\mathbf{x}_{,i}^T C^{-1}\mathbf{x}_{,i} - \sum_{k=1}^{i-1}(\mathbf{x}_{,i}^T\mathbf{M}_k)^2}} & \text{if } 1 < i \leq N_\theta, \end{cases} \quad (3.23)$$

where  $C$  is the covariance matrix of the data vector and  $\mathbf{x}_{,i}$  denotes the derivative of the data vector with respect to the parameter  $\theta_i$ . The Gram-Schmidt process is applied to ensure the vectors  $\{\mathbf{M}_i\}$  are orthonormal. As demonstrated in [93] and further supported by [94], MOPED provides an optimal and lossless compression for Gaussian-distributed data, provided the covariance matrix is independent of the parameters.

### 3.5.2 MSE

There are multiple strategies for training the neural compressor to produce informative summary statistics. A straightforward method involves training the network as a regressor for the parameters, optimizing either an  $\ell_2$  (mean squared error, MSE) or an  $\ell_1$  loss. These loss functions correspond to learning the posterior mean and median, respectively. Previous

work by [17] demonstrated that for a known likelihood, optimal compression is lossless at a fiducial value of the parameter  $\boldsymbol{\theta}$ . In cases where the likelihood is Gaussian and the covariance is independent of  $\boldsymbol{\theta}$ , the optimal score compression becomes linear, equivalent to MOPED compression. Generally, optimal score compression yields summary statistics equivalent in constraining power to the maximum likelihood estimate under the assumed likelihood, i.e.,  $\mathbf{z} = F(\mathbf{x}) = \boldsymbol{\theta}_{\text{MLE}}$ .

Rather than relying on an explicit likelihood, one can use a neural network-based compressor  $\mathbf{z} = F_\varphi(\mathbf{x})$ , where the function  $F_\varphi$  is trained on a dataset  $\{\mathbf{x}^n, \boldsymbol{\theta}^n\}$ . The network is optimized to minimize the following loss function with respect to the network parameters  $\varphi$ :

$$\mathcal{L}_{\ell_2}(\varphi) = \|F_\varphi(\mathbf{x}^n) - [\mathbf{A}\boldsymbol{\theta}^n + b]\|_2^2, \quad (3.24)$$

where  $\mathbf{A}$  and  $b$  are fixed scaling and shift parameters applied to normalize the targets for efficient training, depending on the neural network architecture.

While score compression is equivalent to maximum likelihood estimation for an assumed likelihood, the training objective corresponds to a mean posterior estimate, but without assuming a specific likelihood and using an implicit prior derived from the training labels  $\boldsymbol{\theta}^n$ . An alternative loss function that is commonly used is the  $\ell_1$  norm,

$$\mathcal{L}_{\ell_1}(\varphi) = \|F_\varphi(\mathbf{x}^n) - [\mathbf{A}\boldsymbol{\theta}^n + b]\|_1, \quad (3.25)$$

which minimizes the least absolute deviation (LAD) or mean absolute error (MAE), and is theoretically equivalent to estimating the posterior median. However, it tends to yield a more lossy compression [106].

### 3.5.3 VMIM

The regression-based approaches mentioned in the previous section do not necessarily guarantee that the resulting summary statistics are maximally informative. Under specific conditions, such as fixed fiducial parameter values and the assumption of Gaussianity, this might hold true, but it does not apply in the general case.

An alternative and more principled method is to train the deep compressor by directly maximizing the mutual information  $I(\mathbf{z}, \boldsymbol{\theta})$  between the summary statistics  $\mathbf{z}$  and the parameters  $\boldsymbol{\theta}$ . Mutual information serves as a measure of the amount of information gained about  $\boldsymbol{\theta}$  from observing  $\mathbf{z}$ , thereby providing a more robust objective for learning informa-

tive statistics. We can define the mutual information formally with the KL divergence,

$$\begin{aligned}
I(\mathbf{z}, \boldsymbol{\theta}) &= D_{\text{KL}} [p(\mathbf{z}, \boldsymbol{\theta}) \parallel p(\mathbf{z})p(\boldsymbol{\theta})] \\
&= \int d\boldsymbol{\theta} d\mathbf{z} p(\mathbf{z}, \boldsymbol{\theta}) \log \left( \frac{p(\mathbf{z}, \boldsymbol{\theta})}{p(\mathbf{z})p(\boldsymbol{\theta})} \right) \\
&= \int d\boldsymbol{\theta} d\mathbf{z} p(\mathbf{z}, \boldsymbol{\theta}) \log \left( \frac{p(\boldsymbol{\theta}|\mathbf{z})}{p(\boldsymbol{\theta})} \right) \\
&= \int d\boldsymbol{\theta} d\mathbf{z} p(\mathbf{z}, \boldsymbol{\theta}) \log p(\boldsymbol{\theta}|\mathbf{z}) - \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}) \\
&= \mathbb{E}_{p(\mathbf{z}, \boldsymbol{\theta})} [\log p(\boldsymbol{\theta}|\mathbf{z})] - \mathbb{E}_{p(\boldsymbol{\theta})} [\log p(\boldsymbol{\theta})] \\
&= \mathbb{E}_{p(\mathbf{z}, \boldsymbol{\theta})} [\log p(\boldsymbol{\theta}|\mathbf{z})] - H(\boldsymbol{\theta}).
\end{aligned} \tag{3.26}$$

In the context of data compression, a compressed summary  $\mathbf{z}$  is obtained through a neural network  $F_\varphi$ ,  $\mathbf{z} = F_\varphi(\mathbf{x})$ , yielding

$$\varphi^* = \arg \max_{\varphi} I(F_\varphi(\mathbf{x}), \boldsymbol{\theta}). \tag{3.27}$$

The architecture of  $F_\varphi$  can be chosen freely, depending on the problem at hand. If the data vector consists of an image, usually a 3D convolutional neural network (CNN) is used. For one dimensional vectors, a fully connected network or 1D CNN can be employed.

The mutual information in Eq. 3.26 is generally intractable, and its estimation remains an open challenge in both statistics and machine learning. One can instead use tractable lower bounds on the mutual information, enabling its optimization in scenarios such as the training of deep neural networks. One choice is the variational lower bound introduced by [27],

$$I(\mathbf{z}, \boldsymbol{\theta}) \geq \mathbb{E}_{p(\mathbf{z}, \boldsymbol{\theta})} [\log q_\phi(\boldsymbol{\theta}|\mathbf{z})] - H(\boldsymbol{\theta}). \tag{3.28}$$

This bound relies on a variational conditional distribution  $q_\phi(\boldsymbol{\theta}|\mathbf{z})$  that approximates the true posterior  $p(\boldsymbol{\theta}|\mathbf{z})$ , for example using a normalizing flow, and becomes an equality when  $q_\phi(\boldsymbol{\theta}|\mathbf{z})$  exactly matches  $p(\boldsymbol{\theta}|\mathbf{z})$ . Taking into account that the entropy  $H(\boldsymbol{\theta})$  is constant for a fixed training set, the mutual information optimization objective can be stated as

$$\arg \max_{\varphi, \phi} \mathbb{E}_{p(\mathbf{x}, \boldsymbol{\theta})} [\log q_\phi(\boldsymbol{\theta}|F_\varphi(\mathbf{x}))]. \tag{3.29}$$

This approach is called Variational Mutual Information Maximization (VMIM). The optimization loss is

$$\mathcal{L}_{\text{VMIM}}(\varphi, \phi) = -\frac{1}{N_{\text{sim}}} \sum_{n=1}^{N_{\text{sim}}} \log q_\phi(\boldsymbol{\theta}_n|F_\varphi(\mathbf{x}_n)), \tag{3.30}$$

which has the same structure as the NPE loss function. One can then embed the neural network  $F_\varphi$  to a normalizing flow and train with NPE loss function, where we optimize together both the weights of the compression network  $\varphi$  and of the normalizing flow  $\phi$  by gradient descent. After training, you obtain a compressor  $F_\varphi$  which outputs a summary statistics which has maximum mutual information with respect to the parameters of interest.

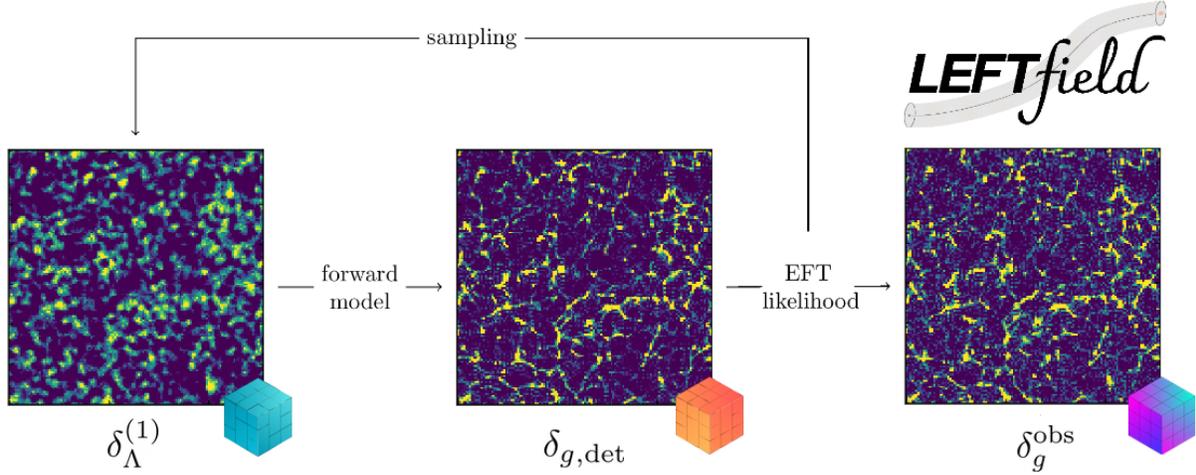


Figure 3.3: Schematic representation of field-level inference through the forward model `LEFTfield`. The initial conditions  $\delta_{\Lambda}^{(1)}$  are evolved with the forward model to obtain  $\delta_{g,\text{det}}$ , which is further used in the EFT likelihood together with  $\delta_{g,\text{obs}}$ . New parameters are proposed and the process is repeated to sample the posterior until convergence. *Credits: Julia Stadler.*

### 3.6 Field-level inference

In field-level inference, the goal is to infer the underlying cosmological and bias parameters  $\boldsymbol{\theta}$  together with the initial conditions directly from observed large-scale structure data, by working at the level of the entire galaxy density field. The posterior distribution for the parameters is given by

$$\mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}] = \int \mathcal{D}\delta_{\Lambda}^{(1)} \mathcal{P}_{\text{prior}}[\boldsymbol{\theta}, \delta_{\Lambda}^{(1)}] \mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}, \delta_{\Lambda}^{(1)}], \quad (3.31)$$

where the EFT likelihood  $\mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}, \delta_{\Lambda}^{(1)}]$  was derived in Sec. 2.4.4, which takes form for discrete  $k$

$$\ln \mathcal{P}[\delta_g^{\text{obs}}|\boldsymbol{\theta}, \delta_{\Lambda}^{(1)}] = -\frac{1}{2} \sum_{k \neq 0}^{k_{\text{max}}} \left[ \frac{1}{\sigma_{\varepsilon}^2(k)} |\delta_g(\mathbf{k}) - \delta_{g,\text{det}}[\boldsymbol{\theta}, \delta_{\Lambda}^{(1)}](\mathbf{k})|^2 + \ln[2\pi\sigma_{\varepsilon}^2(k)] \right]. \quad (3.32)$$

The practical steps involved in this procedure are as follows (see Fig. 3.3):

- **Discretization:** Discretize the density field on a grid or lattice, with the Nyquist frequency setting the cutoff scale  $\Lambda$ .
- **Sampling Initial Conditions:** Draw realizations of the initial density field  $\delta_{\text{in}}$  from the prior distribution, assumed to be Gaussian in the absence of PNGs.
- **Forward Evolution:** Evolve the initial field forward in time using gravitational evolution with `LEFTfield` to obtain  $\delta_{g,\text{det}}$ .

- **Likelihood Evaluation:** Compute the likelihood by comparing the evolved density field to the observed galaxy field.
- **Iteration:** Repeat the process to sample from the posterior distribution of the cosmological parameters and the initial conditions until convergence is achieved.

One major challenge of field-level inference is the extremely high dimensionality of the parameter space: even at relatively coarse grid resolutions, one must sample millions of parameters (the values of  $\delta_{\Lambda}^{(1)}$  at each grid point).

Efficient sampling is therefore crucial. A key technique employed is Hamiltonian Monte Carlo (HMC), which leverages gradient information to efficiently explore the high-dimensional posterior, enabling practical inference despite the large number of parameters. It requires the gradients of the operators with respect to the initial conditions, which is also straightforwardly computed by `LEFTfield`. In the next section, we give an overview on how this sampling algorithm works.

### 3.6.1 Hamiltonian Monte Carlo

**Hamiltonian Dynamics.** The Hamiltonian  $H(q, p)$  is a function of the  $d$ -dimensional position  $q$  and  $d$ -dimensional momentum  $p$  of a given system. The Equations of Motion (EoM) read

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad (3.33)$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad (3.34)$$

for  $i = 1, \dots, d$ . These define a mapping  $T_s$  from the state of the system from time  $t$  to  $t + s$  [153].

First, the mapping  $T_s$  is *reversible*, and its inverse  $T_{-s}$  can be obtained by inverting the signs of Eqs. 3.34. This will be important to show that MCMC updates that use the dynamics leave the desired distribution invariant.

Hamiltonian dynamics are *conserved*, i.e., the Hamiltonian is invariant under the dynamics. We will see that acceptance probability is one if  $H$  is kept invariant, although in MCMC we will not be able to keep it strictly invariant.

By Liouville's Theorem, we have that *volume preservation* in phase space, i.e.,  $(q, p)$  space [21]. This is a huge advantage since we don't have to account for a change of volume in the acceptance probability for Metropolis updates. Otherwise, we would have to calculate the Jacobian of the transformation, what might turn to be unfeasible.

Finally, we have the *symplecticness* of Hamiltonian dynamics. For the Jacobian  $J$ , defined in the analogous formulation of the EoM,

$$\frac{dz}{dt} = J \nabla H(z), \quad (3.35)$$

where  $z = (q, p)$ , we have that the *symplecticness* condition is such that the determinant  $B_s$  of the mapping  $T_s$  satisfies

$$B_s^T J^{-1} B_s = J^{-1}, \quad (3.36)$$

from which follows that  $\det(B_s)^2 = 1$ .

**The Leapfrog Method.** Now we will discretize and solve the Hamiltonian equations in each timestep  $\varepsilon$  with the *Leapfrog method*. We assume that

$$H(q, p) = U(q) + K(p), \quad (3.37)$$

where  $U(q)$  is the *potential energy* of the system and  $K(p)$  is the *kinetic energy*. We can assume that  $K(p) = p^T M^{-1} p$ , where  $M$  is a diagonal mass matrix, such that

$$K(p) = \sum_{i=1}^d \frac{p_i^2}{2m_i}. \quad (3.38)$$

The Leapfrog method constitutes in

$$p_i(t + \varepsilon/2) = p_i(t) - (\varepsilon/2) \frac{\partial U}{\partial q_i}(q(t)), \quad (3.39)$$

$$q_i(t + \varepsilon) = q_i(t) + \varepsilon \frac{p_i(t + \varepsilon/2)}{m_i}, \quad (3.40)$$

$$p_i(t + \varepsilon) = p_i(t + \varepsilon/2) - (\varepsilon/2) \frac{\partial U}{\partial q_i}(q(t + \varepsilon)), \quad (3.41)$$

where  $p_i/m_i$  can be replaced by  $\partial K/\partial p_i$  for a generic kinetic energy  $K$ . This method *preserves volume* exactly, since these are shear transformations. It is also *reversible*, since we can obtain the same results once we negate  $p$ , apply the same number of steps and then negate  $p$  again.

**The Hamiltonian Monte Carlo Algorithm.** If we want to use the Hamiltonian dynamics to sample from a distribution, we have to translate the *density function* for this distribution to a *potential energy* function, where the *original variables* of interests can now be seen as *position variables*, and the introduce the companion *momentum variables* to go with them. Then we will be able to simulate a Markov chain where each iteration resamples the momentum and does a Metropolis update with the proposal obtained by using Hamiltonian dynamics.

**MCMC from Hamiltonian dynamics.** Since the Hamiltonian defines a energy function of the joint state of  $q$  and  $p$ , we can write a joint *canonical distribution* over these states with probability density function

$$P(q, p) = \frac{1}{Z} e^{-H(q,p)/T}, \quad (3.42)$$

where  $Z$  is the normalization constant and  $T$  is the temperature of the system. Since  $H$  is invariant under Hamiltonian dynamics, we have that Hamiltonian trajectories move within a hyper-surface of constant probability density.

We can express the posterior distribution as a canonical distribution using the following definition of potential energy

$$U(q) = -\log[\pi(q)L(q|D)], \quad (3.43)$$

where  $\pi(q)$  is the prior density, and  $L(q|D)$  is the likelihood function given data  $D$ .

**The two steps of the HMC algorithm.** We can use HMC to sample from any  $R^d$  distribution for which the density function can be evaluated and for which the partial derivatives of its log exists. We will assume here that the probability density is non-zero everywhere.

We can freely choose the distribution of the momentum variables  $p$  which are independent of  $q$ , specifying the distribution via the kinetic energy function  $K(p)$ . It is usually used a quadratic kinetic energy as in Eq. 3.38, which leads  $p$  to have a zero-mean multivariate Gaussian distribution. Most often, the components  $p_i$  are specified to be independent with variance  $m_i$ .

In the first step, the new position variables are randomly chosen from its Gaussian distribution, independently from the position ones. This leaves the canonical joint distribution invariant, since it doesn't change  $q$  and draws  $p$  from its conditional distribution given  $q$  (its marginal distribution, since both variables are independent). In the second step, we use Hamiltonian dynamics to perform a Metropolis update. See the Implementation section.

**Illustration.** We will now show the benefits of HMC over the usual Metropolis algorithm to sample a given distribution. Here we will use Gaussian distributions, but usually the highest efficiency of HMC can be seen more clearly in more complex distributions [37].

Let's analyze the HMC sampling from a distribution for two variables that is bivariate Gaussian, with means of zero, standard deviations of one, and correlation 0.95.

We can see from Fig. 3.4 that the trajectory of our variable of interest, the position, is very different from a usual random walks. This avoidance of random walk motions is one of the greatest advantages of the HMC method over the Metropolis one [153]. The consistency of this motion, starting at the yellow dots, e.g. going above and then changing sign, is due to the presence of the momentum variable. Since the gradient is small, the projection of the momentum variable changes only slowly. Besides the large scale motion, there are some small oscillations back and forth due to the correlation between the variables.

Since we have to keep this oscillations over control, the timestep used has to be also small. A too large timestep, e.g.  $\varepsilon = 0.45$  in this case, would make the oscillations in the Hamiltonian to grow with no bounds. Note that if the discrete implementation was exact, the oscillations in the Hamiltonian should be constant, as a product of energy conservation under the Hamiltonian dynamics.

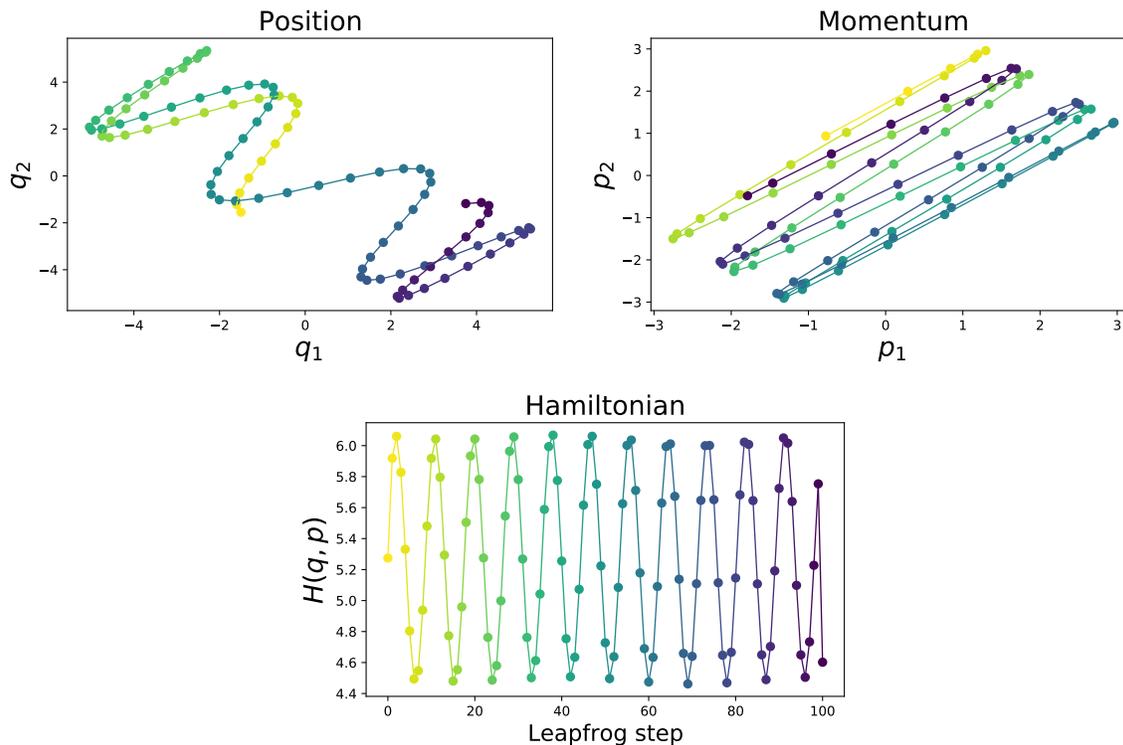


Figure 3.4:  $L = 100$  Leapfrog steps with size  $\varepsilon = 0.25$ . The initial state is  $q = (-1.50, -1.55)$  and  $p = (-0.78, 0.94)$ . The rejection was 0.05.

After changing the Leapfrog timestep size and number, we can turn the correlation to 0.98 and obtain the following results.

We can compare the performance of the HMC sampling over the usual random-walk Metropolis one [62]. In Figure 3.6 we see every 20th state from 400 iterations of random-walk Metropolis. We use a bivariate Gaussian proposal distribution with the current state as mean, zero correlation, and the same standard deviation for the two coordinates. We use the same stepsize of 0.18 used for HMC proposals for the standard deviation here, so that the change in state in these random-walk proposals is comparable to that for a single leapfrog step for HMC.

The most prominent difference is clearly the huge difference between rejection rates of the HMC and Metropolis algorithms; HMC had much lower rejection rates. Also, HMC produces a larger change of states than the Metropolis algorithm does. These facts, together with the aforementioned avoidance from random walks, illustrates the greatest benefits of HMC.

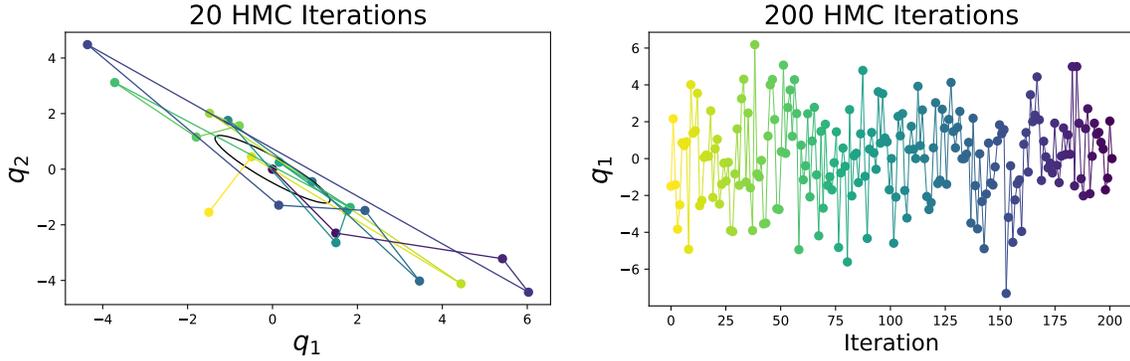


Figure 3.5:  $L = 20$  Leapfrog steps with size  $\varepsilon = 0.18$ . The initial state is  $q = (-1.50, -1.55)$ . The rejection was 0.025.

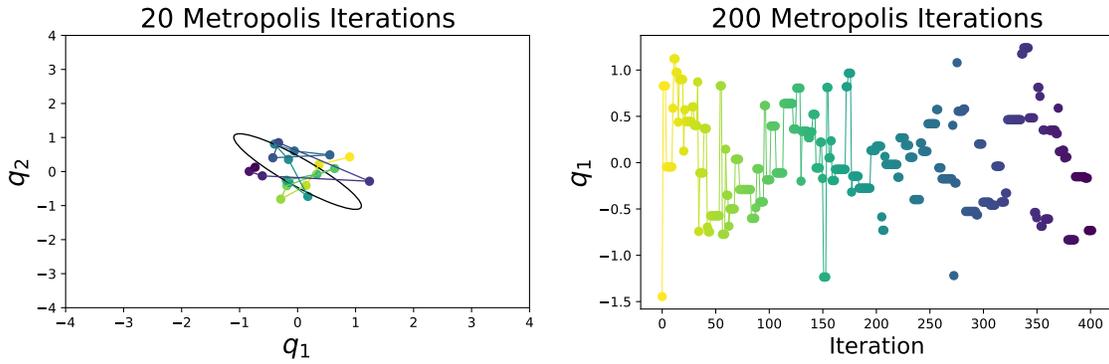


Figure 3.6:  $L = 100$  Leapfrog steps with size  $\varepsilon = 0.25$ . The initial state is  $q = (-1.50, -1.55)$ . The rejection was 0.39.

### 3.6.2 Marginalization over galaxy bias

The likelihood function in Eq. (3.32) can be rewritten as:

$$\mathcal{L}_{\text{FBI}}^{\text{expl.}} \left( \delta_g^{\text{obs.}} \mid \hat{s}, \alpha, \{b_O\}, \{\sigma_\epsilon\} \right) = \text{N exp} \left[ -\frac{1}{2} \sum_{k>0}^{k_{\text{max}}} \ln \sigma_\epsilon^2(k) \right] \times \exp \left\{ -\frac{1}{2} C + \sum_O b_O B_O - \frac{1}{2} \sum_{O,O'} b_O b_{O'} A_{OO'} \right\}, \quad (3.44)$$

where  $\mathbb{N}$  represents a normalization constant independent of parameters. Following [73], we define the scalar quantity  $C$ , the vector  $B_O$ , and the matrix  $A_{OO'}$  as

$$\begin{aligned} C &= \sum_{\mathbf{k}>0}^{k_{\max}} \frac{1}{\sigma_\epsilon^2(k)} \left| \delta_g^{\text{obs.}} - \sum_{O_{\text{unmarg.}}} b_{O_{\text{unmarg.}}} O_{\text{unmarg.}}(\mathbf{k}) \right|^2, \\ B_O &= \sum_{\mathbf{k}>0}^{k_{\max}} \frac{\left[ \delta_g^{\text{obs.}}(\mathbf{k}) - \sum_{O_{\text{unmarg.}}} b_{O_{\text{unmarg.}}} O_{\text{unmarg.}}(\mathbf{k}) \right] O^*(\mathbf{k})}{\sigma_\epsilon^2(k)}, \\ A_{OO'} &= \sum_{\mathbf{k}>0}^{k_{\max}} \frac{O(\mathbf{k}) O'^*(\mathbf{k})}{\sigma_\epsilon^2(k)}. \end{aligned} \quad (3.45)$$

Eq. (3.44) effectively isolates the  $n$  galaxy bias parameters  $\{b_O\}$  to be marginalized, leaving the rest  $\{b_O\}_{\text{unmarg.}}$  unmarginalized. Assuming Gaussian priors, integrating out  $\{b_O\}$  in Eq. (3.44) using Gaussian integrals results in

$$\begin{aligned} \mathcal{L}_{\text{FBI}}^{\text{expl.}} \left( \delta_g^{\text{obs.}} \mid \hat{s}, \alpha, \{b_O\}_{\text{unmarg.}}, \{\sigma_\epsilon\} \right) &= \left( \prod_O \int db_O \right) P \left( \delta_g^{\text{obs.}} \mid \hat{s}, \{b_O\} \right) \\ &= \mathbb{N}(2\pi)^{n/2} \left| A_{OO'} \right|^{-1/2} \exp \left[ -\frac{1}{2} \sum_{\mathbf{k}>0}^{k_{\max}} \ln \sigma_\epsilon^2(k) \right] \\ &\times \exp \left\{ -\frac{1}{2} C(\{b_O\}) + \frac{1}{2} \sum_{O,O'} B_O(\{b_O\}) (A^{-1})_{OO'} B_{O'}(\{b_O\}) \right\}. \end{aligned} \quad (3.46)$$

In the field-level inference approach, all bias coefficients  $\{b_O\}$  can be analytically marginalized, reducing the number of free parameters in the MCMC sampling. This enhances sampling efficiency by lowering the dimensionality of the parameter space. The validity of Eq. (3.46) has been tested on mock datasets, as documented in Appendix F of [118].

### 3.7 Inferring $\sigma_8$ from galaxies

The cosmological parameter  $\sigma_8$  quantifies the amplitude of matter density fluctuations in the Universe on a scale of  $8 h^{-1}$  megaparsecs (Mpc), where  $h$  is the dimensionless Hubble parameter. Specifically, it measures the root-mean-square (RMS) variation of the matter density field when smoothed as,

$$\sigma_8^2 \equiv \left\langle \left( \delta_R^{(1)}(\mathbf{x}) \right)^2 \right\rangle = \int_0^\infty \frac{dk k^2}{2\pi^2} P_L(k, z=0) W_R^2(k), \quad (3.47)$$

where  $P_L$  is the linear power spectrum and  $W_R$  is a spherical window function with radius  $R = 8 h^{-1} \text{Mpc}$ . A higher value of  $\sigma_8$  indicates a Universe with more pronounced clustering of matter, meaning structures like galaxies and clusters formed more efficiently. This parameter is critical in constraining models of cosmic structure formation and is closely

related to the growth of large-scale structure and the nature of dark matter and dark energy. This is evident when we look at the equation for structure growth,

$$D'' + aHD' = 4\pi G\bar{\rho}D, \quad (3.48)$$

as  $\sigma_8$  reflected on the matter density probes both the expansion history by  $H$  and gravitational effects by  $G$ .

The galaxy power spectrum contains

$$P_g(k) = b_1^2 P_L(k) \propto (b_1 \sigma_8)^2. \quad (3.49)$$

At leading order, this introduces a degeneracy between  $b_\delta$  and  $\sigma_8$ , which must be broken using nonlinear clustering or redshift-space distortions (RSD). However, RSD constraints are limited to the combination  $f\sigma_8$ , where  $f = d \ln D_+ / d \ln a$  is the growth rate.

Nonlinear clustering provides a means to break this degeneracy. The equivalence principle ensures that galaxies and matter co-move on large scales [66], leading to predictable advection contributions in nonlinear structures, thus enabling a direct inference of  $\sigma_8$ . Particularly, the galaxy bispectrum can be written as

$$B_g^{\text{LO}}(k_1, k_2, k_3) \equiv \langle \delta_h(\mathbf{k}_1) \delta_h(\mathbf{k}_2) \delta_h(\mathbf{k}_3) \rangle'_{\text{LO}} \quad (3.50)$$

$$= b_1^3 B_m^{\text{LO}}(k_1, k_2, k_3) + B_\varepsilon^{\{0\}} \quad (3.51)$$

$$+ \left\{ b_1^2 \left[ b_2 + 2b_{K^2} \left( (\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2)^2 - \frac{1}{3} \right) \right] P_L(k_1) P_L(k_2) + 2b_1 P_{\varepsilon\delta}^{\{0\}} P_L(k_2) \right\} + 2 \text{ perm..} \quad (3.52)$$

The leading-order matter bispectrum entering Eq. (3.52) is given by

$$B_m^{\text{LO}}(k_1, k_2, k_3) = 2F_2(\mathbf{k}_1, \mathbf{k}_2) P_L(k_1) P_L(k_2) + 2 \text{ perm.}, \quad (3.53)$$

where the  $F_2$  kernel is defined in Eq.(2.32). We can see then that

$$\begin{aligned} B_g^{\text{tree}}(k_1, k_2, k_3) &\supset \langle \delta^{(1)}(k_1) \delta^{(1)}(k_2) \delta^{(2)}(k_3) \rangle + 2 \text{perm.} \\ &= b_1^3 P_L(k_1) P_L(k_2) + 2 \text{perm.} \\ &\propto (b_1 \sigma_8)^3 \sigma_8. \end{aligned} \quad (3.54)$$

The first information gain comes from the fact that the linear bias  $b_1$  is unique for both  $\delta^{(1)}$  and  $\delta^{(2)}$ . Moreover, with addition of  $B_{\delta^2}$  and  $B_{K^2}$ , we have terms with different scale dependencies which are proportional to  $b_1$ , allowing for further degeneracy breaking.



# Chapter 4

## EFTofLSS meets simulation-based inference: $\sigma_8$ from biased tracers

*This chapter is an adaptation of [217] (published in JCAP). Authors: Beatriz Tucci & Fabian Schmidt.*

### 4.1 Introduction

In this paper, we take advantage of our understanding of galaxy clustering on large scales to test the SBI performance based on `LEFTfield`. Using this tool as a first step is of crucial importance, since we tackle model misspecification by keeping all modes under control, cutting small scales where we should not trust our simulator and using the bias expansion on the scales where it is known to be robust. The simulator uses Lagrangian Perturbation Theory (LPT), which is fast to evaluate and therefore allows a detailed study of convergence with respect to the number of simulations and also careful posterior diagnostics.

Using the bias expansion constructed in `LEFTfield`, we employ the combination of the galaxy power spectrum and bispectrum to break the degeneracy between the bias parameters and the amplitude of fluctuations  $\sigma_8$ . Our main goal is to assess the potential impact on cosmological inference when we consider the entire data vector distribution through SBI rather than relying on the assumption of a Gaussian likelihood. We also provide a comprehensive description of how simulation-based inference works in practice, as this is one of the first applications of this technique to galaxy clustering. In the context of SBI, this is the first time that the complete second-order bias expansion is used. Our work is thus complementary to recent work [90, 91, 149], which uses halo occupation distribution (HOD) as a bias model, and includes the power spectrum on smaller scales. Note that our forward model is numerically much less costly than the one used in [90, 91, 149], allowing us to generate many more independent simulated data realizations.

## 4.2 Methods

### 4.2.1 LEFTfield and SBI

The forward model we use to get from the initial conditions, cosmological, noise and bias parameters to the final data is `LEFTfield`, which evolves a given initial (linear) density field up to  $n$ -th order in LPT [190]. This represents a powerful tool for cosmological inference using SBI. As already discussed, statistical inference in SBI is dependent on the quality of the simulator and the simulation budget. Using the EFT-based approach, we keep all the modes under control up to a certain scale  $\Lambda$ , such that SBI only learns the features we trust from our simulator. We therefore combine two big advantages of EFTofLSS for SBI: (i) theoretically robust treatment of bias; and (ii) fast evaluation thanks to the fact that we only need to follow the evolution of modes up to  $\Lambda$ , making it a perfectly suitable tool for analysing convergence and coverage of SBI algorithms. For example, using a Subhalo Abundance Matching (SHAM) [219] approach would be less theoretically robust, while being much more costly in comparison. The price to pay of course is that our forward model is limited to scales where perturbation theory is valid.

Our model differs from those commonly used in the EFTofLSS literature where the basis for the  $n$ -point functions are generated by methods such as CLASS-PT [53] and PyBird [61], since we directly measure the  $n$ -point functions on the simulated galaxy density field. By doing so, we are also simulating the full (non-Gaussian) distribution of our data vector, as generated by the LPT-based simulator, and SBI can provide us with the correct distribution for the summary statistics considered. Each simulation realization is run with a different seed for the density and stochastic fields, so that our model correctly accounts for cosmic variance. Note that SBI models the *probability density* of the data vector, and is therefore also in contrast to emulators [20], where the *expectation value* of the summary statistics is learned.

**LEFTfield settings.** In this paper, we use second-order LPT (2LPT) and CIC for the density assignment throughout. Unless otherwise stated, we employ second-order Lagrangian bias with operators

$$\mathcal{O}^L = \left\{ \delta, \nabla^2 \delta, \left( \text{tr} [\mathbf{M}^{(1)}] \right)^2, \text{tr} [\mathbf{M}^{(1)} \mathbf{M}^{(1)}] \right\}, \quad (4.1)$$

and stochastic parameters (see Sec. 2.5 and Appendix A.1)

$$\mathcal{O}^{\text{stoch.}} = \left\{ \varepsilon, \varepsilon^2, \varepsilon \delta \right\}, \quad (4.2)$$

neglecting higher-derivative stochastic corrections. We set  $\sigma_8^{\text{fid}} = 0.85$ , while all other cosmological parameters are fixed to  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ ,  $h = 0.7$  and  $n_s = 0.967$ .

### 4.2.2 Summary statistics

The usual procedure to extract information from the galaxy density field is through its  $n$ -point functions. Since its one-point function (the mean density) is trivial, most of the information can be captured by its two- and three-point functions or, equivalently, their Fourier transforms, namely the power spectrum and bispectrum. Under the homogeneity and isotropy conditions, the bispectrum is described by three parameters which characterize the shape and the scale of the triangles, thus encoding important cosmological information beyond the two-point function which only depends on scale [196]. Especially in the context of primordial non-Gaussianities, it is convenient to separate the triangle shapes into e.g. “squeezed” and “equilateral” configurations, but we do not make such distinctions as we work with all triangle configurations.

The tree-level galaxy power spectrum displays a degeneracy between the amplitude of density fluctuations, parametrized by  $\alpha$  in our forward model, and the linear bias  $b_1$ , while the bispectrum has the power of breaking this degeneracy [77, 78, 142, 196]. The intuition is that, at leading order in perturbations, the galaxy bispectrum will depend on a sum of different powers of the bias parameters and the linear power spectrum, allowing for a determination of the linear and second-order bias parameters which is independent of the power spectrum normalization. We refer the reader to [107] and [66] (Sec. 4.1.1) for an illustration of the different shape contributions of the matter and galaxy bispectrum, respectively.

To measure such  $n$ -point functions on the grids generated by `LEFTfield`, we use the bispectrum estimator introduced by [195]. We start by defining the quantities [87]

$$I_k(\mathbf{x}) = \int_{\|\mathbf{p}-k\| < \frac{\Delta k}{2}} \frac{d\mathbf{p}}{(2\pi)^3} \delta_g(\mathbf{p}) e^{i\mathbf{x}\cdot\mathbf{p}}, \quad J_k(\mathbf{x}) = \int_{\|\mathbf{p}-k\| < \frac{\Delta k}{2}} \frac{d\mathbf{p}}{(2\pi)^3} e^{i\mathbf{x}\cdot\mathbf{p}}, \quad (4.3)$$

where all  $\mathbf{p}$  modes within a  $k$ -shell of width  $\Delta k$  centered on  $k$  are integrated. The expressions for  $I_k(\mathbf{x})$  and  $J_k(\mathbf{x})$  therefore represent the inverse Fourier transform over a  $k$ -shell of the galaxy overdensity field, which is calculated following the procedure of Sec. 2.5, and a unit field, respectively. The power spectrum estimator is defined as

$$\hat{P}(k_1) = \left( \frac{L^3}{N^6} \right) \frac{\sum_{i=1}^{N^3} I_{k_1}^D(x_i) I_{k_1}^D(x_i)}{\sum_{j=1}^{N^3} J_{k_1}^D(x_j) J_{k_1}^D(x_j)}, \quad (4.4)$$

where the superscript  $D$  refers to the discretized version of the fields,  $k_1$  is the Fourier bin,  $N$  is the grid size and  $L$  is the box size, while the bispectrum estimator in turn is

$$\hat{B}(k_1, k_2, k_3) = \left( \frac{L^6}{N^9} \right) \frac{\sum_{i=1}^{N^3} I_{k_1}^D(x_i) I_{k_2}^D(x_i) I_{k_3}^D(x_i)}{\sum_{j=1}^{N^3} J_{k_1}^D(x_j) J_{k_2}^D(x_j) J_{k_3}^D(x_j)}. \quad (4.5)$$

It is now evident that the power spectrum comes essentially for free when estimating the bispectrum. As already discussed in the introduction, we do not suffer from binning effects, since bin averaging is automatically taken into account in the SBI framework. This

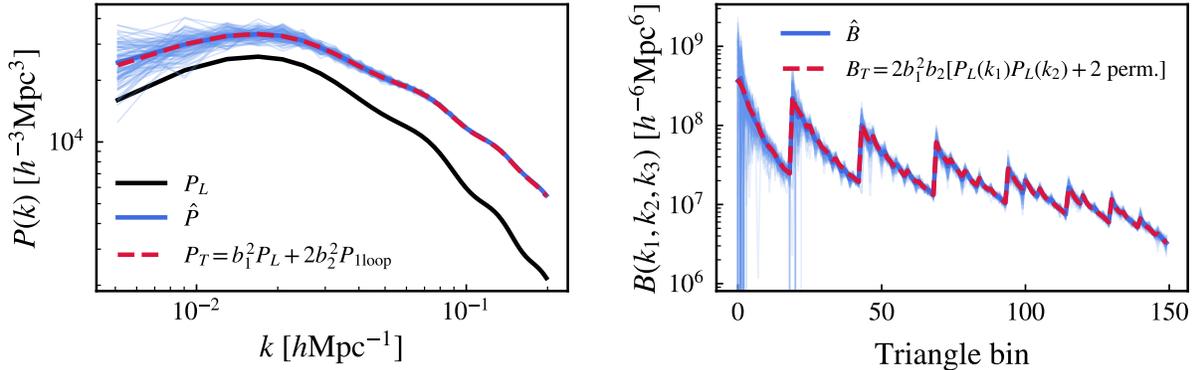


Figure 4.1: Comparison of power spectrum and bispectrum estimators to theory for the linear forward model in Eq. (4.6), where we use  $\Lambda = 0.3h\text{Mpc}^{-1}$ ,  $k_{\min} = 2\pi L^{-1}h\text{Mpc}^{-1}$  and  $k_{\max} = 0.2h\text{Mpc}^{-1}$  for both cases. The blue lines show the  $n$ -point function estimator from LEFTfield as described in this section, which agree with the theory predictions indicated by the red curves. *Left*: we set  $b_1 = b_2 = 1$  and use 50 linear bins for the power spectrum. The linear power spectrum  $P_L$  is displayed in black for reference. *Right*: we set  $b_1 = 1$  and  $b_2 = 0.1$ , in order to suppress next-to-leading-order corrections, and use 150 triangle bins for a better visualization of the bispectrum, which are constructed from 10 linear bins in  $k$ . The lowest triangle bin indexes contain the smaller  $k$ -modes, hence their large variance, where  $k_1 \leq k_2 \leq k_3$ , where  $k_1$  corresponds to the outermost loop index, while  $k_3$  is the innermost index.

allows us to use both “open” and “closed” triangle configurations, as opposed to likelihood-based approaches where the binning effect is important for the open triangles [103, 159]. “Open” triangles are those for which the  $k$ -bins do not respect the momentum conservation constraint  $|k_3 - k_2| < k_1 < k_2 + k_3$ , although there are individual triples of modes associated to the bin which do satisfy the relation.

Considering the most costly model in this work, i.e., the Euclid configuration described in Section 4.3.2, using LEFTfield to generate  $N_{\text{sim}} = 10^5$  density grids with 2LPT, construct a second-order bias expansion and compute their corresponding power spectra and bispectra with data vector size of  $D = 49$  (6 bins for the power spectrum and 43 triangle bins for the bispectrum) takes roughly 19500 core hours when using 2GHz Intel Xeon Gold 6138 CPUs.

We show in Figure 4.1 the comparison of our estimators to their respective theoretical predictions. We use here a linear forward model instead of 2LPT, which is given directly from Eq. (2.107) for  $\alpha = 1$ , where we construct the final density field in a box with size  $L = 2000h\text{Mpc}^{-1}$  as

$$\delta_g(\mathbf{k}, z) = b_1 \delta^{(1)}(\mathbf{k}, z) + b_2 \left( [\delta^{(1)}]^2(\mathbf{k}, z) - \langle [\delta^{(1)}]^2 \rangle(z) \right), \quad (4.6)$$

choosing the redshift  $z = 0$ . The second-order bias term generates a nonzero bispectrum. The prediction for the power spectrum estimator is then  $P_T(k) = b_1^2 P_L(k) + 2b_2^2 P_{1\text{loop}}(k)$ ,

where

$$P_{1\text{loop}}(k) = \int_{||\mathbf{p}|-k| < \frac{\Delta k}{2}} \frac{d\mathbf{p}}{(2\pi)^3} P_L(|\mathbf{p}|) P_L(|\mathbf{k} - \mathbf{p}|) e^{i\mathbf{k}\cdot\mathbf{p}} \quad (4.7)$$

and  $P_L$  has support up to  $\Lambda$ . For the bispectrum, its leading-order theory prediction in this case reads  $B_T(k_1, k_2, k_3) = 2b_1^2 b_2 [P_L(k_1)P_L(k_2) + 2 \text{ perm.}]$ . We have also tested the LPT predictions for the power spectrum and bispectrum estimator (not shown), where in the latter case we compare the closed triangle bins only.

### Simulation-based inference

We have tested some of the NDE algorithms implemented in the publicly available SBI package [211], and we chose to work with NPE from SBI as our baseline. There are several reasons for our choice; first, this method has been empirically observed to perform better for the inference problems considered in this work, especially for a limited simulation budget and large dimensions of the data parameter vectors. Second, it does not require an extra MCMC step as in NLE and NRE, or a retraining of the model for each observation such as for non-amortized algorithms, allowing for faster posterior diagnostics. Lastly, amortized posteriors such as the ones obtained by NPE tend to be more conservative [96].

We use the SNPE method of [85] with 10 atoms for atomic proposals and MAF with 5 autoregressive layers (i.e., stacked MADEs), each constructed using two fully-connected tanh layers with 50 hidden units. We train the models by stochastically minimizing the loss using the Adam optimizer [112] with learning rate of  $5 \times 10^{-4}$  and batch size of 50, where 10% of the samples are used for validation and the training is stopped if the validation set loss did not improve after 20 consecutive epochs.

Training the most complex case considered here (Section 4.3.2, with data vector size of  $D = 49$ ,  $N_\theta = 8$  parameters and simulation budget of  $N_{\text{sim}} = 10^5$ ) takes roughly one hour on a single 2GHz Intel Xeon Gold 6138 CPU, and no significant improvement was observed when using a GPU. Generating  $10^5$  samples from the estimated posterior takes less than one second, allowing for efficient calibration analysis since the NPE posterior is amortized.

## 4.3 Results

In this section, we present our results regarding two cases: first, a simple, linear forward model, described in Sec. 4.3.1, where we compare ABC and NPE. In Sec. 4.3.2, we then use a 2LPT forward model and estimate the posterior with NPE, as summarized in Figure 4.2.

### 4.3.1 Linear forward model

We start reporting our results with a very simple model, where we aim to infer the linear bias  $b_1$  and the noise amplitude  $P_\varepsilon$  from the galaxy power spectrum. For that, we consider

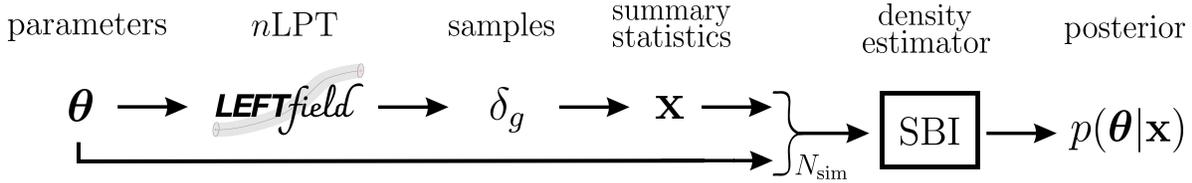


Figure 4.2: Flowchart of our pipeline for SBI, assuming an  $n$ LPT forward model. The cosmological, bias and noise parameters  $\theta$  are sampled from the prior and used as input for LEFTfield (Sec. 2.5), which constructs a model for the galaxy overdensity field  $\delta_g$  through  $n$ LPT (2LPT in this work). We then measure the summary statistics  $\mathbf{x}$ , which in our case is the galaxy power-spectrum concatenated with the bispectrum (Sec. 4.2.2). We repeat this process  $N_{\text{sim}}$  times, and use the resulting pairs  $\{(\theta_n, \mathbf{x}_n)\}_{n=1}^{N_{\text{sim}}}$  as input to the SBI neural density estimation (Sec. 3.4.2), which then gives us the posterior  $p(\theta|\mathbf{x})$  that can in turn be evaluated at a given observed data  $\mathbf{x}_o$ .

a *linear forward model*, where instead of using LPT we use Eq. (2.107) with  $\alpha = 1$  to construct the galaxy overdensity field as

$$\delta_g(\mathbf{k}, z) = b_1 \delta^{(1)}(\mathbf{k}, z) + \varepsilon(\mathbf{k}), \quad (4.8)$$

where  $\varepsilon(\mathbf{k})$  is a Gaussian random field of zero mean and variance  $P_\varepsilon$ . We use a box of size  $L = 2000 h^{-1} \text{Mpc}$ , and choose redshift  $z = 0$ , without any loss of generality in this case.

We define the observed data  $\mathbf{x}_o$  as a fiducial power spectrum  $\bar{P}(k)$  measured from a particular realization of the model evaluated at the ground-truth parameters  $\bar{b}_1 = 1.5$  and  $\bar{P}_\varepsilon = 10^3 h^{-3} \text{Mpc}^3$ . We then generate  $N_{\text{sim}}$  simulated datasets, where  $\theta = \{b_1^2, P_\varepsilon\}$  are sampled from a Gaussian prior as  $\theta \sim \mathcal{N}(\bar{\theta}, 10^{-1} \times \bar{\theta})$  and  $\mathbf{x} = \{P(k)\} = \{P(k_1), P(k_2), \dots, P(k_D)\}$ , where  $D$  is the number of power-spectrum bins  $N_{\text{bin}}$ , are the power spectra of the simulated datasets. It is important to stress that, for each simulation, the realization of both  $\delta^{(1)}(\mathbf{k})$  and  $\varepsilon(\mathbf{k})$  are also varied. We infer  $b_1^2$  instead of  $b_1$  since its posterior can be evaluated analytically assuming a Gaussian likelihood and prior. Note that the SBI posterior for  $b_1^2$  does not necessarily converge to the analytical one, since we relax the Gaussian likelihood assumption in this case.

The ABC metric we consider is

$$\rho(\mathbf{x}, \mathbf{x}_o) = \frac{1}{(D - N_\theta - 1)} \sum_{i=1}^D \frac{(\mathbf{x}^i - \mathbf{x}_o^i)^2}{\text{Cov}[\bar{\mathbf{x}}^i]}, \quad (4.9)$$

where the sum runs over the summary statistics bins ( $k$ -bins for the power spectrum). This metric is inspired by the reduced  $\chi^2$ , and is expected to be close to optimal in the present case, in addition to being straightforward to interpret. Here,  $D = N_{\text{bin}} = 4$ ,  $N_\theta = 2$  and  $\text{Cov}[\bar{P}(k)] = 2\langle \bar{P}(k) \rangle^2 / N_k$ , where  $\langle \bar{P}(k) \rangle$  is the mean of the fiducial power spectrum over initial conditions evaluated at fiducial parameters  $\bar{\theta}$  and  $N_k$  is the total number of modes which lie in the respective  $k$ -bin. As explained in Sec. 3.4.1, the proposed parameters  $\theta$  are

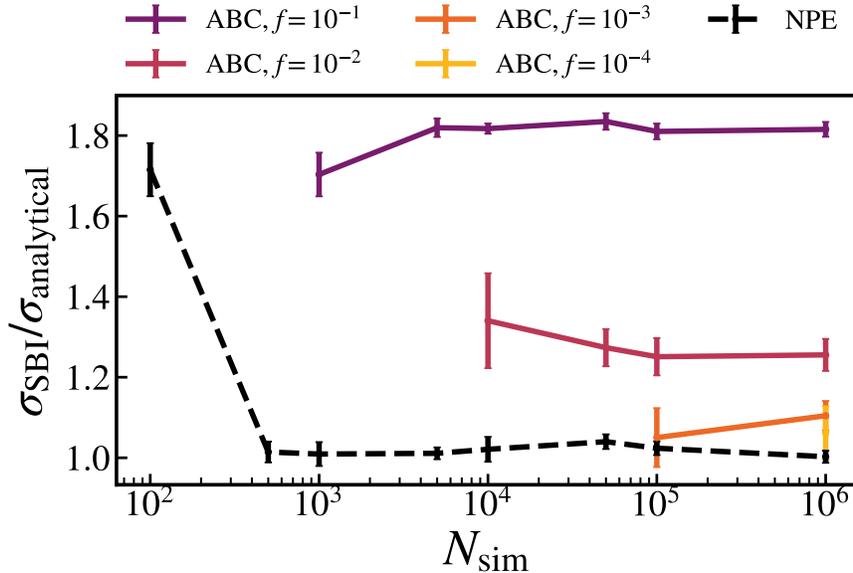


Figure 4.3: Posterior standard deviation of the linear bias parameter squared  $b_1^2$  obtained by SBI divided by its analytical prediction as a function of the total simulation budget  $N_{\text{sim}}$ . Continuous lines show the results obtained from ABC, while the dashed line corresponds to NPE. The value of  $f$  denotes the fractional percentage of samples which were accepted in ABC, where we use  $f \times N_{\text{sim}}$  for posterior estimation. The error bars correspond to 10 different realizations of the observational mock data  $\mathbf{x}_o$ .

then accepted or rejected according to this metric evaluated at the simulated data vector  $\mathbf{x}$ , and the ABC posterior should converge to the true one in the  $\epsilon \rightarrow 0$  limit. However, due to computational efforts, one can never truly reach this limit.

We take the approach of [137], where instead of defining a specific threshold  $\epsilon$  we accept the parameters corresponding to the lowest quantile  $f$  of metric values. We show in Figure 4.3 the SBI posterior standard deviation for  $b_1^2$  normalized by the analytical prediction, where we compare ABC to NPE. We use a minimum of  $10^2$  samples for estimating the posterior variance, leading to a different number of points for different choices of  $f$ . The corresponding threshold value  $\epsilon$  for each percentage fraction  $f = \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$  is  $\epsilon = \{16.2, 3.6, 1.1, 0.4\}$ .

First, we can see that the NPE posterior converges much faster in terms of  $N_{\text{sim}}$  than the ABC one, as expected since the effective number of samples used in ABC is  $f \times N_{\text{sim}}$ , although the total simulation cost is  $N_{\text{sim}}$ , while NPE uses all simulations at hand for posterior estimation. Further, we find that one has to be careful when choosing the value for the threshold  $\epsilon$ ; values that are too large will lead to significantly overestimated error bars, regardless of the number of simulations used. If the summary statistics used is approximately Gaussian-distributed with approximately known covariance, such as in our

case, defining the ABC metric via the reduced  $\chi^2$  makes the former interpretable, with convergence expected for values of  $\epsilon$  of order unity, which is indeed confirmed by Figure 4.3 given the values of  $\epsilon$  listed above. In the simple case we considered here, we can see that the posterior standard deviation converges to the Gaussian prediction, and we can conclude that for this particular case the non-Gaussianity of the power spectrum at low- $k$  does not significantly impact the constraints.

Figure 4.3 serves as a useful cross-check of the different SBI approaches. In the following, due to the difficulty in achieving converged posteriors via ABC, we will use NPE throughout.

### 4.3.2 2LPT forward model

We now move to a forward model consisting of second-order LPT and second-order bias expansion as described in Section 2.5, where the galaxy overdensity field is given by Eq. (2.115). Here,  $\mathbf{x}$  is the concatenated vector of power spectrum and bispectrum bins of dimension  $D = N_{\text{bin}} + N_{\text{tr}}$ , where  $N_{\text{tr}}$  denotes the number of triangle bins for the bispectrum. We use the configuration of the Euclid satellite [18], which will cover a volume of  $63 h^{-3} \text{Gpc}^3$  with mean number density of galaxies  $\bar{n}_g = 5.2 \times 10^{-4} h^{-3} \text{Mpc}^3$  at mean redshift  $z = 1.4$ .

We consider two cutoffs for generating the datasets. For  $\Lambda = 0.1 h\text{Mpc}^{-1}$ , the grid size is  $N_c = 128^3$  and the dimension of the data vector is  $D = 33$ , while for  $\Lambda = 0.2 h\text{Mpc}^{-1}$ ,  $N_c = 256^3$  and  $D = 49$ . The fiducial values of the bias and noise parameters are set to

$$\begin{aligned} \bar{\boldsymbol{\theta}} &= \{ \bar{\alpha}, \bar{b}_1, \bar{b}_{\nabla^2\delta}, \bar{b}_{\text{tr}[M^{(1)}M^{(1)}]}, \bar{b}_{\sigma\sigma}, 10^{-3}\bar{P}_\epsilon, \bar{c}_{\epsilon\delta}, \bar{c}_{\epsilon^2} \} \\ &= \{ 1, 1.5, -1.84, -0.26, -0.79, 1.92, 0.75, 0.17 \}, \end{aligned} \quad (4.10)$$

where the fiducial noise parameters are set to their corresponding Poisson expectation (see Appendix A), inspired by galaxy number densities of current and upcoming spectroscopic surveys, and we choose to sample  $10^{-3}P_\epsilon$  instead of  $P_\epsilon$  in order to have all parameters to be of order unity, which is desirable for the neural density estimation. The fiducial Eulerian linear bias is set to  $\bar{b}_1 = 1.5$ , while  $\bar{b}_{\text{tr}[M^{(1)}M^{(1)}]}$  and  $\bar{b}_{\sigma\sigma}$  are determined from  $\bar{b}_1$  through the bias relations (see Appendix B).

Regarding  $\bar{b}_{\nabla^2\delta}$ , we expect its value to be close to zero, but with variance between two and three times the Lagrangian radius of an average Euclid galaxy. To set its fiducial value, we first find the minimum mass  $M_{\text{min}}$  which satisfies the integral

$$\langle b_{\mathcal{O}}(z, M_{\text{min}}) \rangle = \int_{M_{\text{min}}}^{\infty} dM \frac{dn(z, M)}{dM} b_{\mathcal{O}}(z, M) \quad (4.11)$$

for  $\langle b_1(z, M_{\text{min}}) \rangle = \bar{b}_1 = 1.5$  using the Tinker mass function for  $n(z, M)$  [213] and the Tinker linear bias predictions for  $b_1(z, M)$  [214]. We then use Eq. (4.11) for the higher-derivative bias evaluated at the obtained  $M_{\text{min}}$ , where  $b_{\nabla^2\delta}(z, M)$  is determined by  $-R_L^2(M)/2.5$  [122], to set its fiducial value as the calculated mean, i.e.,  $\bar{b}_{\nabla^2\delta} = \langle b_{\nabla^2\delta}(z, M_{\text{min}}) \rangle$ .

We compare our results to a Fisher forecast, which approximates the posterior with a Gaussian via the ensemble average of its curvature of around the maximum. The Fisher

information matrix for the likelihood is defined as

$$F_{\alpha\beta}^{(\mathcal{L})} \equiv - \left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle \Big|_{\bar{\boldsymbol{\theta}}}, \quad (4.12)$$

which for a Gaussian likelihood reduces to

$$F_{\alpha\beta}^{(\mathcal{L})} = \sum_{i,j=1}^D \left[ \frac{\partial x_i}{\partial \theta_\alpha} (\text{Cov}[\mathbf{x}])_{ij}^{-1} \frac{\partial x_j}{\partial \theta_\beta} \right] \Big|_{\bar{\boldsymbol{\theta}}}, \quad (4.13)$$

where the mean of the derivatives as well as the covariance of the data vector  $\mathbf{x}$  over initial conditions realizations are evaluated at the fiducial parameters values  $\bar{\boldsymbol{\theta}}$ . We also include the prior Fisher matrix  $F_{\alpha\beta}^{(p)} = \delta_{\alpha\beta}/\sigma_\alpha^2$ , where  $\delta_{\alpha\beta}$  is the Kronecker delta function and  $\sigma_\alpha$  is the standard deviation considered for the parameter  $\theta_\alpha$  in a Gaussian prior. The total Fisher information matrix for the posterior is then  $F = F^{(\mathcal{L})} + F^{(p)}$ . The marginal posterior in any subspace of the parameter space is then controlled by the corresponding restriction of  $F^{-1}$ . Specifically, the 1-sigma uncertainty on  $\theta_\alpha$  is approximated by  $\sqrt{(F^{-1})_{\alpha\alpha}}$ . As previously discussed, the power spectrum and bispectrum are not exactly Gaussian distributed, and one of the main goals of this paper is to explore the final parameter posterior when this assumption is relaxed. However, the Fisher prediction provides a guideline for interpreting our results.

Since we aim to use NPE instead of sequential methods for amortizability and since SBI requires many simulations around high density posterior regions, a careful choice of priors (i.e. neither too wide nor too narrow) is crucial. We hence choose the mean and variance of our priors in the following way. First, we run SNPE with batches of  $10^3$  simulations starting from the prior

$$\begin{aligned} \alpha &\sim \mathcal{N}(1.2, 1^2), & b_1 &\sim \mathcal{N}(1.25, 2^2), & b_{\sigma\sigma} &\sim \mathcal{N}(0, 2^2), & b_{\text{tr}[M^{(1)}M^{(1)}]} &\sim \mathcal{N}(0, 2^2), \\ b_{\nabla^2\delta} &\sim \mathcal{N}(0, 25^2), & 10^{-3} P_\varepsilon &\sim \mathcal{N}(2, 1^2), & c_{\varepsilon\delta} &\sim \mathcal{N}(0, 2^2), & c_{\varepsilon^2} &\sim \mathcal{N}(0, 2^2), \end{aligned} \quad (4.14)$$

where we choose the centers of  $\alpha$  and  $b_1$  in such a way that their multiplication equals  $1.5 = \bar{\alpha} \times \bar{b}_1$ . After SNPE convergence, we sample from the final posterior to choose the prior for NPE to be a Gaussian centered on the sample mean with variance given by at least two times the SNPE sample variance. For all inference cases, we overplot the prior with the posterior to guarantee that the latter is not bounded by the prior ranges. We refer the reader to Sec. 4.4 for examples comparing SNPE and NPE.

### Fixed cosmology: impact of the likelihood form

We start with the case where  $\alpha$  is fixed to unity, while all the first and second order bias and noise parameters are sampled. We name as *full SBI* results from the forward model where the measured data vector from `LEFTfield` are used directly for posterior estimation. We compare these results to a *Gaussian-likelihood* model, where we generate samples from the data vector as  $\mathbf{x} \sim \mathcal{N}(\langle \mathbf{x} \rangle, \text{Cov}[\bar{\mathbf{x}}])$ , where  $\langle \mathbf{x} \rangle$  denotes the mean of the data vector over

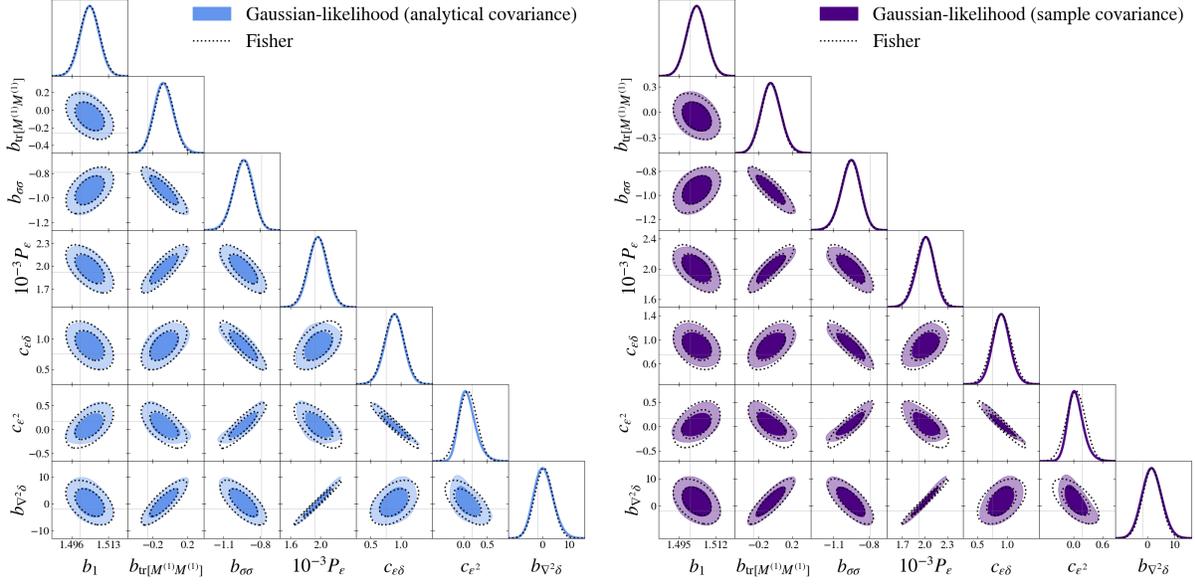


Figure 4.4: Parameter posterior for the case where the cosmological parameter  $\alpha$  is fixed. The left and right contour plots correspond to the models where the data vector is sampled from a Gaussian with analytical and sample covariance, respectively. For all cases, the neural density estimator method NPE was used with a simulation budget of  $N_{\text{sim}} = 10^5$ .

initial conditions realizations calculated at the proposed parameter  $\theta$  and the covariance is calculated for the data vector at the fiducial parameters  $\bar{\mathbf{x}}$ . While the name Gaussian likelihood refers to the fact that the data is Gaussian distributed by construction, we emphasize that the posterior is also obtained with NPE as in the full-SBI case. We of course expect to obtain the same result when performing a Monte Carlo sampling based on the same Gaussian likelihood for the data vector.

Regarding the covariance  $\text{Cov}[\bar{\mathbf{x}}]$ , we consider two distinct cases. First, we consider an *analytical*, diagonal covariance for the data vector, namely

$$\text{Cov}[\bar{P}(k)] = 2\langle\bar{P}(k)\rangle^2/N_k, \quad (4.15)$$

where  $\langle\bar{P}(k)\rangle$  is the mean of the fiducial power spectrum over initial conditions evaluated at fiducial parameters  $\bar{\theta}$  and  $N_k$  is the total number of modes which lie in the respective  $k$ -bin, and

$$\text{Cov}[\bar{B}(k_1, k_2, k_3)] = s_B \langle\bar{P}(k_1)\rangle\langle\bar{P}(k_2)\rangle\langle\bar{P}(k_3)\rangle/k_f^3 N_t, \quad (4.16)$$

where  $s_B$  is the triangle shape symmetry factor (6, 2 and 1 for equilateral, isosceles and scalene triangles, respectively),  $k_f = 2\pi L^{-1}$  is the fundamental frequency and  $N_t$  is the number of triangle configurations inside each triangle bin. Second, we measure the *sample* covariance from  $N_{\text{cov}}$  simulations with parameters fixed at their fiducial values.

To summarize, in both *Gaussian-likelihood* cases, the likelihood of the data vector is characterized completely by its first and second moments (namely the mean and covariance), where the covariance is fixed to the fiducial point in parameter space. Compared

to the analytical covariance, the sample covariance case adds more realism, since cross-correlations between different elements of the data vector (i.e. off-diagonal elements of the covariance) are included. In contrast, the *full SBI* forward model captures the full distribution of the data vector, including its higher-order moments, and as function of the position in parameter space.

We show in Figure 4.4 the estimated posteriors for the Gaussian likelihood model at  $k_{\max} = \Lambda = 0.1 h\text{Mpc}^{-1}$ , which closely resembles the Fisher prediction. For all parameter posteriors considered in this work, we always sample  $10^5$  posterior samples for plotting. The Fisher derivatives are taken numerically for the full case, with steps given by the standard deviation of the NPE posterior, while for the Gaussian-likelihood case it is evaluated analytically at the fiducial values from the mean cross-spectra of the bias expansion basis. The small deviations come from the fact that, although the likelihood and the prior are Gaussian, the posterior is not necessarily Gaussian, as the parameters enter nonlinearly in the prediction for the expectation value of the data vector. Note that, although the Fisher prediction is evaluated at the fiducial parameters points, we shift the mean of the Fisher contours to the maximum a posteriori posterior values for better comparison. This results shows that NPE can successfully recover the expected posterior in this idealized case, showing that this method is therefore robust for the data vector size and number of simulations used.

Figure 4.5 compares the Gaussian-likelihood posteriors with the full-SBI case. First of all, we can notice that the Gaussian likelihood cases are very similar, which indicates that off-diagonal terms do not play an important role on the scales  $k \leq \Lambda = 0.1 h\text{Mpc}^{-1}$  considered here. The difference between these and the full-SBI posteriors is slightly larger, but still small, which also indicates that the non-Gaussianity of the data vector also does not affect the constraints. Figure 4.6 shows the corresponding result when going to smaller scales,  $k \leq \Lambda = 0.2 h\text{Mpc}^{-1}$ . The posterior differences are somewhat larger than on smaller scales, as expected, though still not dramatic.

We show the simulation-based calibration (SBC) tests of the posterior in Figure 4.7 (see Appendix 3.4.4). As previously discussed, a healthy posterior should lead to uniformly distributed rank statistics for all parameters, and we show two different visualizations which can help us in identifying possible problems with the estimated posterior. First, if the SBI underestimates the true posterior variance for some parameter, one expects a “U-shaped” rank distribution, or equivalently a CDF lying below the grey shaded area of the 95% confidence interval of a uniform distribution. Conversely, if the posterior overestimates the errors, one would get a centrally peaked distribution, or a CDF above the grey area. Since our ranks are uniformly distributed and all CDFs lie inside the grey region, we conclude that our posterior passed the calibration test, although we re-emphasize that this is only a necessary and not a sufficient condition.

We show convergence of the obtained posterior with respect to the number of simulations for the full case in Figure 4.8. The results indicate that adding more simulations than  $N_{\text{sim}} = 10^5$ , the simulation budget used in the previous figures, does not change the full-SBI posteriors. The conclusions hold equivalently for the Gaussian likelihood cases. Note that the sudden upturns for a low simulation budget only indicates that the budget

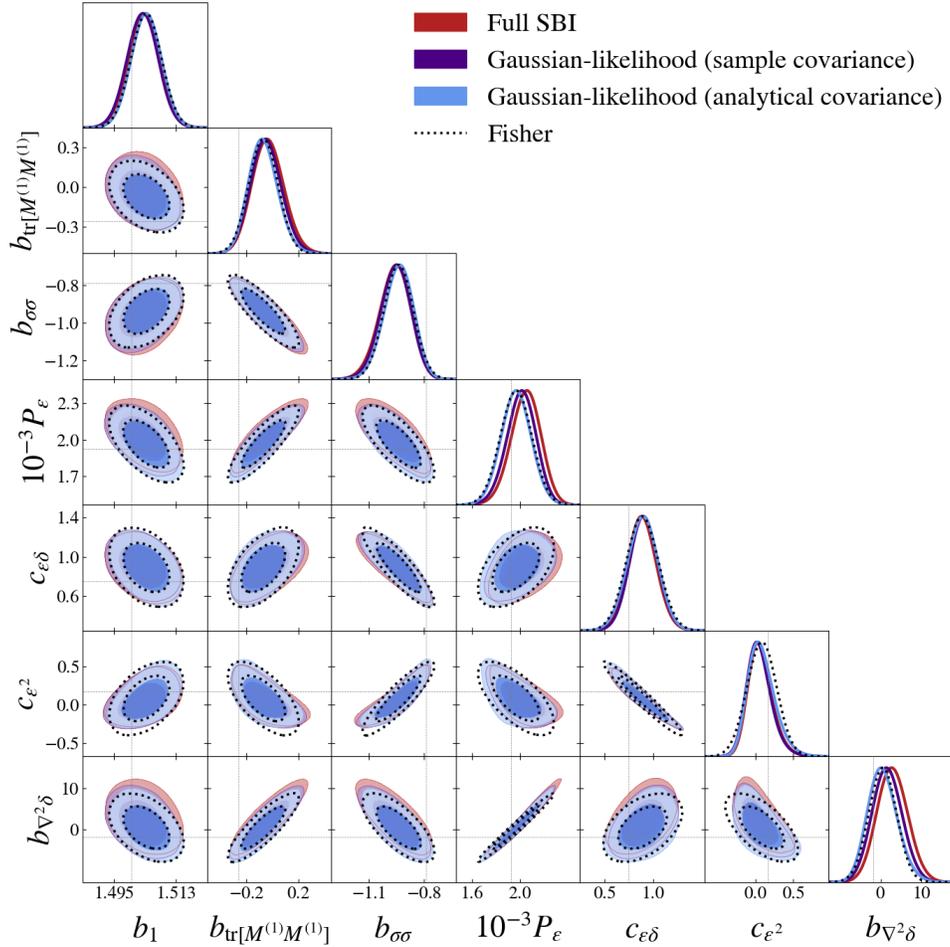


Figure 4.5: Parameter posterior for the case where the cosmological parameter  $\alpha$  is fixed. Blue and purple contours correspond to the models where the data vector is sampled from a Gaussian likelihood with analytical and sample covariance, respectively, where  $N_{\text{cov}} = 10^4$ . The red contours show the full SBI results, with no Gaussian assumption. Dotted lines indicate the Fisher prediction with sample covariance for reference. For all cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 10^5$ , scale cut of  $k_{\text{max}} = \Lambda = 0.1 h\text{Mpc}^{-1}$  and data vector dimension  $D = 33$ .

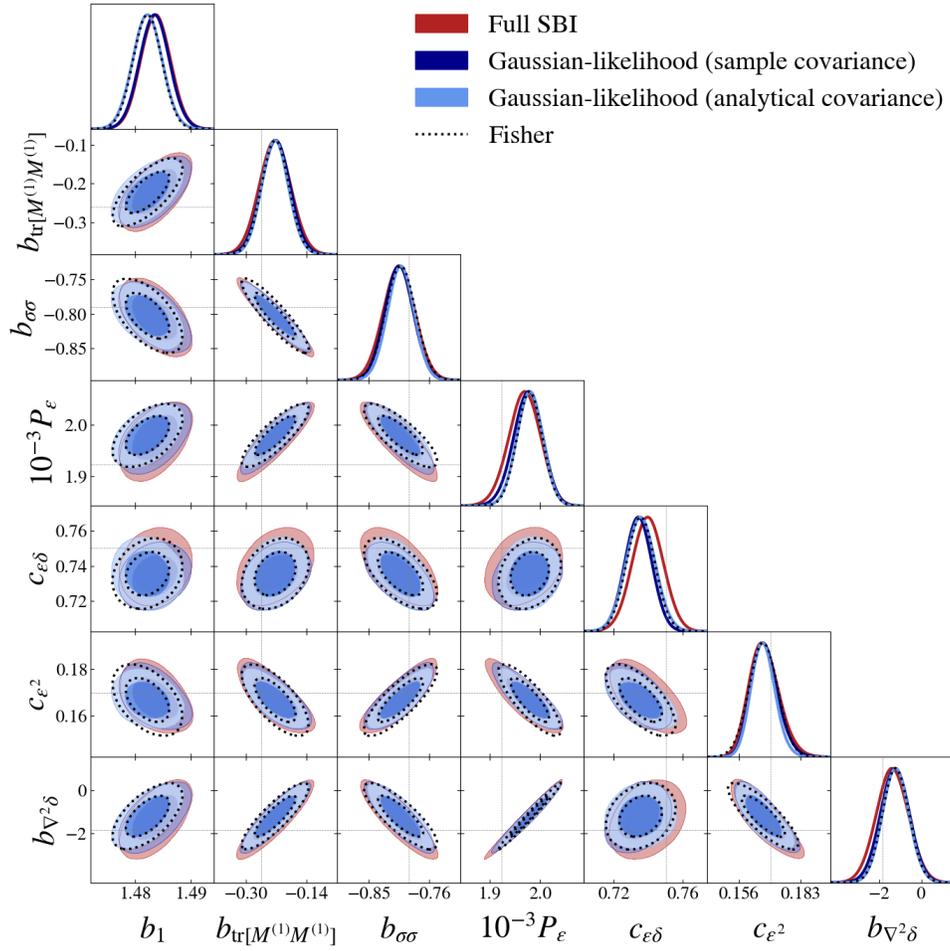


Figure 4.6: Same as Fig. 4.5, but for  $k_{\text{max}} = \Lambda = 0.2h\text{Mpc}^{-1}$  and  $D = 49$ .

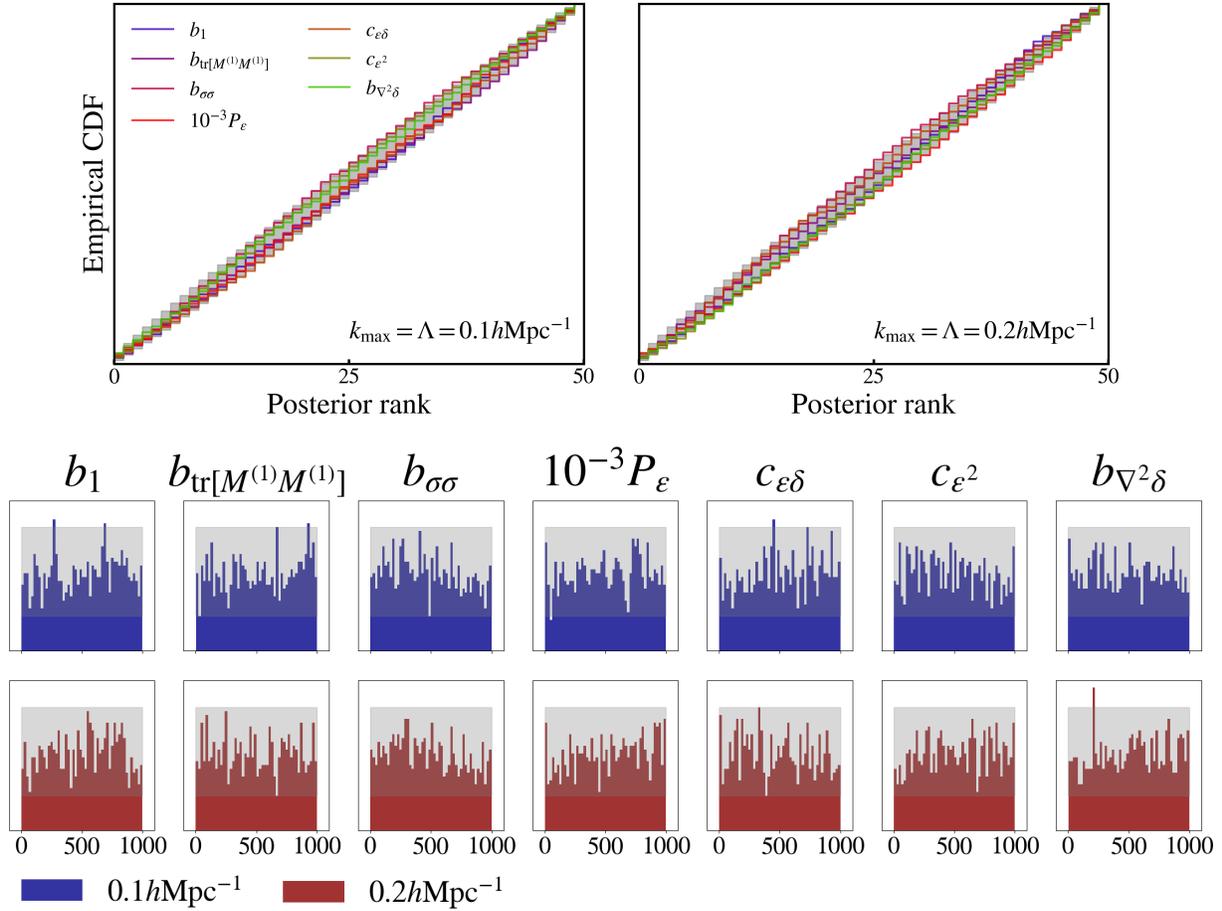


Figure 4.7: SBC (simulation-based calibration) tests for the full-SBI case where the cosmological parameter  $\alpha$  is fixed with  $N_{\text{sim}} = 10^5$ . The upper panels show the CDF of the ranks distributions for each parameter, where the grey area show the 95% confidence interval of a uniform distribution, while the lower panels show the rank distribution, where the grey areas denote the 99% confidence interval of a uniform distribution.

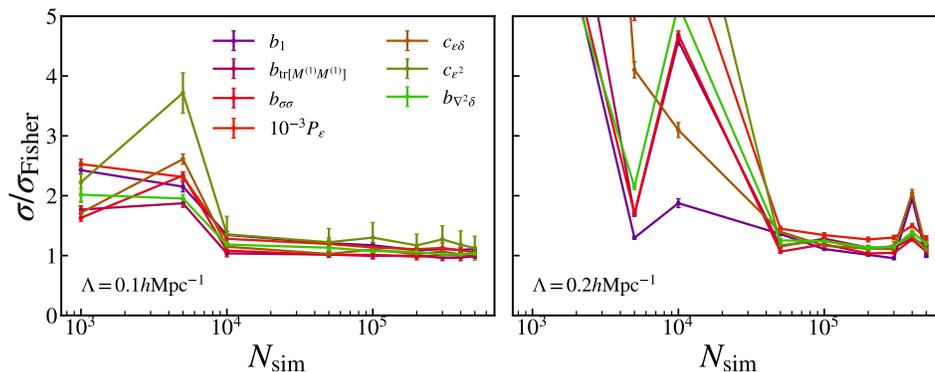


Figure 4.8: Convergence of the full-SBI parameter posterior with the number of simulations  $N_{\text{sim}}$  for the case where the cosmological parameter  $\alpha$  is fixed. The standard deviation of the posterior samples are normalized by their respective Fisher prediction for better comparison. Note that it should not necessarily converge to one. The errors indicate posterior evaluation at 10 different data observations, with same fiducial parameters but different initial conditions realizations.

was not sufficient for convergence, and that running SBI again on the same dataset could lead to different posteriors.

### Inferring cosmology

We now turn to constraining the bias and noise parameters together with the cosmological parameter  $\alpha$ , where we recall that  $\alpha \equiv \sigma_8/\sigma_8^{\text{fid}}$  and set its fiducial value to  $\bar{\alpha} = 1$ . We show in Figure 4.9 how we are able to infer these parameters independently with the help of the bispectrum at  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$ , and that SBI is successful in this context even considering the dimensionality of the problem and the nontrivial parameter degeneracies.

We can notice that the analytical covariance underestimates the errors in this case, a result that has already been discussed in the galaxy clustering literature [41, 54]. We however find no significant difference between the Gaussian likelihood case with sample covariance and the full-SBI one, a result which is in agreement with the current intuition that adding off-diagonal terms to the covariance would be more important than considering a non-Gaussian likelihood. Our conclusion is that, for this particular case, the non-Gaussianity of the data vector (especially at low- $k$  modes) does not lead to any considerable effect on the posterior densities. The conclusion still holds for  $k_{\text{max}} = 0.2h\text{Mpc}^{-1}$ , as displayed in Figure 4.10, although differences between full-SBI and Gaussian-likelihood results become more noticeable. The  $1\text{-}\sigma$  errorbars for the full-SBI, Gaussian-likelihood with sample covariance, Gaussian-likelihood with analytical covariance and the Fisher forecast are 0.173, 0.173, 0.169 and 0.21  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$ , respectively, while 0.131, 0.111, 0.111 and 0.100 for  $k_{\text{max}} = 0.2h\text{Mpc}^{-1}$ .

We emphasize here that this conclusion is highly problem-dependent, and it does not necessarily extend to other cosmological parameters, scale ranges or summary statistics. For example, since most of the constraining power on  $f_{\text{NL}}$  comes from low- $k$  modes, one could expect it to be more sensitive to the Gaussian likelihood assumption. At the other extreme, on highly nonlinear small scales, one can also expect deviations from Gaussianity, while some summary statistics might not even be well described by an analytical form in any scale range.

In general, relaxing the Gaussian assumption in the likelihood is a non-trivial analytical task, but SBI methods provide us a straightforward route. The investigation of whether relaxing the Gaussian likelihood assumption leads to an increase of the error bars has been previously discussed in the literature; for example, [89] reported no significant difference when using the power spectrum only. In contrast, Figure 3 of [125] show that when using the two-point correlation function, the posterior obtained by SBI displays larger error bars than when considering a Gaussian likelihood in the context of lognormal fields.

The SBC results are shown in Figure 4.11, which indicate that our posterior estimates pass the calibration test also for this case. The convergence is analyzed in Figure 4.12. We can notice a slower convergence with respect to the number of simulations when compared to the case where  $\alpha$  is fixed, which might be related to the fact that the two-dimensional contours of  $\alpha$  and the other parameters are more non-Gaussian in this case, due to the increased degeneracies relative to the  $\alpha$ -fixed case, and therefore the normalizing flow needs more simulations to fit the posterior.

We show in Sec. 4.4 an analysis of how our results can be affected by data vector normalization and the number of sample covariance estimates for the Gaussian-likelihood

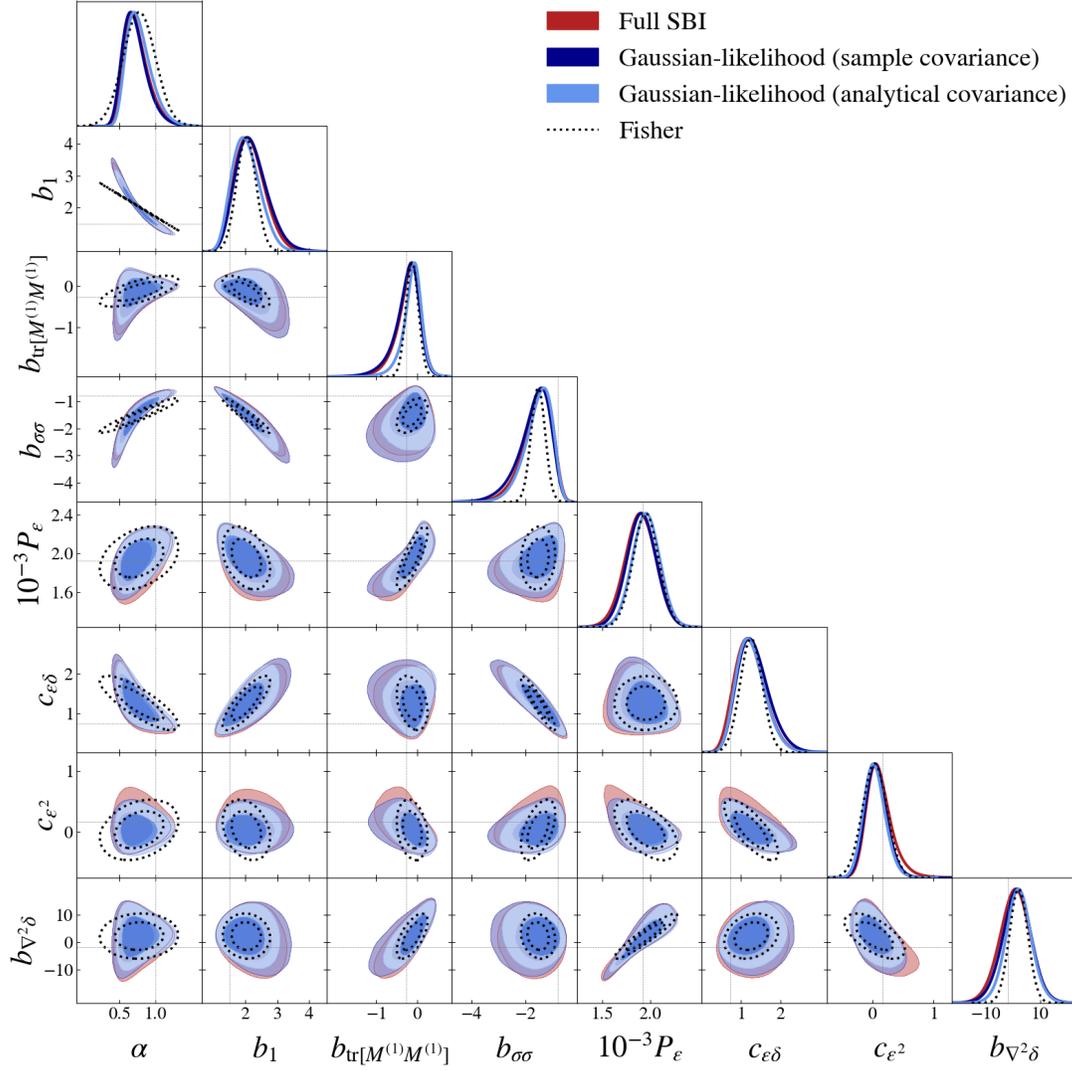


Figure 4.9: Parameter posterior for the case where the cosmological parameter  $\alpha$  is sampled together with the bias and noise parameters. The contour colors, simulation budget and data vector size are the same as in Fig. 4.5. For all cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 10^5$ , scale cut of  $k_{\text{max}} = \Lambda = 0.1h\text{Mpc}^{-1}$  and data vector dimension  $D = 33$ .

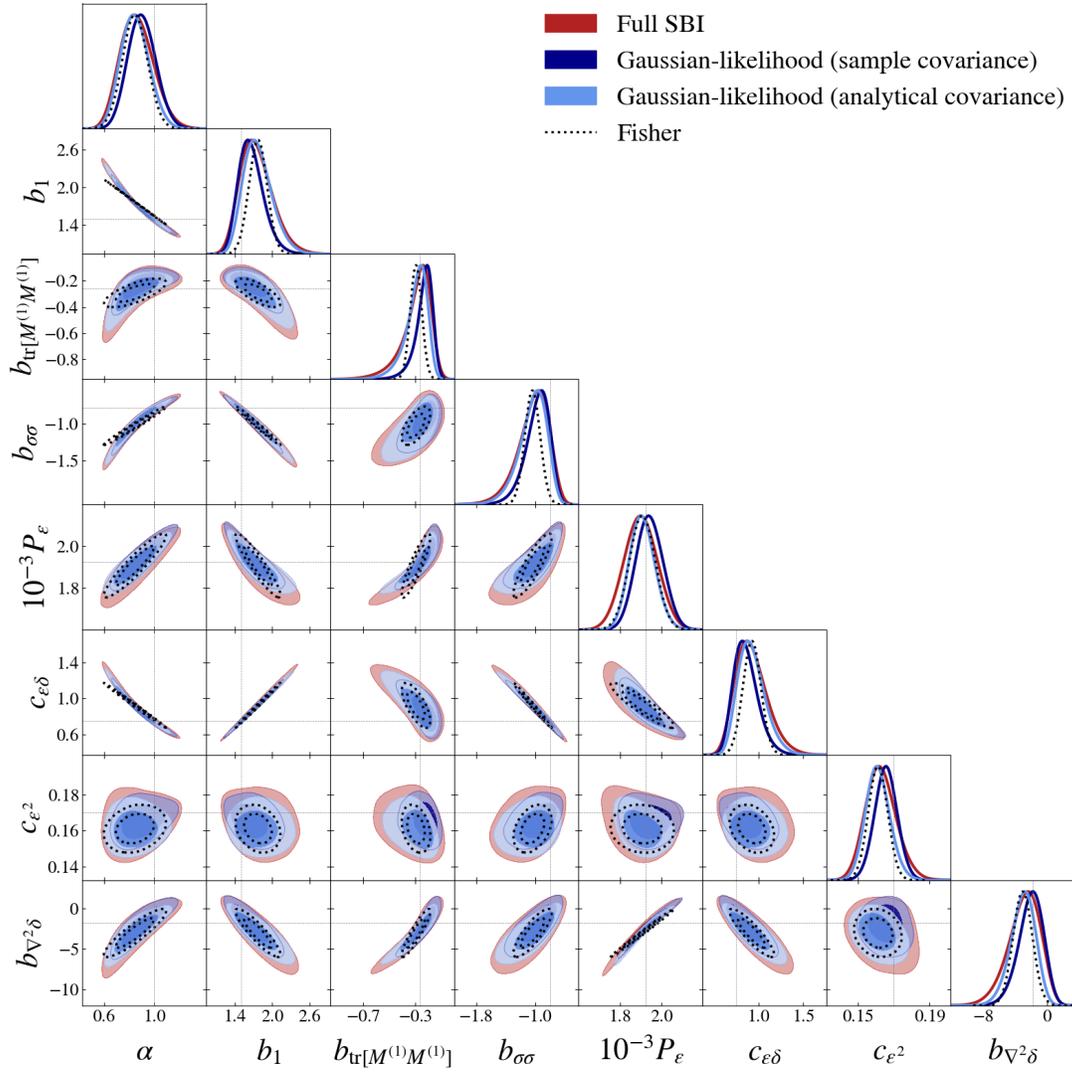


Figure 4.10: Same as Fig. 4.9, but for  $k_{\text{max}} = \Lambda = 0.2h\text{Mpc}^{-1}$  and  $D = 49$ .

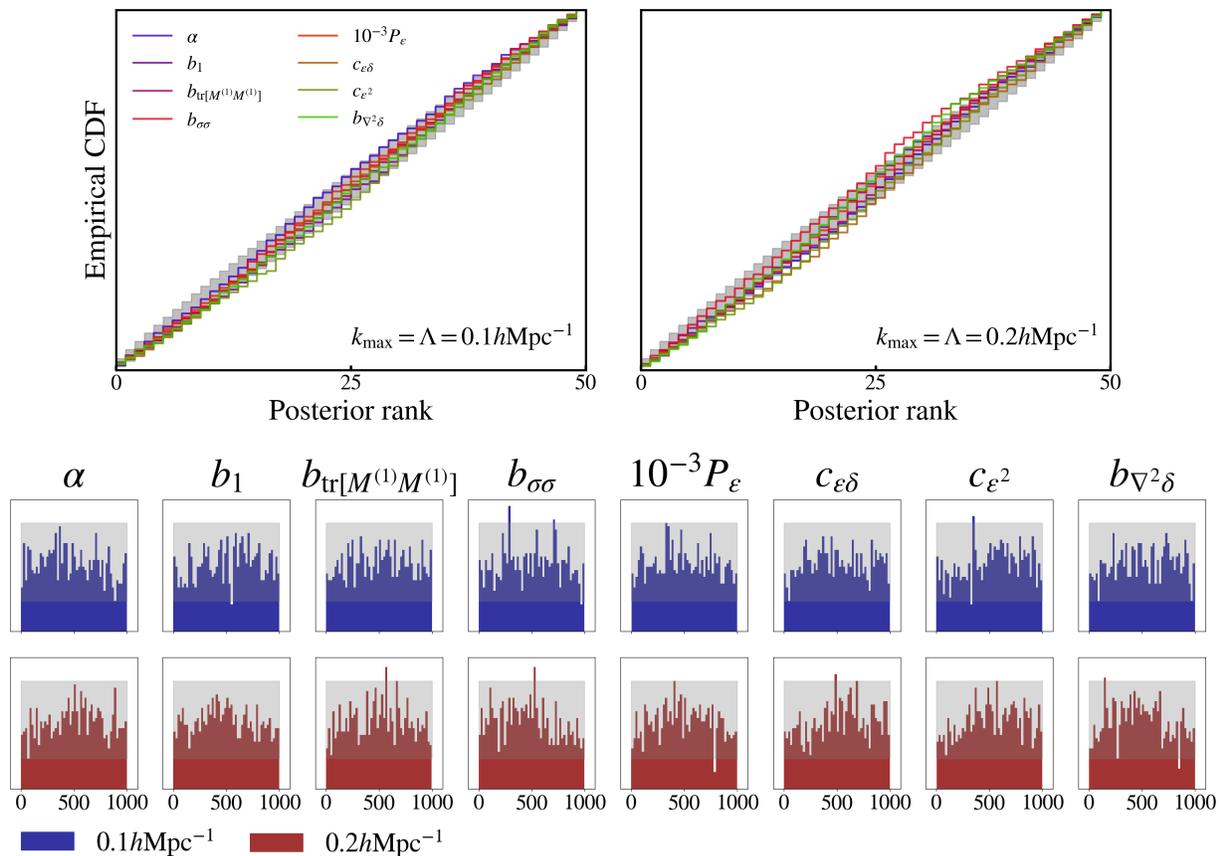


Figure 4.11: Same as Fig. 4.7 but for the case where the cosmological parameter  $\alpha$  is sampled as well, with  $N_{\text{sim}} = 10^5$ .

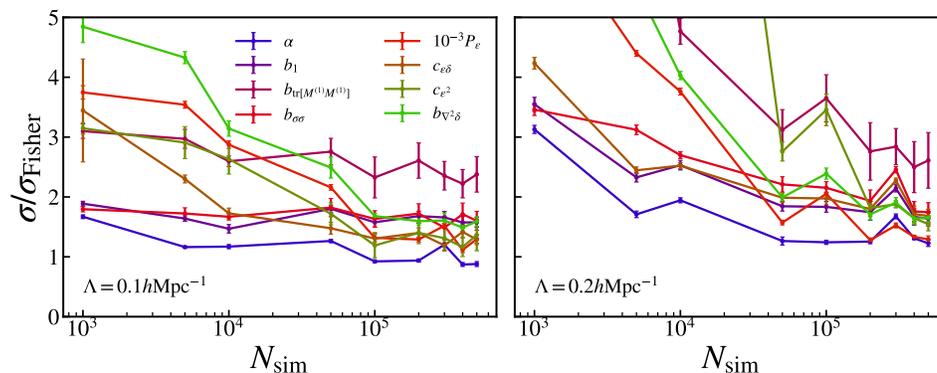


Figure 4.12: Same as Figure 4.8, but showing convergence for the full case where the cosmological parameter  $\alpha$  is sampled.

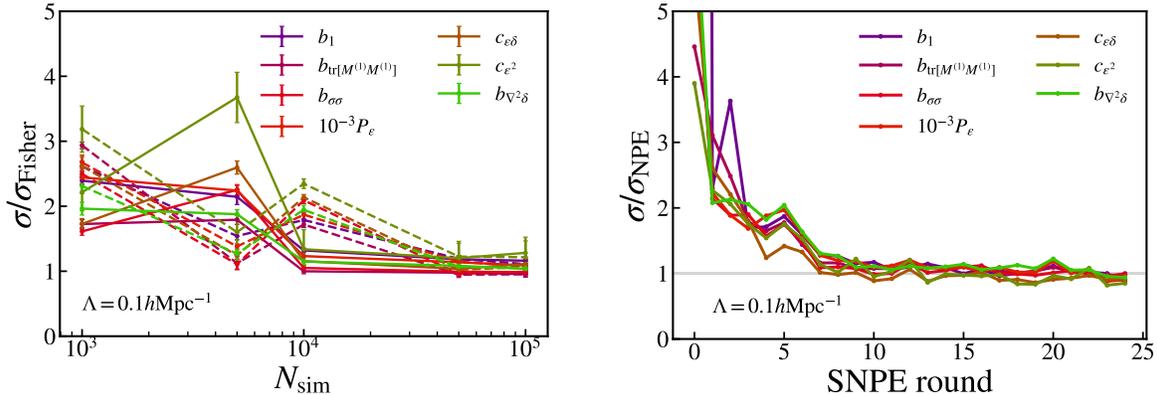


Figure 4.13: *Left*: Convergence of the standard deviation of the posterior of each parameter with respect to the simulation budget  $N_{\text{sim}}$  normalized by their Fisher prediction. Solid lines denote the case where the data vector corresponds to the raw spectra, while dashed lines display the results using the normalized spectra. *Right*: Standard deviation of the parameters from SNPE starting from the prior of Eq. (4.14) as a function of its rounds, where  $10^3$  simulations are sampled at each round, divided by the NPE standard deviation of each parameter obtained from  $N_{\text{sim}} = 10^5$ .

case, besides providing a comparison of bispectrum with power-spectrum-only constraints. We also compare the results with SBI density estimators other than NPE (namely SNPE and NLE), as well as the impact of a more complex network architecture or using an ensemble of networks instead of a single realization for posterior estimation.

## 4.4 Tests of the inference

In this section, we analyse some details of our pipeline focusing on a particular case, namely the one where the cosmological parameter  $\alpha$  is fixed and  $k_{\text{max}} = \Lambda = 0.1 h\text{Mpc}^{-1}$ . We will be referring to the full case using NPE and  $N_{\text{sim}} = 10^5$  unless stated otherwise; note that the following conclusions hold qualitatively the same for the other cases (using a simulation budget  $N_{\text{sim}}$  that guarantees convergence).

**Data vector normalization.** It is well known that usually machine learning techniques perform better with normalized values. Since our parameters  $\theta$  are already of order one, instead of using the data vector  $\mathbf{x}$  as the raw power spectrum and bispectrum for density estimation, we also test normalizing them as  $P(k)/P_L(k)$  and  $B(k_1, k_2, k_3)/[P_L(k_1)P_L(k_2) + 2 \text{ perm.}]$ . However, as we can see in the left panel of Figure 4.13, the normalization does not lead to faster convergence in this case.

**SNPE.** The right hand side of Figure 4.13 shows the results of SNPE, the sequential version of NPE. As we can see, the standard deviations converge to the ones corresponding

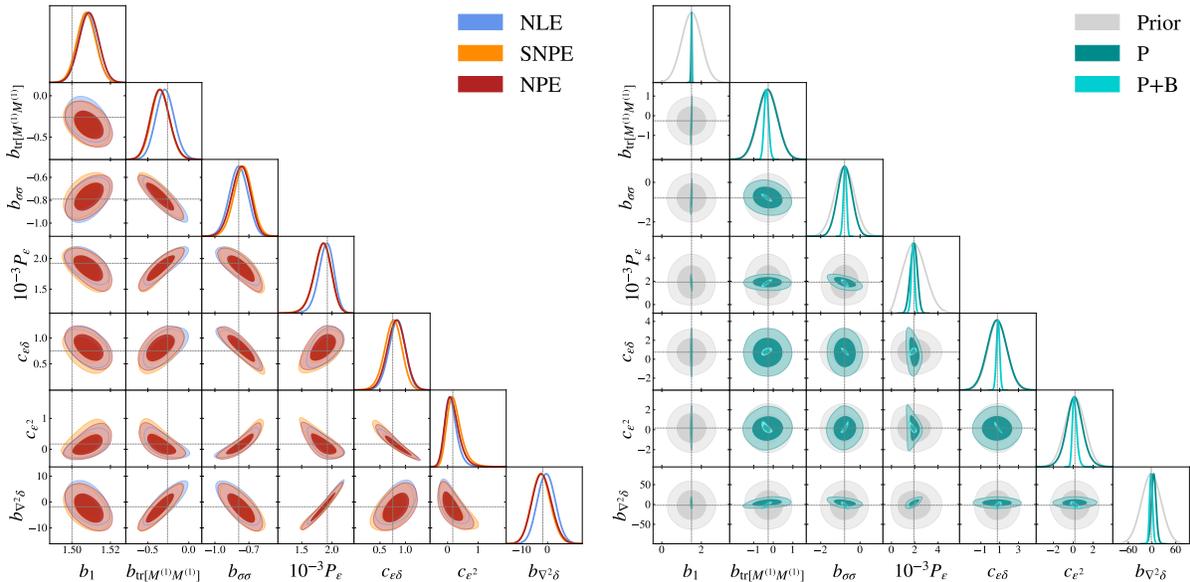


Figure 4.14: *Left*: posteriors obtained with NLE, NPE and SNPE (at round 22). *Right*: Parameter posterior comparing the prior range in grey with the constraints from power spectrum only (dark cyan) and the power spectrum combined with the bispectrum (cyan).

to the final NPE posterior, what further confirms our posterior convergence. As aforementioned, although this case uses less simulations for convergence (e.g., after 15 rounds, where convergence seems to be safely reached, one would have used a total of only  $N_{\text{sim}} \sim 10^4$  simulations), performing SBC tests on such non-amortized posteriors is a heavy computational task.

**Neural density estimation algorithm.** We compare the posteriors obtained from NPE, NLE and SNPE (at round 22) on the left side of Figure 4.14. As expected, all posteriors look very similar. The small deviations could be due to a not sufficiently flexible model or convergence issues. We explore the impact of network architecture below.

**Bispectrum constraining power.** We can also compare our analysis with a power-spectrum only analysis, i.e. the case where the data vector consists solely of the power spectrum in the same  $k$  bins. We can see on the right hand side of Figure 4.14 that, as expected,  $b_1$  and  $10^{-3} P_\epsilon$  are the most well-constrained parameters when considering only the power spectrum. Although the power spectrum does include higher-order (loop) contributions, these have very similar shape and are of limited use in disentangling the second-order bias parameters. As a result, essentially all of these parameters turn out to be prior dominated, i.e. the bispectrum is essential for constraints on higher-order bias and  $\sigma_8$ .

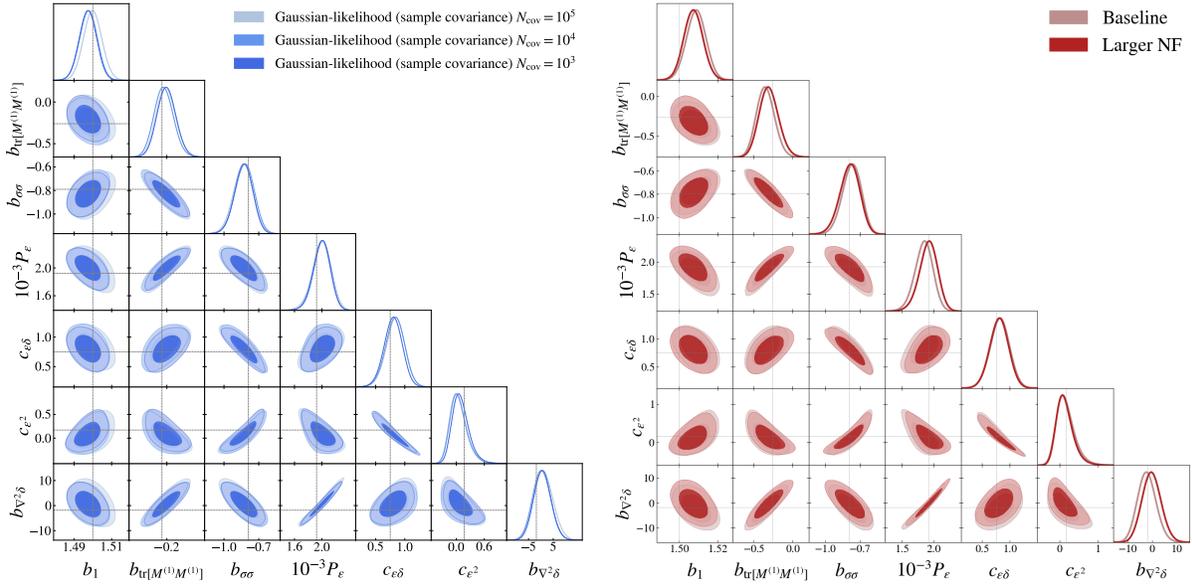


Figure 4.15: *Left*: impact of the number of simulations for covariance estimation on the final posterior for the Gaussian-likelihood case with sample covariance. *Right*: impact of doubling the number of hidden units and number of transforms.

**Number of covariance estimates.** On the left hand side of Figure 4.15 we can see how  $N_{\text{cov}}$ , the number of simulations used to estimate the covariance, changes the contours for the Gaussian-likelihood case with sample covariance. As expected, fewer simulations tend to underestimate the errors, although the effect is minor in this case.

**Network architecture.** We have also tested how increasing the complexity of the flow can affect the final posterior densities. We show in the right panel of Figure 4.15 that a more complex model, denoted as “larger NF” (normalizing flow), where we increased the hidden units from 50 to 100 and the number of transforms from 5 to 10, leads to basically the same posteriors.

**Ensemble network.** Figure 4.16 shows the same results as Figure 4.9, but with an ensemble of networks instead of a single realization. For that, we independently train 10 different posteriors for the three different cases with the same simulation budget of  $N_{\text{sim}} = 10^5$ , and then sample  $10^4$  posterior samples for each of the 10 estimated posteriors. This can give us an estimate of the error associated with the posterior estimation itself, and as we can see the errors are indeed larger; however, the trends are very similar and further confirm our previous findings.

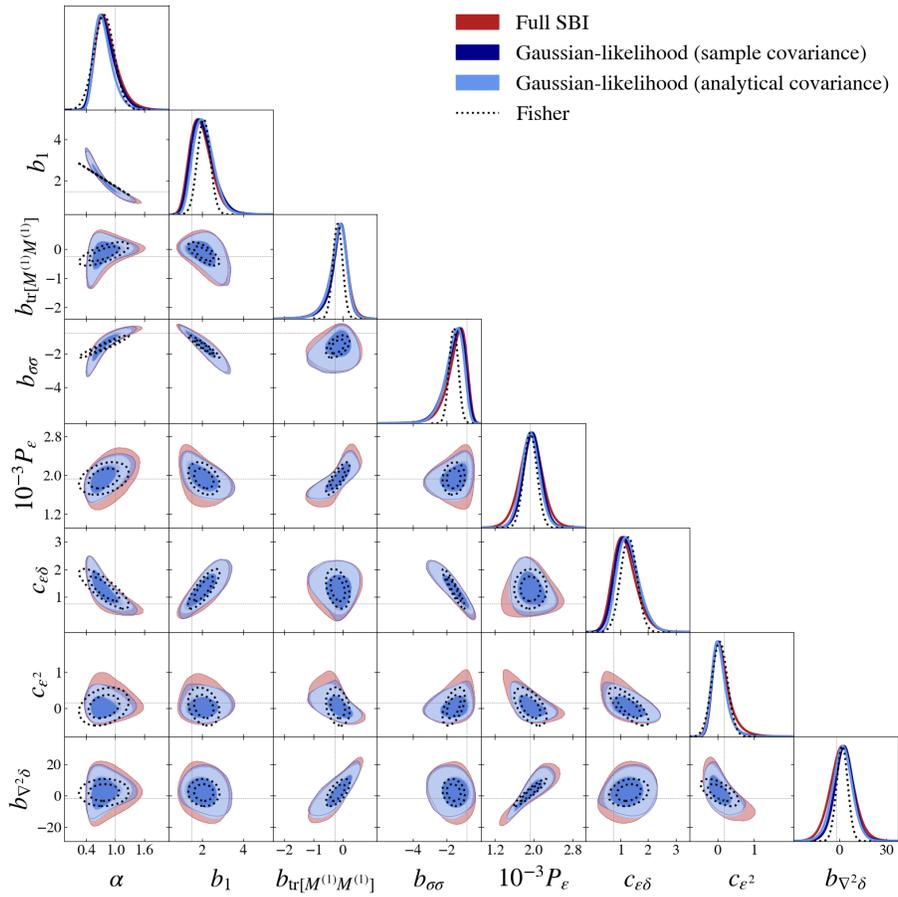


Figure 4.16: Same as Fig. 4.9, but using an ensemble of networks instead of a single posterior estimation.

## 4.5 Conclusions

We have explored for the first time how SBI performs within the context of a forward model based on the EFTofLSS and the bias expansion with the goal of constraining cosmological, noise and bias parameters from the galaxy power spectrum and bispectrum. First, we have demonstrated that SBI can successfully recover the expected posterior in controlled cases where the data vector is sampled from a Gaussian likelihood. Considering an Euclid-like mock tracer sample, we conclude that the non-Gaussianity of the lowest order  $n$ -point functions on large scales does not impact the constraints on  $\sigma_8$  for the scale ranges considered in this work, namely  $k_{\max} = 0.1h\text{Mpc}^{-1}$  and  $0.2h\text{Mpc}^{-1}$ . In these cases, adding off-diagonal terms to the covariance has a larger impact than relaxing the Gaussian likelihood assumption. Our estimated posteriors passed the SBC test and, within this specific context, we estimated the number of simulations needed for convergence to be of the order of  $10^5$  simulations.

It is important to emphasize that, while SBI serves as a versatile and powerful tool for cosmological inference, our utilization of machine-learning is strictly limited to the statistical aspect of our inference process, specifically density estimation. This entails essentially a fitting of the posterior distribution or likelihood function derived from a set of simulations, allowing us to overcome the limitations of standard inference techniques that rely on analytical approximations for the selected summary statistics. Most importantly, this work is performed within the framework of a forward model rigorously constructed to ensure accuracy on large scales. In this context, our understanding of the underlying physics benefits from years of dedicated study within the EFTofLSS community, providing us with a high level of control and intuitive comprehension, especially when employing the power spectrum and bispectrum as summary statistics.

It is paramount to validate and establish the reliability of SBI in well-understood regimes, such as those under consideration in this study. This validation process represents a crucial first step before delving into more complicated scenarios, where intuitive comprehension may be lacking. In addition, knowledge of an explicit likelihood that is at least valid for some elements of the data vector (for example on very large scales) allows for the possibility for explicit combinations of analytical likelihoods and SBI [150].

As the next direct step of this work, we will test our inference pipeline on dark-matter halos from N-body simulations and test different compression schemes other than  $n$ -point functions. It would also be interesting to investigate the impact of the details of the forward model that differ in our approach and current EFT-based analysis techniques based from codes such as CLASS-PT or PyBird (see the discussion in [185]). We also aim to use SBI to investigate whether the low- $k$  non-Gaussianities of the lowest order  $n$ -point functions impact the inference of cosmological parameters related to primordial non-Gaussianities.

In future work, we plan to extend our analysis to redshift space, using the `LEFTfield` forward model extension presented in [205], and to sample more cosmological parameters and to improve the realism of the forward model by incorporating systematic effects and masks in `LEFTfield`. We are particularly interested in comparing the results from field-

level cosmological inference, which explicitly marginalizes over the initial conditions using HMC, to SBI inference, both using  $n$ -point functions as summary statistics and other “field-level” compression mechanisms which learn the summary statistics directly from the galaxy density field.

Regarding our SBI pipeline, we are planning to explore prior truncation schemes [111, 145], using techniques such as the Sobol’ sequence [202] and active learning [15]. We would also like to improve our density estimation methods by hyperparameter tuning, besides trying other SBI algorithms such as Neural Ratio Estimation [65, 70, 95, 145, 146]. In particular, it would be interesting to test Truncated Marginal Ratio Estimation (TMNRE) [55, 145] for marginalizing over the bias parameters. Lastly, it would be interesting to also explore other SBI diagnostics besides SBC [127, 136, 230].



# Chapter 5

## How much information can be extracted from galaxy clustering at the field level?

*This chapter is an adaptation of [157] (published in PRL). Authors: Nhat-Minh Nguyen, Fabian Schmidt, Beatriz Tucci, Martin Reinecke & Andrija Kostić.*

My main contribution to this work was performing SBI with `LEFTfield` on dark-matter halos using the power-spectrum and bispectrum, thus providing a fair comparison to field-level inference constraints on the cosmological parameter  $\sigma_8$ . All field-level inference analysis presented here were conducted by the first two authors, Nhat-Minh Nguyen and Fabian Schmidt.

### 5.1 Introduction

Advancements in cosmological surveys of LSS have transitioned from angular clustering measurements using photographic plates to three-dimensional clustering analyses with spectroscopic data. Despite this progress, statistical techniques for analyzing galaxy clustering still predominantly focus on two- and three-point correlation functions, leaving open the question:

*How much cosmological information can be robustly extracted from galaxies and LSS tracers?*

This work addresses this question by comparing constraints on the amplitude of linear matter fluctuations,  $\sigma_8$  [32, 49, 51], obtained from (1) full field-level statistics and (2) power spectrum plus bispectrum (P+B) summary statistics, within the framework of the EFTofLSS.

This work provides the first demonstration that robust  $\sigma_8$  constraints can be extracted from nonlinear clustering of dark-matter halos in N-body simulations using field-level inference. Unlike previous studies relying on neural network compression [90, 129], which may lose information, our approach explicitly samples the full posterior of  $[\sigma_8, \delta^{(1)}]$ , ensuring optimal constraints.

Our analysis advances prior works in three key ways: (1) We marginalize over unknown initial conditions, making our constraints more applicable to real data. (2) We analyze dark-matter halos instead of idealized tracers, testing the robustness of the EFT-based forward model. (3) We perform a direct comparison between field-level and summary statistics constraints using the same data and model, providing a fair assessment of the information content in nonlinear galaxy clustering.

## 5.2 Methods

**Data.** We analyze two halo samples: **SNG**, comprising main halos in the  $\log_{10} M_{200m} = 12.5 - 14.0 h^{-1} M_{\odot}$  mass range at  $z = 0.50$ , identified using **ROCKSTAR** [33] in an N-body simulation with  $\sigma_{8,\text{true}} = 0.850$ , volume  $L^3 = (2000 h^{-1} \text{Mpc})^3$ , and  $N_{\text{particle}} = 1536^3$  particles of mass  $M_{\text{particle}} = 1.8 \times 10^{11} h^{-1} M_{\odot}$  [191]. Additionally, the **Uchuu** sample consists of halos in the  $\log_{10} M_{200m} = 12.0 - 13.5 h^{-1} M_{\odot}$  mass range at  $z = 1.03$ , identified in the **Uchuu** simulation [101], with  $\sigma_{8,\text{true}} = 0.816$ , the same volume, but at a higher resolution of  $N_{\text{particle}} = 12800^3$  and  $M_{\text{particle}} = 3.27 \times 10^8 h^{-1} M_{\odot}$ .

**LEFTfield settings.** We use 2LPT and Eulerian bias up to third order,

$$O \in [\delta, \delta^2, K^2, \delta^3, K^3, \delta K^2, O_{\text{td}}, \nabla^2 \delta], \quad (5.1)$$

and Gaussian stochasticity

$$\sigma_{\epsilon}(k) = \sigma_{\epsilon,0} [1 + \sigma_{\epsilon,k^2} k^2]. \quad (5.2)$$

These bias operators are employed in the Forward-Based Inference (FBI) and Simulation-Based Inference (SBI)  $P+B$  analyses, utilizing **LEFTfield**. The second cubic sharp- $k$  filter to the Eulerian density is applied with a cutoff  $\Lambda_{\text{bias}} = 1.2 k_{\text{max}}$ . Throughout this work, we adopt  $\Lambda = 1.2 \Lambda_{\text{bias}} = 1.44 k_{\text{max}}$ .

**New developments.** Relative to previous work [118, 190, 192, 217], we introduce: (1) a third-order galaxy bias model for improved accuracy, (2) a non-uniform Fast Fourier Transform (NUFFT) [29] for enhanced numerical convergence, and (3) a modified analysis cutoff  $k_{\text{max}} = \Lambda/1.2$  to mitigate higher-derivative contributions.

**Field-Level Bayesian Inference (FBI).** The FBI pipeline explicitly models the likelihood of the galaxy field  $\mathcal{L}_{\text{FBI}}^{\text{expl}}$  following [47], assuming Gaussianity of stochasticity and

marginalizing over  $\epsilon$  analytically [46, 193]. The likelihood takes the form:

$$\mathcal{L}_{\text{FBI}}^{\text{expl.}} \left( \delta_g^{\text{obs.}} \middle| \hat{s}, \alpha, \{b_O\}, \{\sigma_\epsilon\} \right) = -\frac{1}{2} \sum_{\mathbf{k} > 0}^{|\mathbf{k}| < k_{\text{max}}} \left[ \ln 2\pi\sigma_\epsilon^2(k) + \frac{1}{\sigma_\epsilon^2(k)} \left| \delta_g^{\text{obs.}}(\mathbf{k}) - \sum_O b_O O[\alpha, \hat{s}](\mathbf{k}) \right|^2 \right], \quad (5.3)$$

where the sum applies a spherical sharp- $k$  filter up to  $k_{\text{max}}$ . Bias parameters  $\{b_O\}$  are marginalized with weakly informative priors. The posterior is explored using Hamiltonian Monte Carlo (HMC) [151] for  $\hat{s}$  and slice sampling [152] for  $[\alpha, \{\sigma_\epsilon\}]$ .

**Simulation-Based Inference (SBI) P+B** SBI bypasses an explicit likelihood by learning the posterior from simulations [57], following [217]. Parameters  $\theta \equiv [\alpha, \{b_O\}, \{\sigma_\epsilon\}]$  are sampled and used to generate galaxy fields via Eq. (5.3). The power spectrum  $P$  and bispectrum  $B$  are measured:

$$\langle \delta_g(\mathbf{k}) \delta_g(\mathbf{k}') \rangle = P(k) (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}'), \quad (5.4)$$

$$\langle \delta_g(\mathbf{k}_1) \delta_g(\mathbf{k}_2) \delta_g(\mathbf{k}_3) \rangle = B(k_1, k_2, k_3) (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3), \quad (5.5)$$

with P+B data vectors containing  $N_{\text{bin}} + N_{\text{triangle}}$  elements. Neural posterior estimation (NPE) [85] is performed using masked autoregressive flows [164] from `sbi` [211].

**Comparison and Validation.** The target posteriors for FBI and SBI P+B are

$$\mathcal{P}_{\text{FBI}} \left( \alpha, \{b_O\}, \{\sigma_\epsilon\} \middle| \delta_g^{\text{obs.}} \right) \propto \int \mathcal{D}\hat{s} \mathcal{P}(\hat{s}) \mathcal{L}_{\text{FBI}}^{\text{expl.}} \left( \delta_g^{\text{obs.}} \middle| \hat{s}, \alpha, \{b_O\}, \{\sigma_\epsilon\} \right) \mathcal{P}(\alpha, \{b_O\}, \{\sigma_\epsilon\}), \quad (5.6)$$

$$\mathcal{P}_{\text{P+B}} \left( \alpha, \{b_O\}, \{\sigma_\epsilon\} \middle| P[\delta_g^{\text{obs.}}], B[\delta_g^{\text{obs.}}] \right) \propto \mathcal{L}_{\text{P+B}}^{\text{impl.}} \left( P[\delta_g^{\text{obs.}}], B[\delta_g^{\text{obs.}}] \middle| \alpha, \{b_O\}, \{\sigma_\epsilon\} \right) \mathcal{P}(\alpha, \{b_O\}, \{\sigma_\epsilon\}). \quad (5.7)$$

We validate the SBI posterior using simulation-based calibration (SBC) [210] and convergence tests. Our analysis assumes Gaussian noise, and comparisons with alternative stochastic models are detailed in Sec. 5.5. Overall, we find consistency between the different inference methods, reinforcing the robustness of SBI P+B in extracting cosmological information.

**FBI priors.** The priors assumed in the FBI analysis for the amplitude rescaling parameter  $\alpha$ , bias coefficients  $\{b_O\}$  from Eq. (5.1), and noise parameters  $\{\sigma_\epsilon\}$  are:

$$\begin{aligned} \mathcal{P}(\alpha) &= \mathcal{U}(0.5, 1.5), \\ \mathcal{P}(b_\delta) &= \mathcal{N}(1.0, 5.0), \\ \mathcal{P}(b_{\delta^2}) &= \mathcal{P}(b_{K^2}) = \mathcal{P}(b_{\delta^3}) = \mathcal{P}(b_{\delta K^2}) = \mathcal{P}(b_{K^3}) = \mathcal{P}(b_{O_{\text{td}}}) = \mathcal{N}(0.0, 1.0), \\ \mathcal{P}(b_{\nabla^2 \delta}) &= \mathcal{N}(0.0, 5.0), \\ \mathcal{P}(\sigma_{\epsilon,0}) &= \mathcal{U}(0.8\sigma_{\epsilon,\text{Poisson}}, 100.), \quad \mathcal{P}(\sigma_{\epsilon,k^2}) = \mathcal{U}(-10.0, 100.0), \end{aligned} \quad (5.8)$$

## 5. How much information can be extracted from galaxy clustering at the field level?

where  $\mathcal{U}(a, b)$  denotes a uniform distribution between  $a$  and  $b$ , while  $\mathcal{N}(\mu, \sigma)$  is a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ .

The parameter  $\sigma_{\epsilon, \text{Poisson}}$  represents the expected noise level based on Poisson shot noise given the tracer number density. A lower bound of  $0.8\sigma_{\epsilon, \text{Poisson}}$  is imposed to prevent the MCMC chains from drifting to unrealistically low noise levels, a known issue still under investigation. This bound is significantly below previous halo stochasticity estimates [92, 194]. Our priors on  $\{b_O\}$  are consistent with other EFT-based cosmological analyses [38, 103]. Additionally, [38] demonstrates that variations in bias priors do not significantly impact  $\sigma_8$  constraints.

**SBI P+B priors.** For the SBC tests to reliably assess the uncertainty estimates of the inferred SBI posterior, the posterior must be amortized, meaning it should not be constrained to a single observation. However, conducting an amortized NPE using samples drawn from the broad FBI prior would demand an extremely large number of simulations to achieve convergence.

To address this, we adopt the approach outlined in [217]: (1) We begin training with the same priors used in the FBI analysis and implement a sequential NPE (SNPE) across multiple rounds, where each round’s posterior serves as the proposal distribution for the next. This iterative process helps discard prior regions where the posterior has negligible support. (2) For the final NPE posterior, training samples are drawn from a Gaussian distribution that is broader than the posterior obtained in the last SNPE iteration. This choice accelerates the SBC analysis and allows convergence with fewer simulations, while ensuring consistency between the SBI and FBI analyses and preventing any parameter from being dominated by its prior.

An additional refinement in SBI involves rescaling the bias parameters  $b_{O^{(n)}}$ , where  $n$  represents the order of the bias operators defined in Eq. (5.1). Specifically, they are adjusted as  $\alpha^n b_{O^{(n)}}$  during inference to mitigate prior volume effects and reduce degeneracies within the SBI data vector. Furthermore, the higher-derivative bias parameters  $b_{\nabla^2 \delta}$  and  $b_{\nabla^2 \epsilon}$  are rendered dimensionless by normalizing with  $R_*^{-2}$ , where  $R_*$  denotes the characteristic scale of halo formation. In this study, we assume  $R_* = 5h^{-1}\text{Mpc}$ .

### 5.3 Results

Our key findings are illustrated in Fig. 5.1, where we contrast the posterior distributions of  $\alpha$  obtained from FBI and SBI P+B analyses. All methods successfully recover the true value of  $\alpha = 1$  within the 68% confidence level.

For specific scale cutoffs, at  $k_{\text{max}} = 0.1 (0.12)h\text{Mpc}^{-1}$ , the constraints from FBI yield  $\alpha = 0.976 \pm 0.056$  ( $\alpha = 1.013 \pm 0.033$ ), corresponding to relative uncertainties of 5.9% and 3.6%, respectively. This represents an improvement by a factor of 3.5 (5.2) over the SBI P+B constraints, which result in  $\alpha = 1.014 \pm 0.200$  ( $\alpha = 0.872 \pm 0.170$ ). As expected, the enhancement of field-level constraints over summary statistics increases with  $k_{\text{max}}$ , reflecting the greater role of nonlinearities in the forward model at higher wavenumbers.

Both FBI and SBI P+B analyses remain consistent across the two  $k_{\max}$  values for each inference approach, with their results differing by at most  $0.2\text{-}\sigma$  ( $0.8\text{-}\sigma$ ). This agreement stems from their shared forward model, `LEFTfield`, which demonstrates its precision on these scales.

To assess the generality of these findings, we perform an analogous analysis on the publicly available `Uchuu` simulation. The results in Fig. 5.2 confirm this trend: FBI improves constraints on  $\alpha$  by factors of 1.9 (2.5) compared to SBI P+B. More precisely, at  $k_{\max} = 0.1$  ( $0.12$ ) $h\text{Mpc}^{-1}$ , the constraints from FBI are  $\alpha = 0.941 \pm 0.090$  ( $\alpha = 0.993 \pm 0.053$ ), whereas SBI P+B yields  $\alpha = 1.018 \pm 0.168$  ( $\alpha = 0.900 \pm 0.136$ ), demonstrating agreement within  $0.4\text{-}\sigma$  ( $0.6\text{-}\sigma$ ).

The improvement factors in the `Uchuu` sample are slightly lower than those in `SNG`, likely due to reduced contributions from displacement effects in higher-order statistics at higher redshifts, where the growth factor is smaller. Since these contributions enhance  $\sigma_8$  inference at the field level, their diminished impact at higher redshifts may explain the observed reduction in improvement.

For the variant SBI case shown in Sec. 5.5, which mirrors current standard P+B analyses, the improvement factors in  $\alpha$  constraints between FBI and SBI P+B are 3.5 (5.2) for `SNG` and 2.3 (3.5) for `Uchuu` at  $k_{\max} = 0.1$  ( $0.12$ ) $h\text{Mpc}^{-1}$ . This underscores the advantage of field-level inference over summary statistics.

## 5.4 Tests of the inference: FBI

**Galaxy Bias Marginalization.** In this study, we confirm this analytical marginalization by reanalyzing the `SNG` sample at  $k_{\max} = 0.10h\text{Mpc}^{-1}$ . We compare results using a likelihood that marginalizes over all bias terms in Eq. (5.1), except for the linear bias parameter  $b_\delta$ . In Fig. 5.3, we contrast the posteriors obtained with this “ $b_\delta$ -unmarg.” likelihood against those derived from the fiducial “ $\{b_O\}$ -marg.” likelihood in Eq. (3.46). The 1D and 2D marginalized posteriors are fully consistent across both approaches. The “ $b_\delta$ -unmarg.” likelihood reveals that FBI effectively constrains  $b_\delta$  and  $\alpha$ , demonstrating its capability to break the  $b_\delta - \sigma_8$  degeneracy using nonlinear clustering information.

It is worth noting that the “ $b_\delta$ -unmarg.” approach assumes a prior  $\mathcal{P}(b_\delta|\text{unmarg}) = \mathcal{N}(1.0, 5.0)\mathcal{U}(0., 7.)$ , which differs slightly from the priors in the marginalized analysis. However, given the strong constraint on  $b_\delta$ , this minor discrepancy is unlikely to affect the results significantly. Additionally, the “ $b_\delta$ -unmarg.” chains achieved approximately 21 effective samples, significantly fewer than the fiducial chains [Table 5.1]. Therefore, our primary focus remains on the overall consistency between the marginalized and unmarginalized posteriors.

**Eulerian and Lagrangian Bias Representations.** For both the fiducial analyses within FBI and SBI using  $P + B$ , we employ the Eulerian framework for the EFT galaxy bias expansion, as elaborated in Sec. II of the Supplementary Material. An alternative

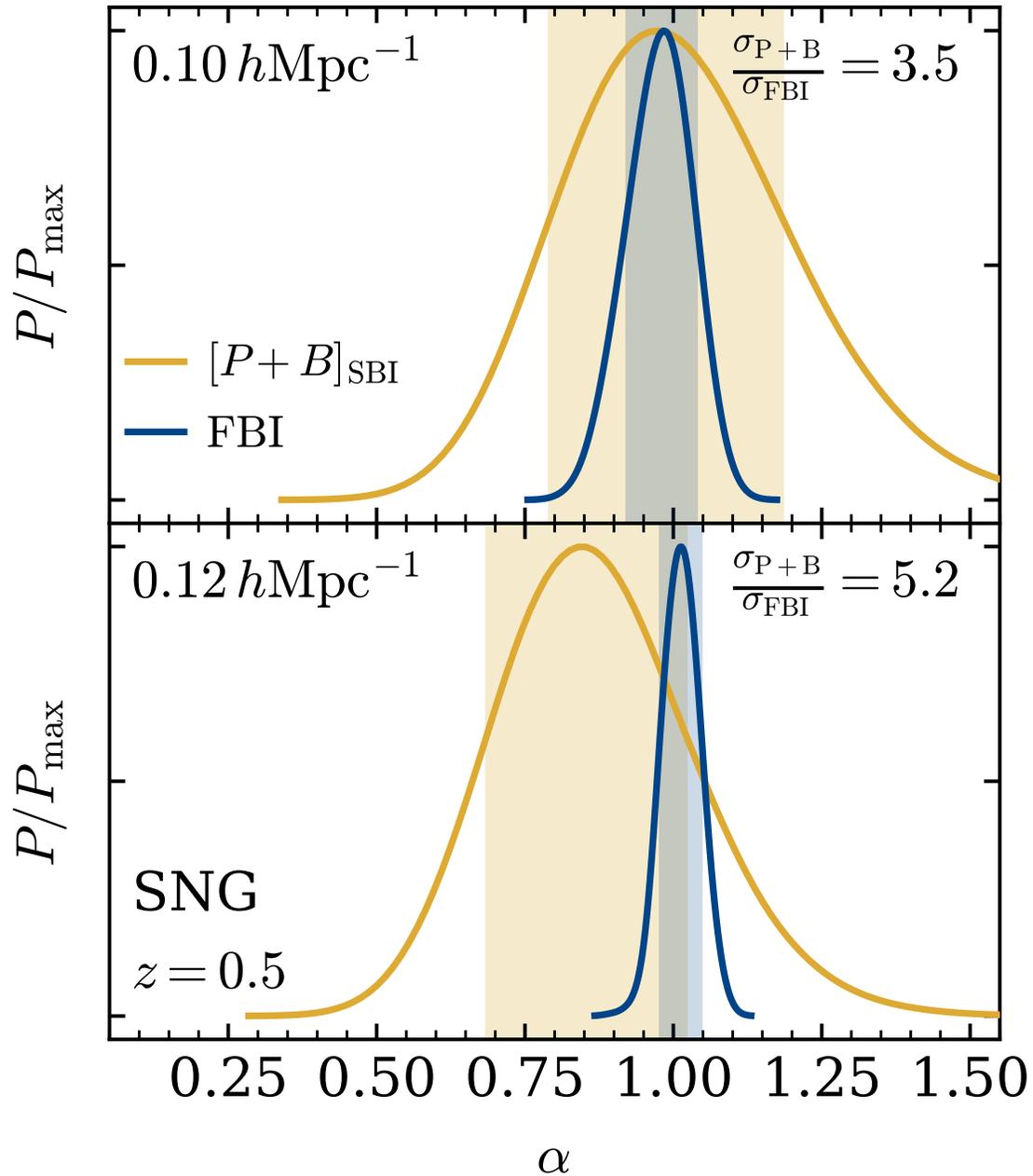


Figure 5.1: Posterior constraints on  $\alpha = \sigma_8 / \sigma_{8,\text{true}}$  from the SNG sample (see text) at  $k_{\text{max}} = [0.10, 0.12] h\text{Mpc}^{-1}$ . Vertical bands denote 68% confidence limits. The upper right corners indicate the ratio of 1- $\sigma$  constraints between FBI (blue) and SBI P+B (yellow).

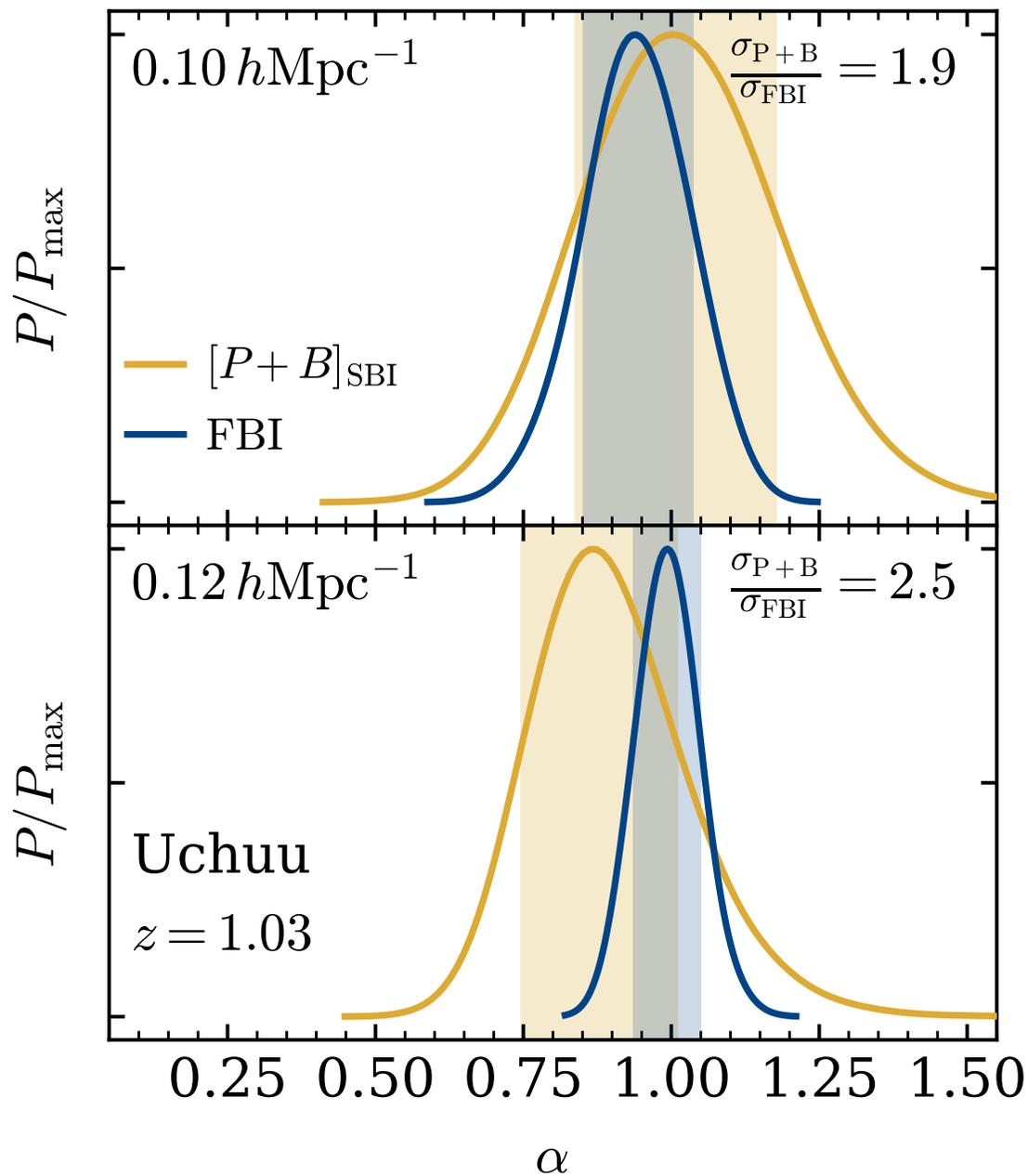


Figure 5.2: Similar to Fig. 5.1, but for the Uchuu sample (see text) at  $k_{\max} = [0.10, 0.12]h\text{Mpc}^{-1}$ .

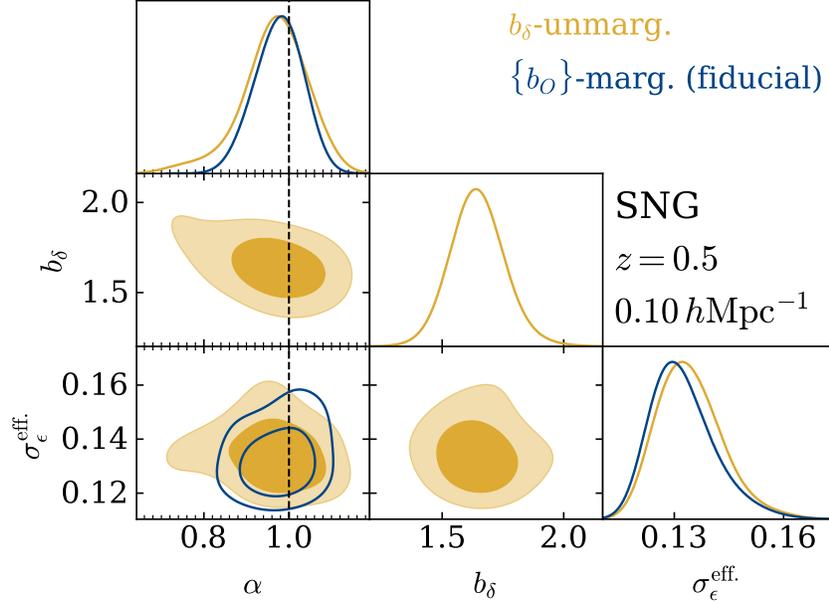


Figure 5.3: Posterior distributions comparing the fiducial  $\{b_O\}$ -marginalized likelihood (blue, open) and the  $b_\delta$ -unmarginalized likelihood (yellow, filled). The figure presents 1D and 2D constraints on  $\alpha = \sigma_8/\sigma_{8,\text{fid.}}$ , the linear bias parameter  $b_\delta$ , and the galaxy stochasticity parameters  $[\sigma_{\epsilon,0}, \sigma_{\epsilon,k^2}]$ . Contours indicate 68% and 95% confidence intervals.

formulation for the EFT bias terms  $O$  and bias expansion exists in Lagrangian space<sup>1</sup>. Specifically, one can define  $O$  at the Lagrangian coordinates  $\mathbf{q} = \mathbf{x}(\tau = 0)$  and transport them to their final Eulerian positions  $\mathbf{x}(\mathbf{q}, \tau)$ , following the methodology outlined in Sec. 3 of [190].

At any fixed order in perturbation theory, both bias expansions yield equivalent results, but deviations emerge at higher orders. Thus, comparing these approaches helps assess the relevance of higher-order corrections in bias modeling.

Fig. 5.4 presents the parameter constraints derived using each bias model for a fiducial halo sample at  $z = 0.5$ , considering a scale cutoff of  $k_{\text{max}} = 0.10 h\text{Mpc}^{-1}$ . The posteriors from both frameworks align well within statistical uncertainties, keeping in mind that the Lagrangian chains contain approximately 13 effective samples, implying limited precision in their posterior moments.

**MCMC Initialization in FBI.** Regardless of the initial conditions, a properly converged Markov Chain Monte Carlo (MCMC) chain samples from the same target distribution. Ensuring convergence is particularly crucial in high-dimensional parameter spaces such as those in FBI analyses.

Here, we explore two initialization strategies. For each FBI run, one MCMC chain is

<sup>1</sup>See Sec. 2.5.2-2.5.3 in [66].

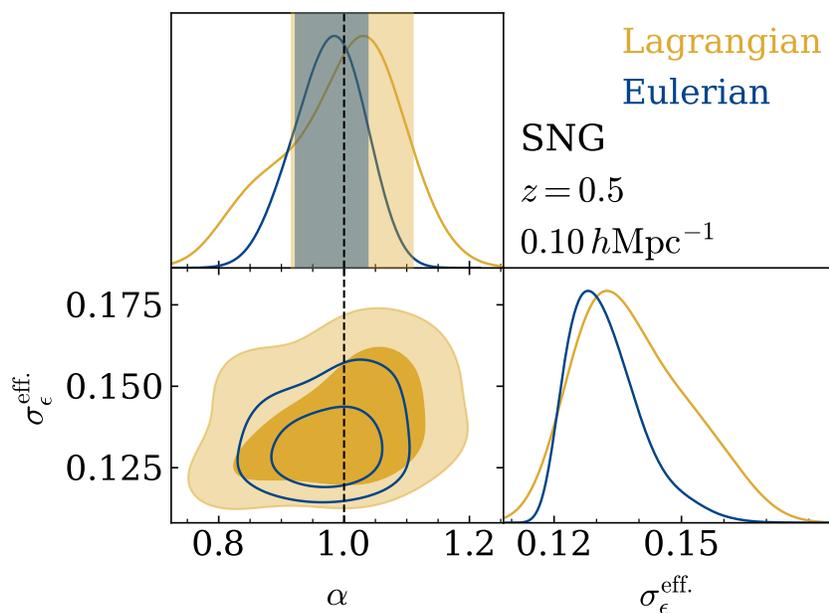


Figure 5.4: Comparison of posterior constraints obtained using the Eulerian bias (blue, open) and Lagrangian bias (yellow, filled). The figure displays constraints on the amplitude rescaling parameter  $\alpha = \sigma_8/\sigma_{8,\text{fid.}}$  and the effective noise amplitude  $\sigma_{\epsilon}^{\text{eff.}} = \sigma_{\epsilon,0}[1 + \sigma_{\epsilon,k^2} k_{\text{max}}^2]$ . Contours denote 68% and 95% confidence regions of the 2D marginal posteriors, while vertical bands represent the 68% intervals of the 1D marginal posteriors.

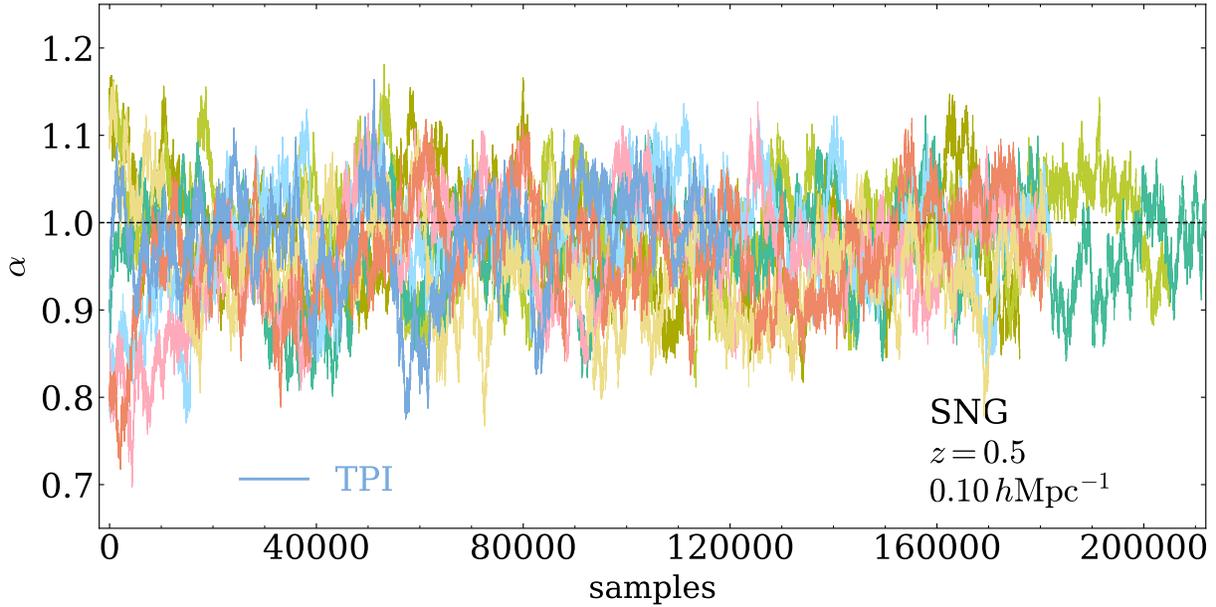


Figure 5.5: Trace plot of the amplitude rescaling parameter  $\alpha = \sigma_8/\sigma_{8,\text{fid}}$  from TPI and RPI chains in the FBI analysis. Different colors indicate distinct initial conditions.

initialized from the true initial conditions (true-phase initialization, TPI)  $\hat{s}_{\text{true}}$ , while additional chains begin from randomly selected initial conditions (random-phase initialization, RPI)  $\hat{s}$ . By comparing TPI and RPI behaviors, we assess convergence.

Fig. 5.5 illustrates the trace plot of the parameter  $\alpha$  in the FBI chains. After the warm-up phase, the TPI (light blue) and RPI chains become indistinguishable, indicating successful posterior sampling.

**Posterior Sampling of Initial Conditions.** Our field-level inference also constrains the initial conditions  $\hat{s}$  of the observed volume. Fig. 5.6 summarizes different statistical properties of the inferred initial condition posterior.

The top panel displays the histogram of  $\hat{s} - \hat{s}_{\text{true}}$ , where  $\hat{s}$  represents posterior samples, and  $\hat{s}_{\text{true}}$  denotes the true initial conditions. The distribution follows a normal  $\mathcal{N}(0, 1)$ , confirming that our inference correctly recovers the first two moments of  $\mathcal{P}(\hat{s}|\delta_g^{\text{obs.}})$ .

In the middle panel, we present the square root of the ratio between the inferred and true initial power spectra, a transfer function that remains consistent with unity across Fourier modes, as expected.

The bottom panel shows the variance of  $\hat{s} - \hat{s}_{\text{true}}$ , which approaches zero for an ideal forward model without noise. The inferred variance matches expectations from Poisson noise in the data (see [118] for further discussion).

**Parameter Posterior Convergence in FBI** Given the high-dimensional nature of FBI, ensuring posterior convergence is essential. Fig. 5.7 compares marginal posteriors of

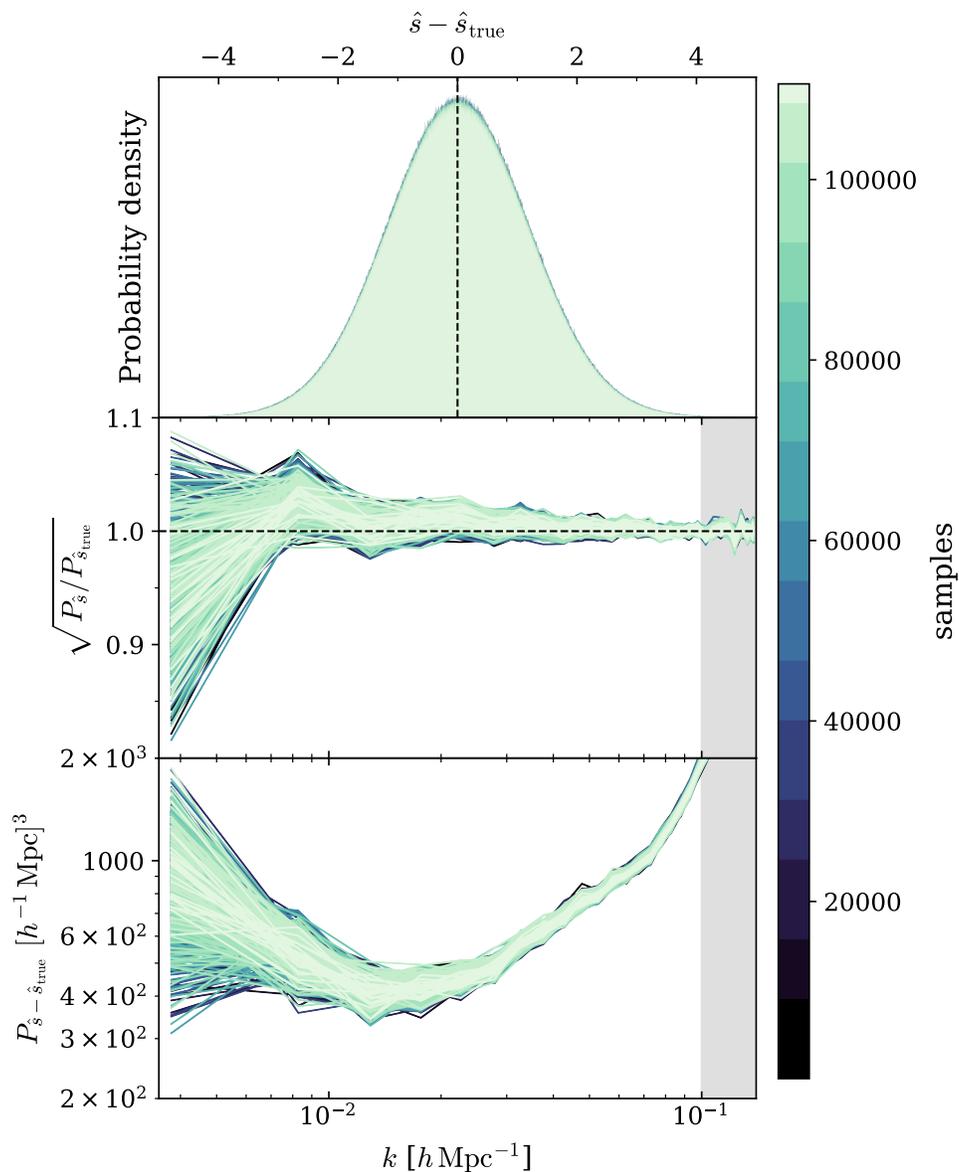


Figure 5.6: Trace plots of various statistics for the posterior initial conditions  $\hat{s}$  from the TPI chain in Fig. 5.5. The warm-up phase is omitted, and samples are thinned by a factor of 100 for clarity. Grey vertical bands mark scales beyond the cutoff  $k_{\max} = 0.1 h\text{Mpc}^{-1}$ . (Top) Histogram of  $\hat{s} - \hat{s}_{\text{true}}$ . (Middle) Square root of the ratio between inferred and true initial power spectra. (Bottom) Variance of  $\hat{s} - \hat{s}_{\text{true}}$ .

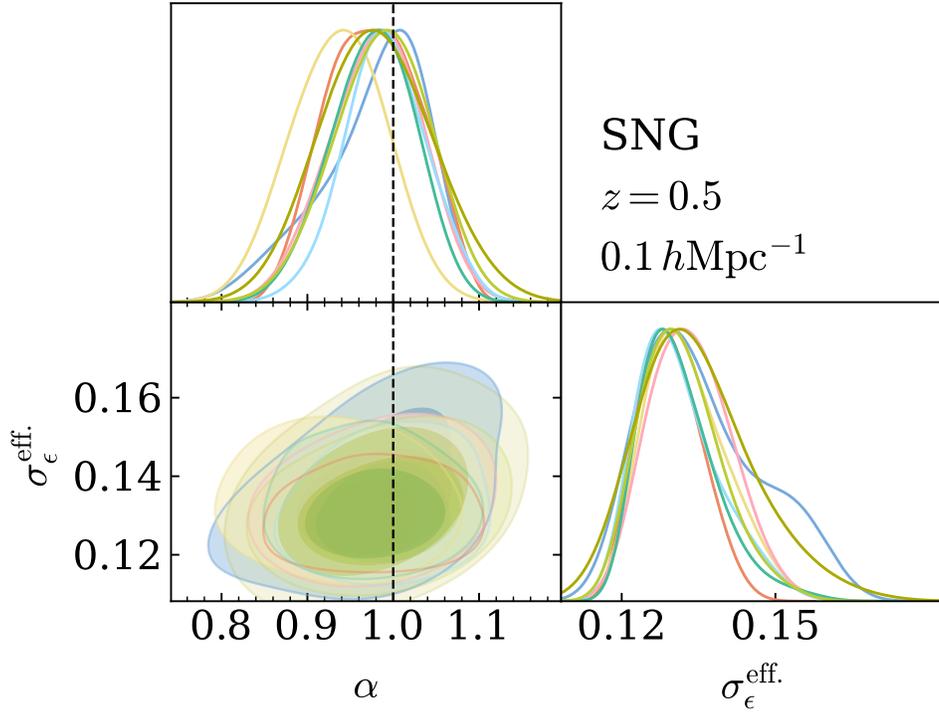


Figure 5.7: Posterior consistency of MCMC chains with different initializations. Colors correspond to the chains in Fig. 5.6. Contours represent the 68% and 95% credible intervals of the 1D and 2D marginal posteriors for  $[\alpha, \sigma_\epsilon^{\text{eff}}]$ .

$\alpha$  and  $\sigma_\epsilon^{\text{eff}}$  from MCMC chains with different initializations, at  $k_{\text{max}} = 0.1 h\text{Mpc}^{-1}$ . The consistency of posterior contours indicates that the chains adequately sample from the underlying distribution.

Since MCMC samples are correlated, the standard error in the mean of  $\alpha$  can be estimated as  $\sigma_\alpha/\text{ESS}$ , where ESS denotes the effective sample size. Table 5.1 provides ESS estimates for different FBI analyses.

## 5.5 Tests of the inference: SBI

**Analysis Setup.** For the considered scale limits, namely  $k_{\text{max}} = 0.1$  ( $0.12$ )  $h\text{Mpc}^{-1}$ , the power spectrum consists of  $N_{\text{bins}} = 15$  ( $18$ ) bins, leading to an SBI data vector dimension of  $D = 443$  ( $714$ ). We allocate a simulation budget of  $N_{\text{sim}} = 5 \times 10^5$  for most posteriors, except for the SNG halo sample at the higher cutoff, which utilizes  $N_{\text{sim}} = 10^6$ , as justified by the convergence analysis. For visualization purposes, we extract  $10^5$  posterior samples.

**Training and Hyperparameters.** We employ the SNPE method [85] using 10 atoms for atomic proposals and masked autoregressive flows [164] with 10 autoregressive layers.

| Analysis                                 | ESS [samples] |
|------------------------------------------|---------------|
| SNG, $k_{\max} = 0.10h\text{Mpc}^{-1}$   | 200           |
| SNG, $k_{\max} = 0.12h\text{Mpc}^{-1}$   | 100           |
| Uchuu, $k_{\max} = 0.10h\text{Mpc}^{-1}$ | 180           |
| Uchuu, $k_{\max} = 0.12h\text{Mpc}^{-1}$ | 60            |

Table 5.1: ESS estimates for  $\alpha$  in various FBI analyses.

Each layer consists of two fully connected tanh layers with 100 hidden units. Training is performed via stochastic minimization of the loss function using the Adam optimizer [112] with a learning rate of  $5 \times 10^{-4}$  and a batch size of 50. We reserve 10% of samples for validation and terminate training if validation loss remains stagnant for 20 consecutive epochs.

**Simulation-Based Calibration (SBC).** To evaluate posterior uncertainties, we apply SBC [210]. A well-calibrated posterior should exhibit uniform rank statistics for all parameters. Deviations from uniformity, such as U-shaped or inverted U-shaped distributions, indicate under- or overestimated posterior variance, respectively. As illustrated in Fig. 5.8, the uniform rank distribution confirms successful calibration.

**Convergence Assessment.** For SBI, posterior convergence is assessed by tracking the evolution of uncertainty in  $\alpha$  as a function of simulation budget,  $N_{\text{sim}}$ . Figure 5.9 depicts this relationship, with the standard deviations of  $\alpha$  normalized by Fisher-based constraints. This analysis extends to all SBI P+B parameters, but for brevity, we focus on  $\alpha$ .

**Fisher Analysis.** The standard deviation of parameter  $\theta_\alpha$  is estimated as  $\sqrt{(F^{-1})_{\alpha\alpha}}$ , where  $F$  denotes the Fisher information matrix, evaluated at the SBI posterior mode,  $\theta_{\text{MAP}}$ . Following [217],  $F$  is computed via numerical differentiation of the mean data vector, obtained by averaging over 1000 `LEFTfield` realizations. Sample covariance is estimated from  $10^5$  simulations evaluated at  $\theta_{\text{MAP}}$ . When considering all third-order bias parameters, the fiducial values of additional parameters are set to zero, and derivative step sizes are chosen based on SBI priors. We confirm convergence with respect to step size, fiducial point, and number of simulations. Figure 5.10 compares the SBI posterior to a Fisher analysis using sample covariance, with  $\log(b_\epsilon)$  displayed instead of  $b_\epsilon$  due to prior constraints.

**Non-Gaussian Stochasticity.** The primary analysis considers only Gaussian stochasticity in the galaxy field. Standard P+B analyses typically include additional stochastic contributions:  $\delta_g^{\text{stoch.}}(\mathbf{x}) = [b_\epsilon + b_{\epsilon\delta}\delta(\mathbf{x}) + b_{\nabla^2\epsilon}\nabla^2]\epsilon_0(\mathbf{x}) + b_{\epsilon^2}[\epsilon_0^2(\mathbf{x}) - \langle\epsilon_0^2\rangle]$ , where  $\epsilon_0$  is a unit Gaussian field [217]. While density-dependent noise dominates over the  $k^2$  term [46], its

## 5. How much information can be extracted from galaxy clustering at the field level?

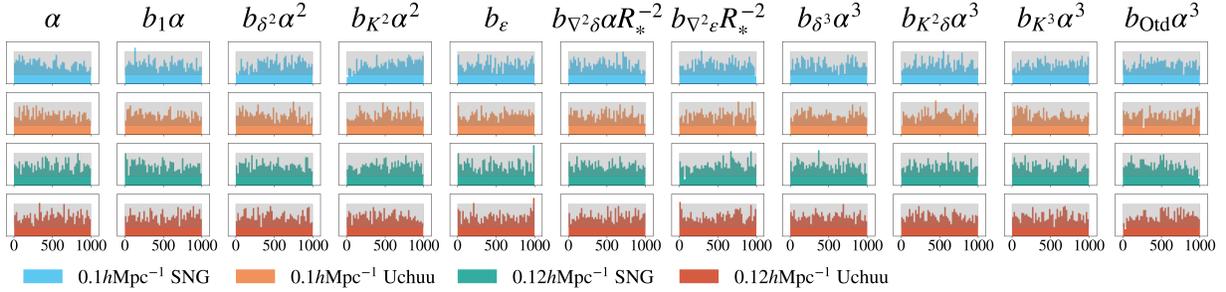


Figure 5.8: Rank distributions from the SBC tests for the SBI posteriors shown in Fig. 5.1 and Fig. 5.2. The grey shaded area indicates the 99% confidence interval of a uniform distribution. Upper, middle and lower panels correspond to the SNG ( $k_{\max} = 0.1h\text{Mpc}^{-1}$ ), Uchuu ( $k_{\max} = 0.1h\text{Mpc}^{-1}$ ), and SNG ( $k_{\max} = 0.12h\text{Mpc}^{-1}$ ) analyses, respectively.

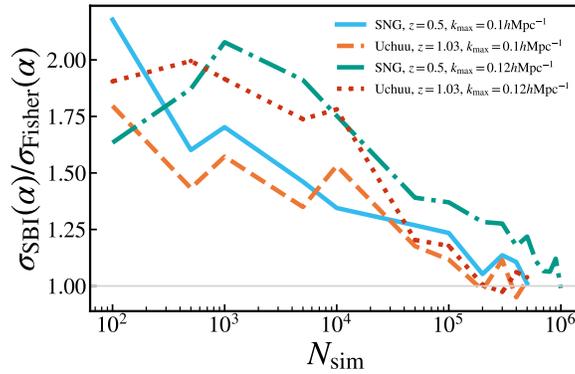


Figure 5.9: Convergence of the standard deviations of  $\alpha$  posteriors in SBI P+B analyses with increasing simulation budget  $N_{\text{sim}}$  used for posterior estimation. Each value is normalized by the corresponding  $\alpha$  constraint from a Fisher analysis, for each of the cases considered.

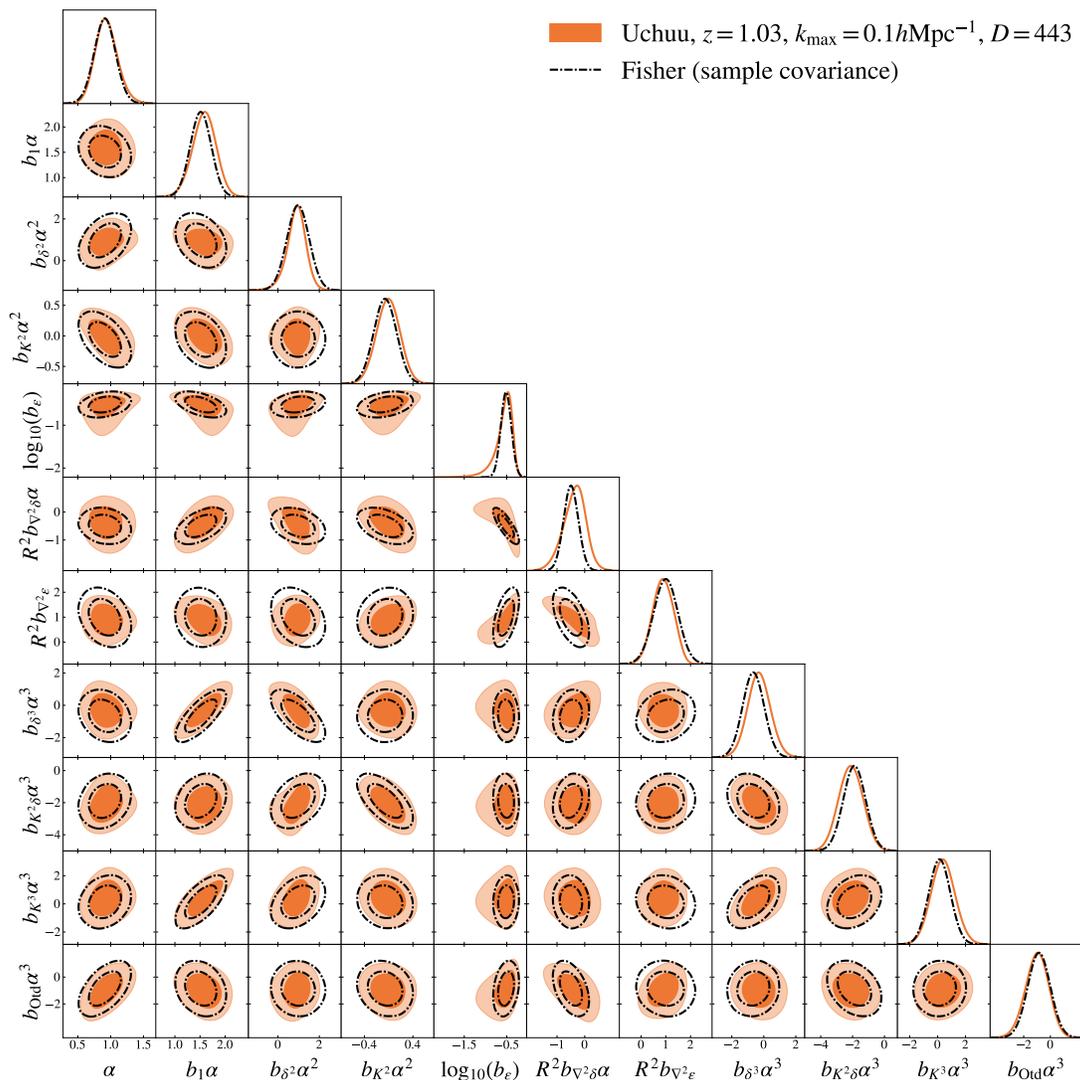


Figure 5.10: Parameter posteriors obtained in the SBI P+B analysis of the Uchuu halo sample at redshift  $z = 1.03$  with  $k_{\max} = 0.1 h\text{Mpc}^{-1}$ . Contours indicate 68% and 95% credible intervals. Dashed lines correspond to the Fisher analysis for this case, using sample covariance for the data vector.

## 5. How much information can be extracted from galaxy clustering at the field level?

implementation in FBI is technically challenging [47, 191]. Typically, only the third-order bias operator relevant at one-loop order [22], denoted  $O_{\text{td}}$ , is constrained.

We conduct an alternative analysis, fixing all third-order bias parameters to zero except  $b_{O_{\text{td}}}$ , which is inferred alongside  $b_{\epsilon_2}$  and  $b_{\epsilon_3}$ . Figure 5.11 shows that constraints on  $\alpha$  remain consistent with the Gaussian-only case. Convergence and calibration tests confirm robustness (not shown).

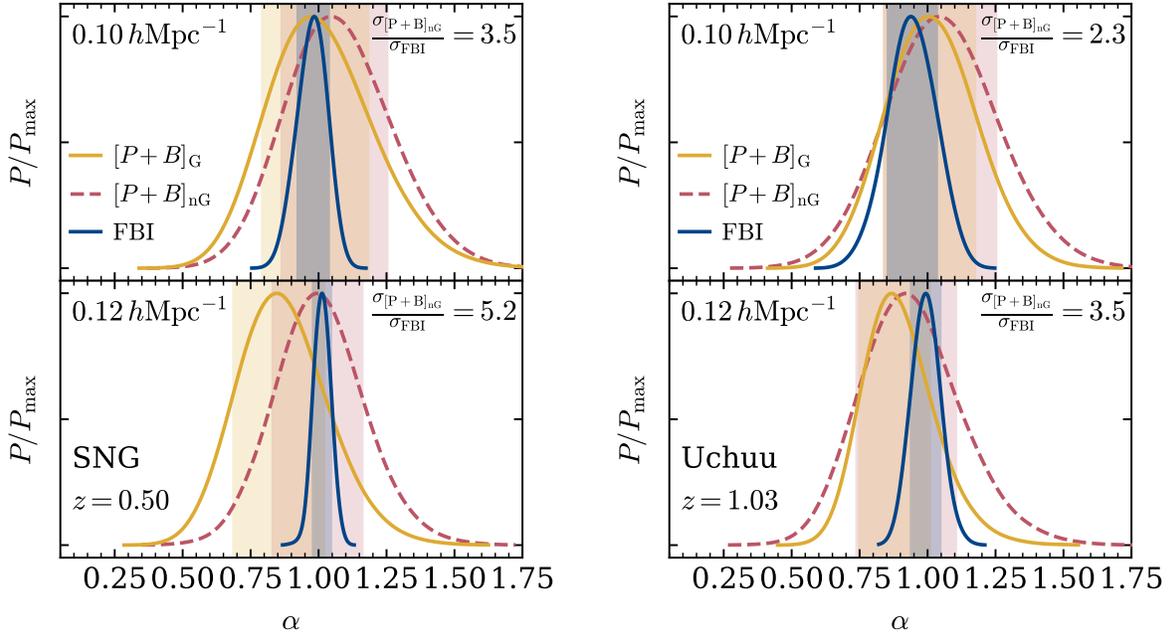


Figure 5.11: Comparison of  $\alpha$  posteriors in SBI P+B analyses, incorporating non-Gaussian stochasticity, as detailed in Section 5.5. Vertical bands denote 68% credible intervals. Upper right corners show the ratio of 1- $\sigma$  constraints between FBI (blue) and SBI with non-Gaussian stochasticity (pink).

## 5.6 Summary and Discussion

This chapter discusses the first constraints on  $\sigma_8$  derived from field-level inference of fully nonlinear biased tracers, specifically N-body halos in their comoving rest frame. These constraints rely on the EFTofLSS validity at quasilinear scales, with a rigorous marginalization over fully nonlinear scales. Although we have not included the effect of redshift-space distortions (RSD) from tracer peculiar velocities, typically used to constrain the combination  $f\sigma_8$ , our analysis still serves as a useful demonstration. First, when combined with  $f\sigma_8$  constraints obtained from *linear* RSD, a separate  $\sigma_8$  constraint enables direct

inference of the growth rate  $f$ , which serves as a sensitive probe of dark energy and gravity [133–135, 225], providing more insights than the combination  $f\sigma_8$  alone. Second, for a galaxy sample affected by line-of-sight-dependent selection bias, the dominant RSD contribution is degenerate with the primary selection bias term [98, 232], in the same way that  $b_1$  is degenerate with  $\sigma_8$  in the analysis here. However, higher-order protected RSD contributions can break the selection bias-growth rate degeneracy [7]. Therefore, even when considering a redshift-space analysis, breaking the  $b_1 - \sigma_8$  degeneracy remains crucial for constraining both  $\sigma_8$  and  $f$ .

We compare our field-level results with simulation-based inference using summary statistics like the power spectrum and bispectrum. By applying the same field-level forward model in both methods, we demonstrate that the field-level approach outperforms the summary statistics significantly [Figs. 5.1 and 5.2]. Although previous studies, e.g., [48, 193], have shown that field-level inference aligns with power spectrum and bispectrum analysis, this correspondence only holds when expanding the field-level likelihood to second-order in perturbation theory. Our forward model, however, includes third-order bias and thus accounts for information from higher  $n$ -point functions. Our results confirm that even on quasilinear scales, substantial cosmological information exists beyond the power spectrum and bispectrum.

In Chapter 6, we investigate whether additional low-order summaries, such as the trispectrum (4-point function), could extract more information.

While our current focus is on dark-matter halos, we show in [38] that this conclusion extends to simulated galaxies as well, consistent with EFT principles. Looking ahead to FBI on observed data, [45, 205] established the theoretical framework for incorporating RSD into the `LEFTfield` forward model and demonstrated its successful implementation. This advancement will facilitate field-level analysis in redshift space, unlocking further cosmological information as discussed above.

We emphasize that we have not attempted to push our analysis to smaller scales but instead targeted converged posteriors with conservative scale cuts at  $k_{\max} \leq 0.12h\text{Mpc}^{-1}$ , in light of the fact that the computational cost of our forward model scales as  $\sim \Lambda^3 \ln \Lambda$ . Even with these scale cuts, our results show that field-level inference provides robust constraints on structure growth, independent of the growth rate  $f$ , at the few-percent level, even within a modest volume of  $8(h^{-1}\text{Gpc})^3$ . This should allow for improved constraints on cosmological parameters in both standard  $\Lambda\text{CDM}$  and extended models, using upcoming DESI [8, 82], Euclid [40, 43], and PFS [209] data.

5. How much information can be extracted from galaxy clustering at the field level?

---

# Chapter 6

## The information content in the galaxy trispectrum

*This chapter is an adaptation of a paper in preparation (to be submitted to JCAP). Author: Beatriz Tucci. Since it is a work in progress, we report the current status and next steps here.*

### 6.1 Trispectrum

Given the evidence presented in Chapter 5 regarding the information content encoded at the field level compared to the galaxy power-spectrum and bispectrum, we now turn our attention to investigating the information content in the galaxy trispectrum. The galaxy trispectrum has been used in cosmological contexts to infer parity violation and primordial non-Gaussianities. However, it has never been employed for full-shape cosmological inference due to challenges related to its high dimensionality.

One of the primary difficulties is estimating the galaxy trispectrum covariance, both in mock data (due to the large number of dimensions) and when attempting to invert the covariance matrix. Furthermore, the full model for the galaxy trispectrum in redshift space remains unclear, as does the identification of the relevant stochastic terms.

In this chapter, we present the first steps towards providing, for the first time, an assessment on the cosmological information content in the galaxy trispectrum. By using SBI, the trispectrum covariance, along with higher-order moments, are automatically incorporated when learning the posterior distribution. Using `LEFTfield`, we explore the information content of the galaxy trispectrum in comparison to using the power-spectrum and bispectrum alone.

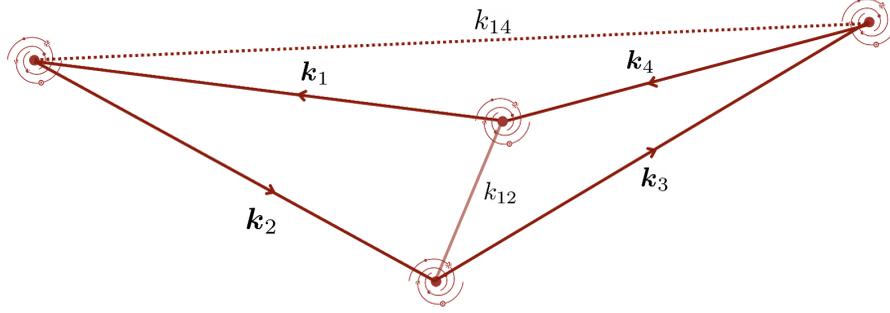


Figure 6.1: Schematic representation of the galaxy trispectrum, where each galaxy is in one of the vertices of the tetrahedron.

## 6.2 Estimators

### 6.2.1 Trispectrum estimator

Let the trispectrum be represented by a tetrahedron defined by four wavevectors  $\mathbf{k}_1$ ,  $\mathbf{k}_2$ ,  $\mathbf{k}_3$ ,  $\mathbf{k}_4$  (see Figure 6.1). The diagonals connecting different pairs of points in this configuration are denoted as follows: the diagonal connecting the beginning of  $\mathbf{k}_1$  to the end of  $\mathbf{k}_2$  is  $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$ , and the diagonal connecting the beginning of  $\mathbf{k}_1$  to the end of  $\mathbf{k}_4$  is  $\mathbf{k}_{14} = \mathbf{k}_1 + \mathbf{k}_4$ . Since this forms a closed tetrahedral structure, the wavevectors satisfy the closure condition,

$$\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 = 0. \quad (6.1)$$

In general, the trispectrum of the real galaxy density field  $\delta_g$  can be written as

$$\langle \delta_g(\mathbf{k}_1) \delta_g(\mathbf{k}_2) \delta_g(\mathbf{k}_3) \delta_g(\mathbf{k}_4) \rangle \equiv (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T(k_1, k_2, k_3, k_4, k_{12}, k_{14}). \quad (6.2)$$

In practice, in order to have a separable estimator, we can define the trispectrum estimator used in [56, 83, 108], which averages over one of the diagonals of the tetrahedron,  $k_{14}$ . The idea is realizing that we want to close the two triangles formed by  $\{\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_{12}\}$  and  $\{\mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_{12}\}$ ,

$$\hat{T}(k_1, k_2, k_3, k_4, k_{12}) \propto \int_{\mathbf{p}_i \in k_i, \mathbf{q} \in k_{12}} \left[ (2\pi)^3 \delta_D(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{q}) (2\pi)^3 \delta_D(\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{q}) \delta_g(\mathbf{p}_1) \delta_g(\mathbf{p}_2) \delta_g(\mathbf{p}_3) \delta_g(\mathbf{p}_4) \right] \quad (6.3)$$

where we introduced the notation

$$\int_{\mathbf{p} \in k_i} \equiv \int_{\|\mathbf{p} - k_i\| < \Delta k/2} \frac{d\mathbf{p}}{(2\pi)^3}, \quad (6.4)$$

which indicates that all  $\mathbf{p}$  modes within a  $k$ -shell of width  $\Delta k$  centered on  $k_i$  are integrated, for  $i = \{1, 2, 3, 4\}$ . We can use FFTs to evaluate Eq. 6.3 if we replace the Dirac deltas by

integrals. To see this, we can define

$$I_{k_i}(\mathbf{x}) = \int_{\mathbf{p} \in k_i} \delta_g(\mathbf{p}) e^{i\mathbf{x} \cdot \mathbf{p}}, \quad J_{k_i}(\mathbf{x}) = \int_{\mathbf{p} \in k_i} e^{i\mathbf{x} \cdot \mathbf{p}}, \quad (6.5)$$

which represent the inverse Fourier transform over a  $k$ -shell of the galaxy overdensity field and a unit field, respectively, and

$$\mathcal{I}_{k_i, k_j}(\mathbf{q}) = \int d\mathbf{x} e^{-i\mathbf{q} \cdot \mathbf{x}} I_{k_i}(\mathbf{x}) I_{k_j}(\mathbf{x}), \quad \mathcal{J}_{k_i, k_j}(\mathbf{q}) = \int d\mathbf{x} e^{-i\mathbf{q} \cdot \mathbf{x}} J_{k_i}(\mathbf{x}) J_{k_j}(\mathbf{x}). \quad (6.6)$$

Our estimator then takes the form

$$\hat{T}(k_1, k_2, k_3, k_4, k_{12}) = \int_{\mathbf{q} \in k_{12}} \mathcal{I}_{k_1, k_2}(\mathbf{q}) \mathcal{I}_{k_3, k_4}(-\mathbf{q}) \times \left[ \int_{\mathbf{q} \in k_{12}} \mathcal{J}_{k_1, k_2}(\mathbf{q}) \mathcal{J}_{k_3, k_4}(-\mathbf{q}) \right]^{-1}. \quad (6.7)$$

For efficiency, we first compute the  $N_{\text{bin}}$   $I_{k_i}$  grids and then save all possible two-by-two combinations of  $\mathcal{I}_{k_i, k_j}$  without permutations, which gives us  $N_{\text{bin}}(N_{\text{bin}} + 1)/2$  grids, and then we sum over each of the given  $k_{12}$  bins. The normalization factor proportional to  $\mathcal{J}_{k_i, k_j}$  has to be calculated only once for a given setting, and it gives us the total number of tetrahedra in a given bin. In order to select only the non-trivial components of the trispectrum and have a closed tetrahedron, we impose the conditions

$$k_1 \leq k_2, \quad k_1 \leq k_3, \quad k_3 \leq k_4, \quad |k_1 - k_2| \leq k_{12} \leq k_1 + k_2, \quad |k_3 - k_4| \leq k_{12} \leq k_3 + k_4. \quad (6.8)$$

Note that the trispectrum has *disconnected* contributions, when both pairs of external legs are the same (e.g.,  $k_1 = k_2$  and  $k_3 = k_4$ ), which can be removed since they have the same information as the power-spectrum [83, 87]. Since we use the trispectrum for SBI analysis, where we are free to choose the summary statistics, we do not subtract the disconnected parts here.

### 6.2.2 Integrated trispectrum

A trispectrum estimator which integrates in both diagonals  $k_{12}$  and  $k_{14}$  is discussed in [86–88], known as the *integrated trispectrum*. We can construct it as

$$\hat{\mathcal{T}}(k_1, k_2, k_3, k_4) = \int d\mathbf{x} I_{k_1}(\mathbf{x}) I_{k_2}(\mathbf{x}) I_{k_3}(\mathbf{x}) I_{k_4}(\mathbf{x}) \times \left[ \int d\mathbf{y} J_{k_1}(\mathbf{y}) J_{k_2}(\mathbf{y}) J_{k_3}(\mathbf{y}) J_{k_4}(\mathbf{y}) \right]^{-1}, \quad (6.9)$$

which after discretization takes the form

$$\hat{\mathcal{T}}(k_1, k_2, k_3, k_4) = \left( \frac{L^9}{N^{12}} \right) \frac{\sum_{i=1}^{N^3} I_{k_1}^D(x_i) I_{k_2}^D(x_i) I_{k_3}^D(x_i) I_{k_4}^D(x_i)}{\sum_{j=1}^{N^3} J_{k_1}^D(x_j) J_{k_2}^D(x_j) J_{k_3}^D(x_j) J_{k_4}^D(x_j)}. \quad (6.10)$$

Since the the integrated trispectrum averages over an extra diagonal, we will use the trispectrum estimator from the previous section in this thesis, unless otherwise stated.

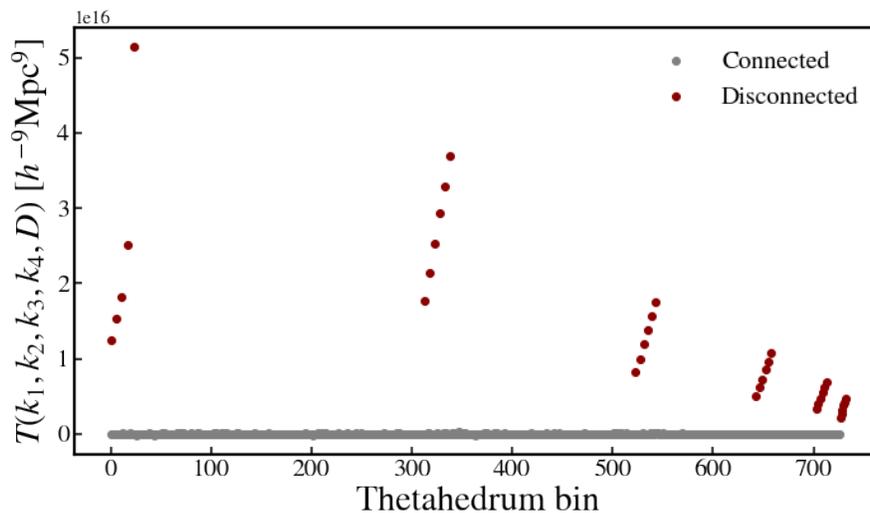


Figure 6.2: Total trispectrum estimator measured in a Gaussian random field.

### 6.3 Validation of the trispectrum estimator

In order to validate our estimator, we use a linear forward model instead of LPT, which is given directly from Eq. (2.107) for  $\alpha = 1$ , where we construct the final density field as

$$\delta_g(\mathbf{k}, z) = b_1 \delta^{(1)}(\mathbf{k}, z) + b_2 \left( [\delta^{(1)}]^2(\mathbf{k}, z) - \langle [\delta^{(1)}]^2 \rangle(z) \right) + b_3 [\delta^{(1)}]^3(\mathbf{k}, z). \quad (6.11)$$

The first test consists of analyzing whether only the disconnected contributions of the trispectrum are non-zero when measured in a Gaussian random field, which can be obtained by setting  $b_1 = 1$  and  $b_2 = b_3 = 0$  in Eq. (6.11). Following [83], we use a box size to  $L = 1000h\text{Mpc}^{-1}$  and scale range  $4k_f < k_1, k_2, k_3, k_4, k_{12} < 10k_f$  with binning width  $\Delta k = k_f$ . We show in Figure 6.2 the measurement averaged over 100 realizations.

Next, we evaluate the configurations dependence of our estimator. If we consider a linear forward model with nonzero  $b_1, b_2$  and  $b_3$ , then at tree level the galaxy trispectrum reads

$$T_g^{\text{LO}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = b_1^2 b_2^2 T_{1122}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) + b_1^3 b_3 T_{1113}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \quad (6.12)$$

where

$$T_{1122}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = 4P(k_1)P(k_2)[P(|\mathbf{k}_1 + \mathbf{k}_3|) + P(|\mathbf{k}_1 + \mathbf{k}_4|)] + 5 \text{ perm.} \quad (6.13)$$

$$T_{1113}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = 6P(k_1)P(k_2)P(k_3) + 3 \text{ perm.} \quad (6.14)$$

Note that, while the theoretical prediction is given as a function of the vector of wavenumbers, our estimator defined in Eq. 6.7 is only a function of the *absolute values*  $k_1, k_2, k_3, k_4, k_{12}$  and integrate the theoretical prediction over  $k_{14}$ . In fact, the trispectrum has only six independent degrees of freedom due to rotational invariance, so that it is completely described by six variables such as the magnitudes  $k_1, k_2, k_3, k_4, k_{12}, k_{14}$  (see, e.g., Fig. 6.3).

### 6.3.1 $T_{1122}$

Let us start by investigating how to express the arguments of  $T_{1122}$  as a function of these absolute values, while  $T_{1113}$  is already expressed in such a way. Writing  $T_{1122}$  explicitly over all permutations, we obtain

$$\begin{aligned}
T_{1122}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = & 4P(k_1)P(k_2) \left[ P(|\mathbf{k}_1 + \mathbf{k}_3|) + P(|\mathbf{k}_1 + \mathbf{k}_4|) \right] \\
& + 4P(k_1)P(k_3) \left[ P(|\mathbf{k}_1 + \mathbf{k}_2|) + P(|\mathbf{k}_1 + \mathbf{k}_4|) \right] \\
& + 4P(k_1)P(k_4) \left[ P(|\mathbf{k}_1 + \mathbf{k}_2|) + P(|\mathbf{k}_1 + \mathbf{k}_3|) \right] \\
& + 4P(k_2)P(k_3) \left[ P(|\mathbf{k}_2 + \mathbf{k}_1|) + P(|\mathbf{k}_2 + \mathbf{k}_4|) \right] \\
& + 4P(k_2)P(k_4) \left[ P(|\mathbf{k}_2 + \mathbf{k}_1|) + P(|\mathbf{k}_2 + \mathbf{k}_3|) \right] \\
& + 4P(k_3)P(k_4) \left[ P(|\mathbf{k}_3 + \mathbf{k}_1|) + P(|\mathbf{k}_3 + \mathbf{k}_2|) \right].
\end{aligned} \tag{6.15}$$

Defining the diagonal  $k_{13} = |\mathbf{k}_1 + \mathbf{k}_3| = |\mathbf{k}_2 + \mathbf{k}_4|$  and noting that  $k_{14} = |\mathbf{k}_1 + \mathbf{k}_4| = |\mathbf{k}_2 + \mathbf{k}_3|$ , we rewrite the expression as

$$\begin{aligned}
T_{1122}(k_1, k_2, k_3, k_4, k_{12}, k_{14}) = & 4P(k_1)P(k_2) [P(k_{13}) + P(k_{14})] \\
& + 4P(k_1)P(k_3) [P(k_{12}) + P(k_{14})] \\
& + 4P(k_1)P(k_4) [P(k_{12}) + P(k_{13})] \\
& + 4P(k_2)P(k_3) [P(k_{12}) + P(k_{13})] \\
& + 4P(k_2)P(k_4) [P(k_{12}) + P(k_{14})] \\
& + 4P(k_3)P(k_4) [P(k_{13}) + P(k_{14})].
\end{aligned} \tag{6.16}$$

We now have the expression as function of given  $k_1, k_2, k_3, k_4, k_{12}, k_{14}$ . Note that, while  $k_{13}^2 = k_1^2 + k_3^2 + 2\mathbf{k}_1 \cdot \mathbf{k}_3$  is in principle a function of the dot product of two vectors, we can express it as a function of the given absolute values only, since

$$\begin{aligned}
k_{13}^2 &= k_1^2 + k_2^2 + k_3^2 + k_4^2 - (k_{12}^2 + k_{14}^2) \\
&= k_1^2 + k_2^2 + k_3^2 + k_4^2 - (k_1^2 + k_2^2 + 2\mathbf{k}_1 \cdot \mathbf{k}_2 + k_1^2 + k_4^2 + 2\mathbf{k}_1 \cdot \mathbf{k}_4) \\
&= k_1^2 + k_3^2 - (k_1^2 + 2\mathbf{k}_1 \cdot \mathbf{k}_2 + k_1^2 + 2\mathbf{k}_1 \cdot \mathbf{k}_4) \\
&= k_1^2 + k_3^2 - 2\mathbf{k}_1 \cdot (\mathbf{k}_2 + \mathbf{k}_4) \\
&= k_1^2 + k_3^2 + 2\mathbf{k}_1 \cdot \mathbf{k}_3.
\end{aligned} \tag{6.17}$$

Finally, we integrate the theoretical expression in  $k_{14}$  to have a prediction for our estimator,

$$T_{1122}(k_1, k_2, k_3, k_4, k_{12}) = \frac{1}{(k_{14}^{\max} - k_{14}^{\min})} \int_{k_{14}^{\min}}^{k_{14}^{\max}} dk_{14} T_{1122}(k_1, k_2, k_3, k_4, k_{12}, k_{14}). \tag{6.18}$$

For a given configuration of  $k_1, k_2, k_3, k_4, k_{12}$ , the maximum and minimum values of the diagonal  $k_{14}$  occur when the tetrahedron is planar (see Fig. 6.3), depending on the

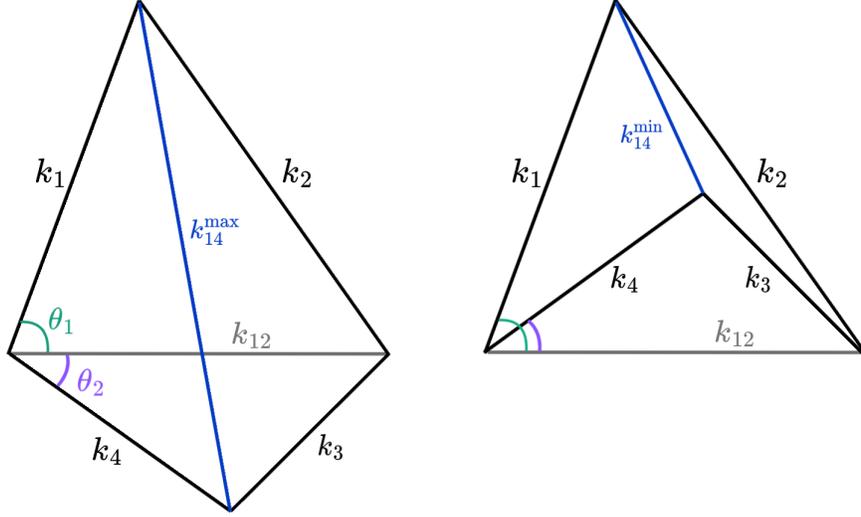


Figure 6.3: Maximum and minimum values that the diagonal  $k_{14}$  can take.

rotation around the diagonal  $k_{12}$ . Using the cosine rule,

$$\theta_1 = \arccos\left(\frac{k_1^2 + k_{12}^2 - k_2^2}{2k_1k_{12}}\right), \quad (6.19)$$

$$\theta_2 = \arccos\left(\frac{k_4^2 + k_{12}^2 - k_3^2}{2k_4k_{12}}\right), \quad (6.20)$$

$$k_{14}^{\max} = \sqrt{k_1^2 + k_4^2 - 2k_1k_4 \cos(\theta_1 + \theta_2)}. \quad (6.21)$$

The minimum diagonal length of  $k_{14}$  occurs when the quadrilateral is flipped around  $k_{12}$ ,

$$k_{14}^{\min} = \sqrt{k_1^2 + k_4^2 - 2k_1k_4 \cos(\theta_1 - \theta_2)}. \quad (6.22)$$

Note that, although Fig. 6.3 depicts a convex quadrilateral, these expressions are valid in general.

To test the performance of our estimator and compare it to the prediction of  $T_{1122}$ , we set  $b_1 = 1$ ,  $b_2 = 0.5$  and  $b_3 = 0$  in Eq. 6.11, box size to  $L = 1000h^{-1}\text{Mpc}$  and we use 8 linear bins in  $k$  from  $k_{\min} = 0.01h\text{Mpc}^{-1}$  to  $k_{\max} = 0.1h\text{Mpc}^{-1}$ . In total, we have 930 tetrahedra configurations, and 882 after removing disconnected bins. We show in Fig. 6.4 the average and standard deviation over 2000 initial conditions realizations against the prediction  $b_1^2 b_2^2 T_{1122}$ , where  $T_{1122}$  is given by Eq. (6.18). We find excellent agreement of our estimator implemented in `LEFTfield` and the theoretical prediction.

### 6.3.2 $T_{1113}$

For  $T_{1113}$ , we set box size to  $L = 1000h^{-1}\text{Mpc}$  and use 5 linear bins in  $k$  from  $k_{\min} = k_f$  to  $k_{\max} = 0.1h\text{Mpc}^{-1}$ . In total, we have 134 configurations, and 128 after removing

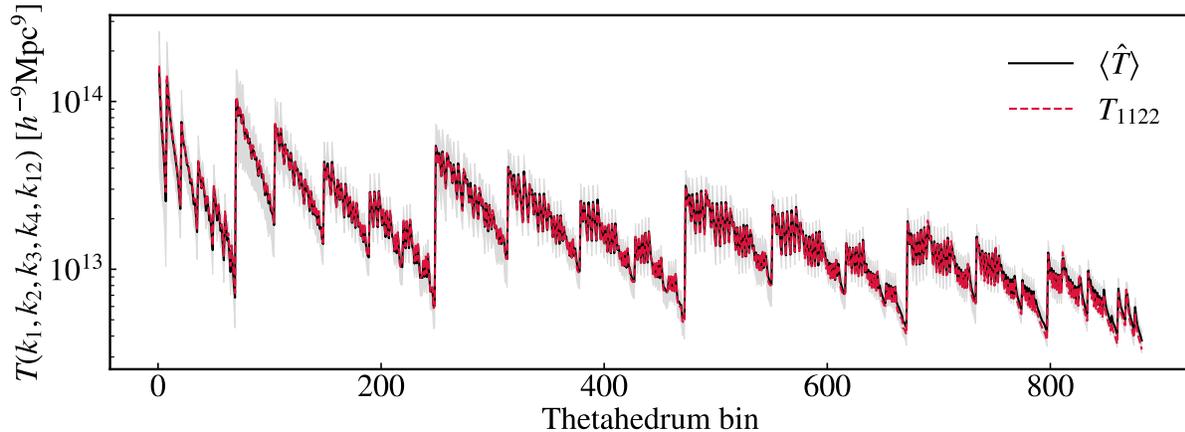


Figure 6.4: Trispectrum estimator averaged over 2000 realizations (in black) and theoretical prediction  $b_1^2 b_2^2 T_{1122}$  (in red) as a function of thetadedrum bin. The standard deviation over the realizations is indicated by the shaded grey area.

disconnected bins. We set  $b_1 = 1$ ,  $b_2 = 0$  and  $b_3 = 0.1$  in Eq. (6.11). Since  $T_{1113}$  depends only on the magnitudes of the external momenta, we validate it as follows: for given  $k_1, k_2, k_3, k_4$ , we calculate the histogram of  $(\langle \hat{T} \rangle - b_1^3 b_3 T_{1113}) / \sigma(\hat{T})$  for all diagonals  $k_{12}$  over 100 realizations, as the distributions should scatter around a unity Gaussian if the estimator is correct. Note that for a given configuration the number of allowed diagonals may vary due to the triangle inequality conditions (see Eq. 6.8). We show the results for randomly selected configurations in Fig. 6.5. We do not see systematic bias in any of the configurations, which further confirms the validity of our estimator.

## 6.4 Trispectrum covariance

Here, we analyze the trispectrum covariance for illustration purposes. We emphasize that SBI does not require an explicit covariance, where the moments of the distributions are learned from the trispectrum measurements in LEFTfield simulations. An analytical approximation can be obtained for the ‘‘Gaussian’’ trispectrum covariance (i.e., ignoring cross-correlations in between different bins) [87],

$$\text{Cov}[T(k_1, k_2, k_3, k_4)] = \frac{(2\pi)^9 R_{1234}}{V_s V_{1234}^q} P(k_1) P(k_2) P(k_3) P(k_4), \quad (6.23)$$

where  $V_s$  is the effective survey volume,  $V_{1234}^q = 16\pi^3 k_1 k_2 k_3 k_4 \Delta k_1 \Delta k_2 \Delta k_3 \Delta k_4$  is the bin integration volume, and  $R_{1234}$  is the symmetric factor that counts the number of

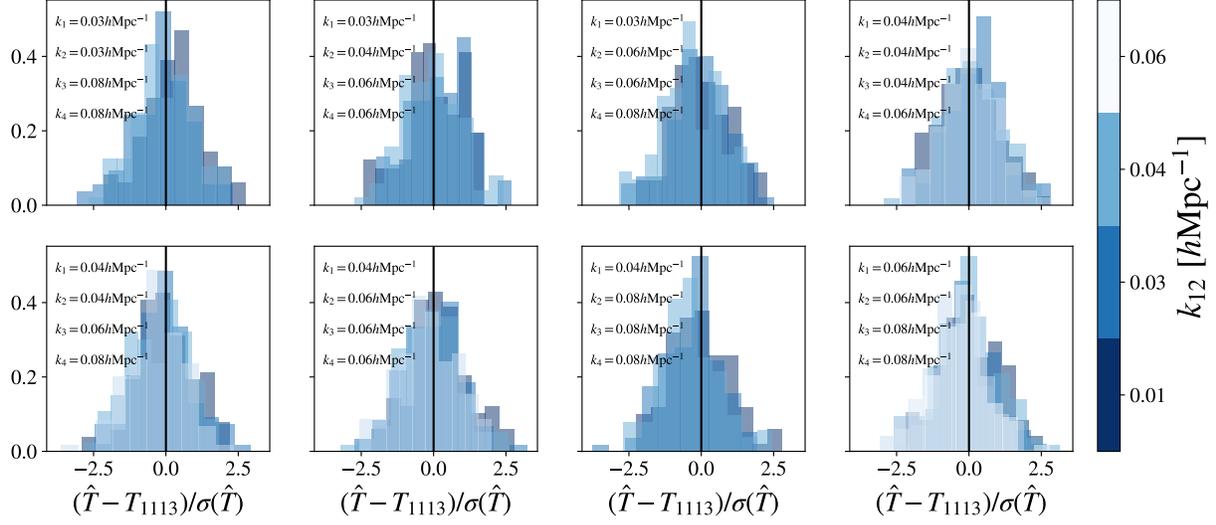


Figure 6.5: Normalized histogram of  $(\langle \hat{T} \rangle - b_1^3 b_3 T_{1113}) / \sigma(\hat{T})$ , where  $\hat{T}$  are the estimated trispectra for  $b_1 = 1$  and  $b_3 = 0.1$  and  $T_{1113}$  is the prediction.

possible permutations of equal sides between the two identical quadrilaterals sets,

$$R_{1234} = \begin{cases} 1 & \text{for } (k_a, k_b, k_c, k_d) \\ 2 & \text{for } (k_a, k_a, k_c, k_d) \\ 4 & \text{for } (k_a, k_a, k_c, k_c) \\ 6 & \text{for } (k_a, k_a, k_a, k_d) \\ 24 & \text{for } (k_a, k_a, k_a, k_a). \end{cases} \quad (6.24)$$

The reduced covariance matrix is a dimensionless form of the covariance matrix that provides a measure of the correlation between different bins or modes, independent of their variances, defined by normalizing each entry by the product of the mean values of the corresponding bins. For a covariance matrix  $C_{ij}$ , the reduced covariance matrix  $r_{ij}$  is given by

$$r_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}}, \quad (6.25)$$

where this normalization ensures that

$$-1 \leq r_{ij} \leq 1, \quad \text{with } r_{ii} = 1. \quad (6.26)$$

We show in Fig. 6.6 the reduced covariance matrix corresponding to the data vector of the galaxy power-spectrum, bispectrum and trispectrum bins. We use the parameter values from the MAP posterior of SNG halos at  $z = 0.5$  and Uchuu halos at  $z = 1.03$  and only vary the initial conditions in LEFTfield. We observe that the data vector is more strongly correlated in between different data vector bins at  $z = 0.5$  than at  $z = 1.03$ , which is

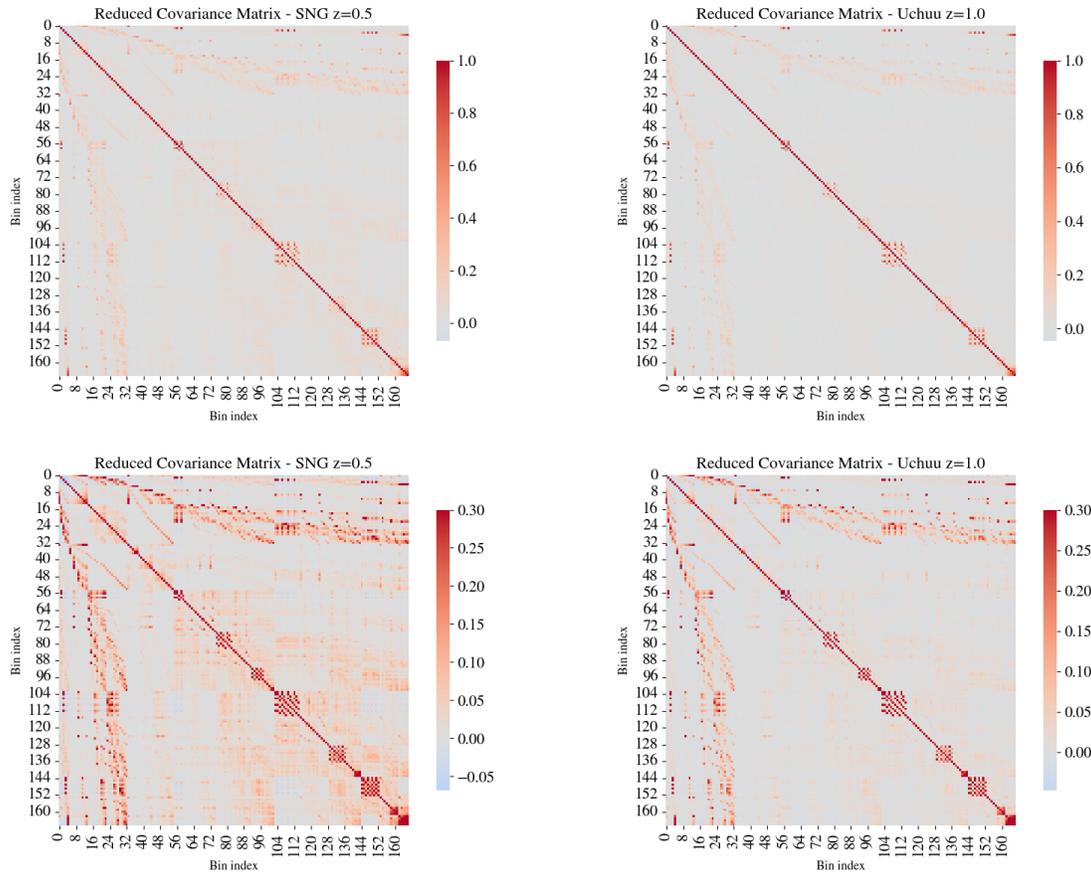


Figure 6.6: Reduced covariance matrix for the galaxy power-spectrum, bispectrum and trispectrum bins. The left columns show the covariance for  $z = 0.5$  and bias parameters fixed to the MAP from SNG halos posterior, while the right column shows the same for the Uchuu halos at  $z = 1.03$ . The colorbar on the upper panels show the value of the reduced covariance matrix for each bin, while the lower panels have a maximum value of 0.3 to contrast the difference between both redshifts. In general, the data vector at  $z = 0.5$  is more strongly correlated among between different bins than at  $z = 1.03$ .

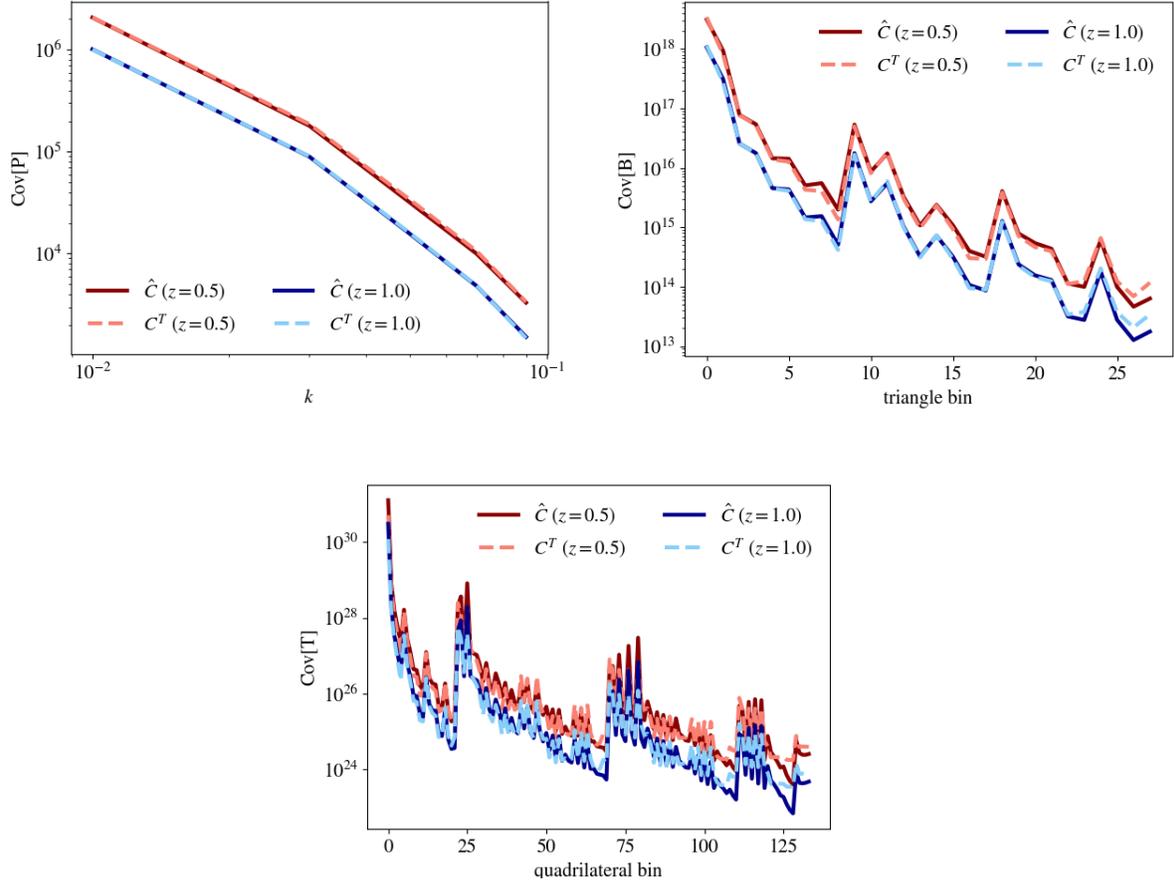


Figure 6.7: Diagonal part of the sample covariance  $\hat{C}$  of the galaxy power-spectrum (*upper left*), bispectrum (*upper right*) and trispectrum (*bottom*) measured from LEFTfield, compared to the corresponding theoretical covariance  $C^T$ . Red lines correspond to redshift  $z = 0.5$  while blue lines indicate  $z = 1.03$ .

expected since gravitational evolution introduces coupling of different modes. This is in agreement with [87], who also found smaller covariance for higher redshift.

In Fig. 6.7, we compare the theoretical covariance of the power-spectrum (Eq. 4.15), bispectrum (Eq. 4.16) and trispectrum (Eq. 6.23), with the estimated covariance from LEFTfield (same procedure as described above). The theoretical predictions matches well with the sample covariance at both redshifts, although we see a bigger deviation at small scales when considering the galaxy bispectrum and trispectrum.

## 6.5 Results

### 6.5.1 LEFTfield settings

The baseline used in this paper is the same as [157] (detailed Chapter 5): 2LPT, third order Eulerian bias and Gaussian stochasticity, including leading order higher-derivative terms, i.e.,

$$\mathbf{O} = \left\{ \delta, \nabla^2 \delta, \delta^2, K^2, \delta^3, K^3, \delta K^2, O_{\text{td}}, \varepsilon, \nabla^2 \varepsilon \right\}. \quad (6.27)$$

We consider a box size of  $L = 2000 h^{-1} \text{Mpc}$  to analyze the same dark-matter halos as well, namely SNG at  $z = 0.5$  and Uchuu at  $z = 1.03$ , focusing on the lower-cutoff case of  $\Lambda = 0.12 h \text{Mpc}^{-1}$  ( $k_{\text{max}} = 0.1 h \text{Mpc}^{-1}$ ). We use the same prior that was used in the power-spectrum and bispectrum analysis considered in [157], the *baseline prior*,

$$\begin{aligned} \alpha &\sim \mathcal{N}(1, 0.3^2), & b_1 \alpha &\sim \mathcal{N}(1.6, 0.5^2), & b_{\delta^2} \alpha^2 &\sim \mathcal{N}(0, 0.5^2), & b_{K^2} \alpha^2 &\sim \mathcal{N}(0, 0.5^2), \\ b_{\delta^3} \alpha^3 &\sim \mathcal{N}(0, 1^2), & b_{\delta K^2} \alpha^3 &\sim \mathcal{N}(0, 1^2), & b_{K^3} \alpha^3 &\sim \mathcal{N}(0, 1^2), & b_{\text{td}} \alpha^3 &\sim \mathcal{N}(0, 1^2), \\ b_{\nabla^2 \delta} \alpha R_\star^{-2} &\sim \mathcal{N}(0, 1^2), & b_{\nabla^2 \varepsilon} R_\star^{-2} &\sim \mathcal{N}(1, 1^2), & b_\varepsilon &\sim \mathcal{U}(0, 0.5). \end{aligned} \quad (6.28)$$

We also consider a *narrower prior* (on the majority of the deterministic bias parameters),

$$\begin{aligned} \alpha &\sim \mathcal{N}(1, 0.3^2), & b_1 \alpha &\sim \mathcal{N}(1.6, 0.25^2), & b_{\delta^2} \alpha^2 &\sim \mathcal{N}(0, 0.25^2), & b_{K^2} \alpha^2 &\sim \mathcal{N}(0, 0.25^2), \\ b_{\delta^3} \alpha^3 &\sim \mathcal{N}(0, 0.25^2), & b_{\delta K^2} \alpha^3 &\sim \mathcal{N}(0, 0.25^2), & b_{K^3} \alpha^3 &\sim \mathcal{N}(0, 0.25^2), & b_{\text{td}} \alpha^3 &\sim \mathcal{N}(0, 0.25^2), \\ b_{\nabla^2 \delta} \alpha R_\star^{-2} &\sim \mathcal{N}(0, 1^2), & b_{\nabla^2 \varepsilon} R_\star^{-2} &\sim \mathcal{N}(1, 1^2), & b_\varepsilon &\sim \mathcal{U}(0, 0.5). \end{aligned} \quad (6.29)$$

### 6.5.2 Trispectrum information content

Here, we report the results of a joint analysis of the galaxy power-spectrum, bispectrum and trispectrum using SBI and LEFTfield, while the analyzed data are the same dark-matter halos that were analyzed in Chapter 5, as mentioned in the previous section. Due to the increased computational demand of the galaxy trispectrum, we show results using 5 linear bins instead of 15 considered in Chapter 5 for the power-spectrum and bispectrum only, keeping all other setting fixed and focusing on the case of  $k_{\text{max}} = 0.1 h \text{Mpc}^{-1}$ . This leads to a data vector of dimensionality  $D = 33$  for the power-spectrum and bispectrum (“PB”) and  $D = 167$  for the power-spectrum, bispectrum and trispectrum (“PBT”). Our main goal in this section is to provide an overview of the current status of this project, while in the final manuscript we plan to provide an analysis with thinner  $k$ -bins and also for  $k_{\text{max}} = 0.12 h \text{Mpc}^{-1}$ .

In Fig. 6.8 (for Uchuu halos) and in Fig. 6.9 (for SNG halos), we can see the impact of adding the galaxy trispectrum for constraining  $\alpha$  together with bias and stochastic parameters using the baseline prior in Eq. (6.28). The improvement factor for adding the trispectrum, in this case, is of 1.2 for the Uchuu halos and unity for SNG, (i.e., the ratio of standard deviation on  $\alpha$  obtained by the posterior samples,  $\sigma_{\text{PB}}(\alpha)/\sigma_{\text{PBT}}(\alpha)$ ). Note that

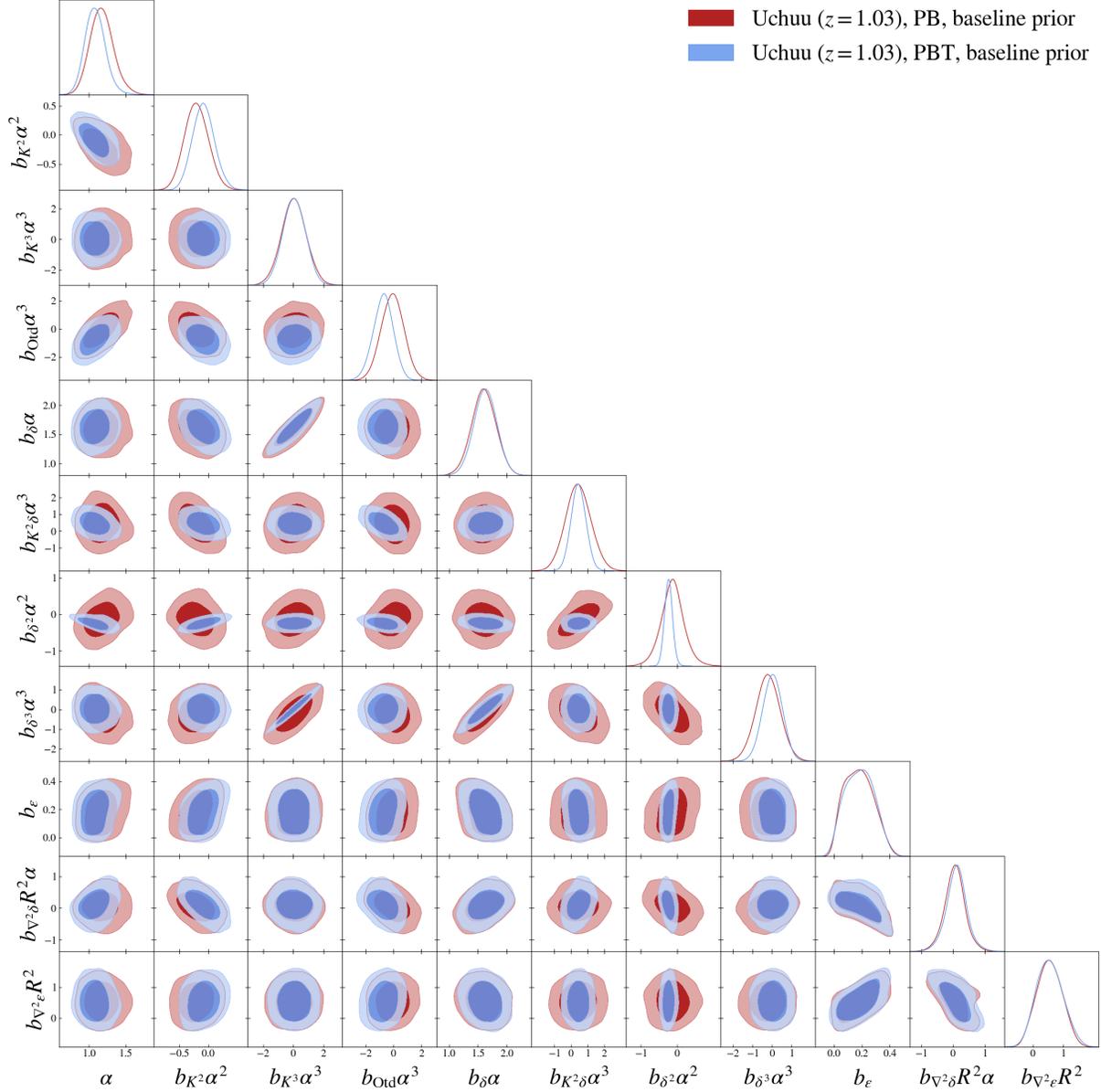


Figure 6.8: Posterior distribution of all model parameters inferred using Uchuu halos at redshift  $z = 1.03$  as the “observed” data vector. Red and blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For both cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 8 \times 10^5$ , scale cut of  $k_{\text{max}} = 0.1 h\text{Mpc}^{-1}$  and the baseline prior given in Eq. (6.28).

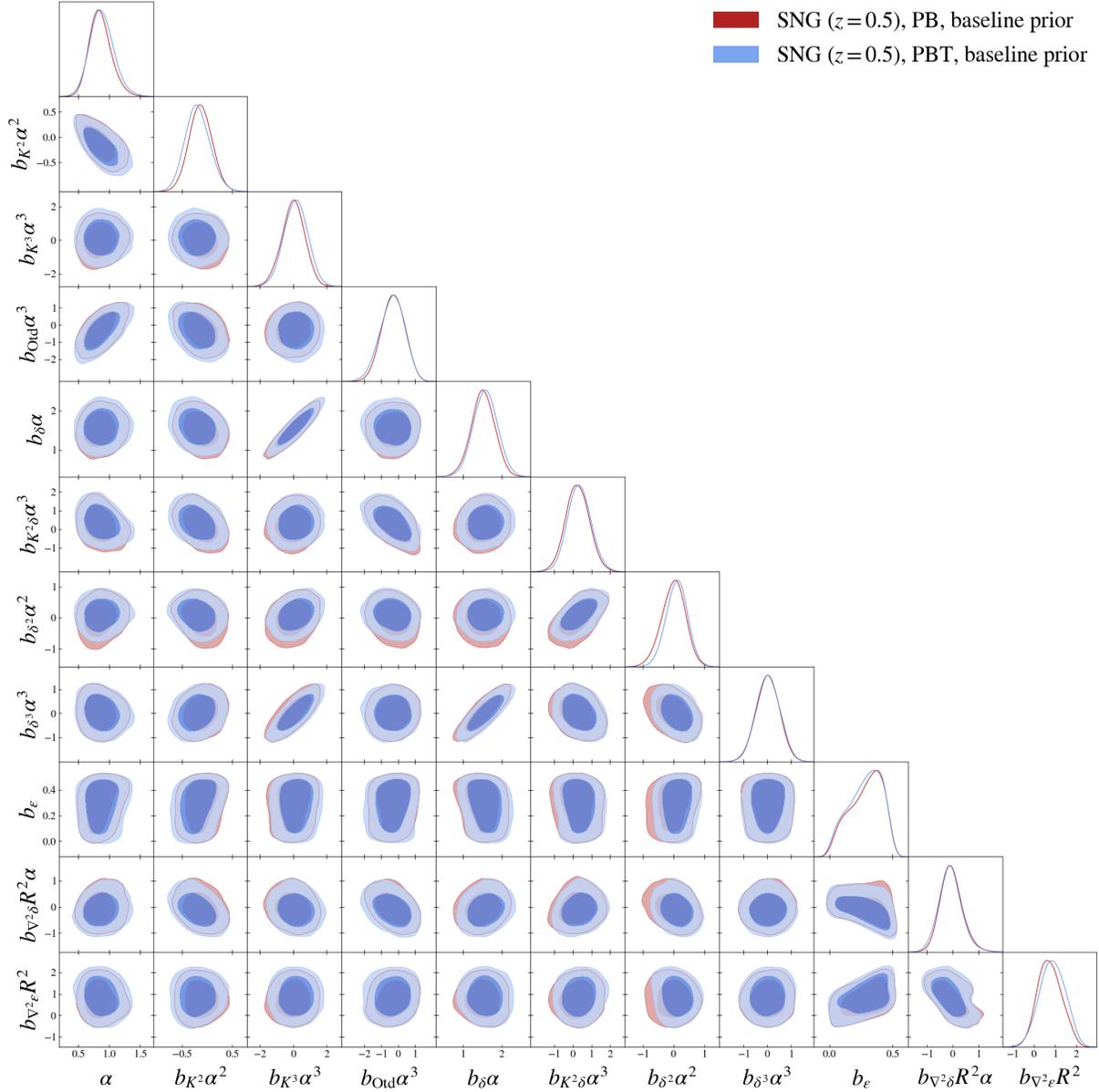


Figure 6.9: Posterior distribution of all model parameters inferred using SNG halos at redshift  $z = 0.5$  as the “observed” data vector. Red and blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For both cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 8 \times 10^5$ , scale cut of  $k_{\text{max}} = 0.1 h \text{Mpc}^{-1}$  and the baseline prior given in Eq. (6.28).

since the cosmologies and redshifts are different, we need to train separate models for the NPE in each case to then evaluate it at the corresponding halo data vector.

The fact that the trispectrum does not add information for the SNG halo sample at  $z = 0.5$ , but does for Uchuu at  $z = 1.03$ , is intriguing. Many tests were made to the baseline in order to test what was the reason for this difference. We tried to include more simulations (up to  $10^6$ ); thinner bins (10 linear bins in  $k$ ); include more wavenumbers ( $k_{\max} = 0.12h\text{Mpc}^{-1}$ ); hyperparameter optimization for NPE (varying batch size, flow complexity, dropout probability, learning rate); different data compression schemes (MSE and VMIM for the loss function, 1D CNN and FCN for the architecture); add the integrated trispectrum instead; sequential SBI (different settings and number of simulations per round); alternative density estimators ((S)NLE, NPSE, flow matching); include higher-order stochastic terms; exclude disconnected bins from the trispectrum; evaluate NPE at different “observed” data vectors (mock data generated from `LEFTfield` with no model mismatch, mock data with lower noise level, Uchuu halo sample at  $z = 0.5$ , mock data with Uchuu  $z = 1.03$  best fit parameters, etc.). All these tests resulted in the same qualitative feature: that the trispectrum seemed to not add information when compared to power-spectrum and bispectrum constraints. We do not show all these tests here to avoid clutter; all of them look like Fig. 6.9.

The only cases in which the trispectrum adds information are when the prior is narrower (see Figs. 6.10 and 6.11, using the prior from Eq. 6.29) and when the parameter space is reduced (specially when excluding third order bias parameters; see Figs. 6.12 and 6.13). We believe that the issue could be related to the fact that high-order bias parameters are too correlated, and that SBI struggles to break this degeneracy. Note, for example, the strong correlations in the  $\{b_1\alpha, b_{s3}\alpha^3, b_{K3}\alpha^3\}$  plane, and that most of the third-order bias parameters are actually prior dominated. Therefore, narrowing the prior and reducing the parameter space seem to help SBI for breaking bias degeneracies, as we discuss in detail below.

**The effect of narrowing the prior.** We show in Table 6.1 the different constraints on  $\alpha$  depending if we use the baseline or narrower priors, corresponding to Fig. 6.10 and Fig. 6.11. Note that we find agreement with the results presented in Chapter 5 for the power-spectrum and bispectrum case (PB), although we emphasize that small differences could come from the fact that we are using wider binning (5 linear bins instead of 15), besides the fact that there is an stochasticity of SBI methods during training. The latter can be easily solved with ensemble networks, as we plan to do once we settle the analysis setup.

Considering the baseline prior, we obtain an improvement factor of 1.2 when including the trispectrum for the Uchuu halos, and no improvement factor for SNG, as previously discussed. However, when using the narrower prior, adding the trispectrum leads to an improvement factor of 1.4 for both halo samples. This is a good indication that the trispectrum can indeed add information even in this large parameter space, although by using the narrower prior we are not able to provide a fair comparison with the field-level

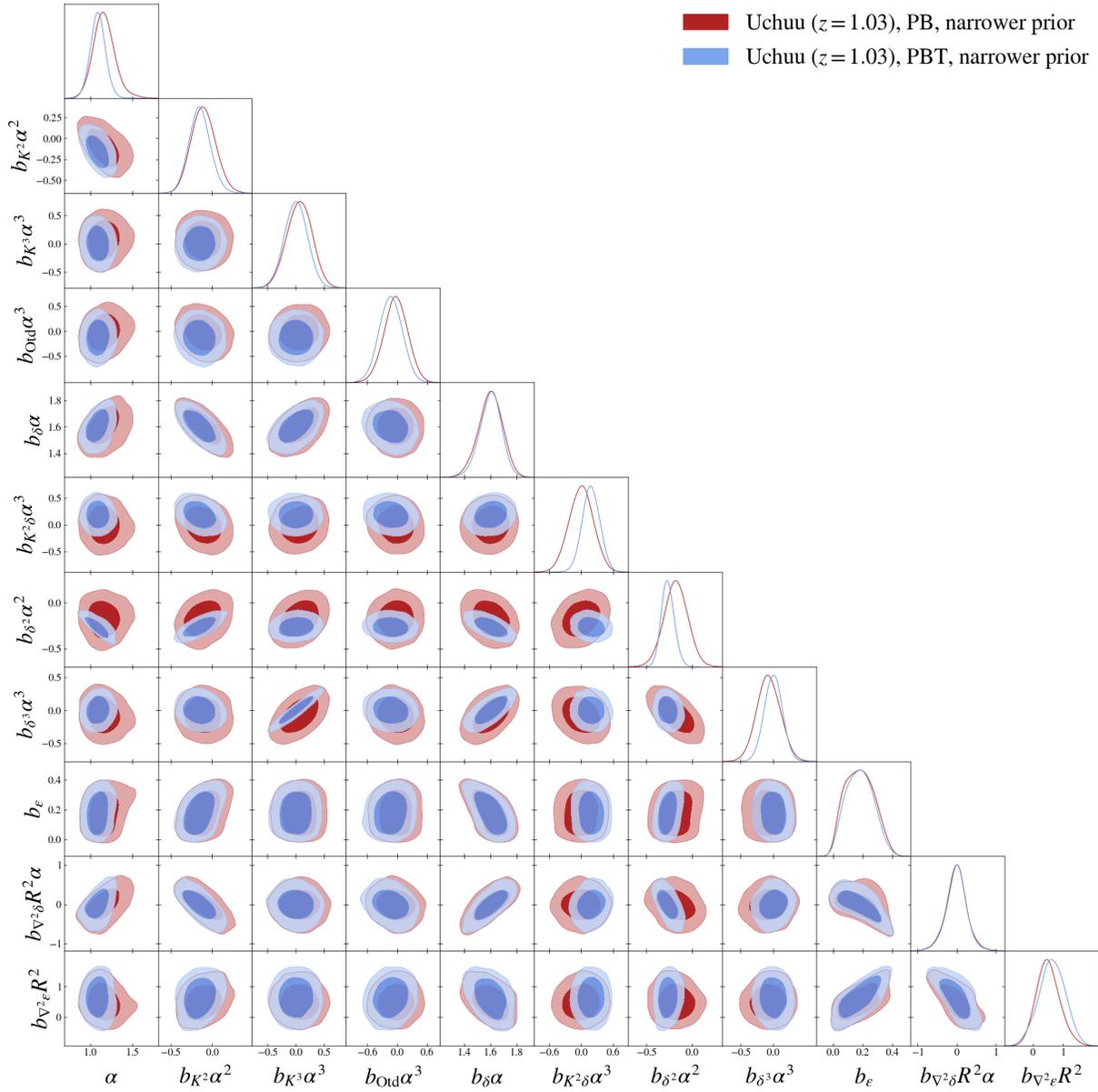


Figure 6.10: Same as Fig. 6.8, but using a simulation budget of  $N_{\text{sim}} = 5 \times 10^5$  and the narrower prior given in Eq. (6.29).

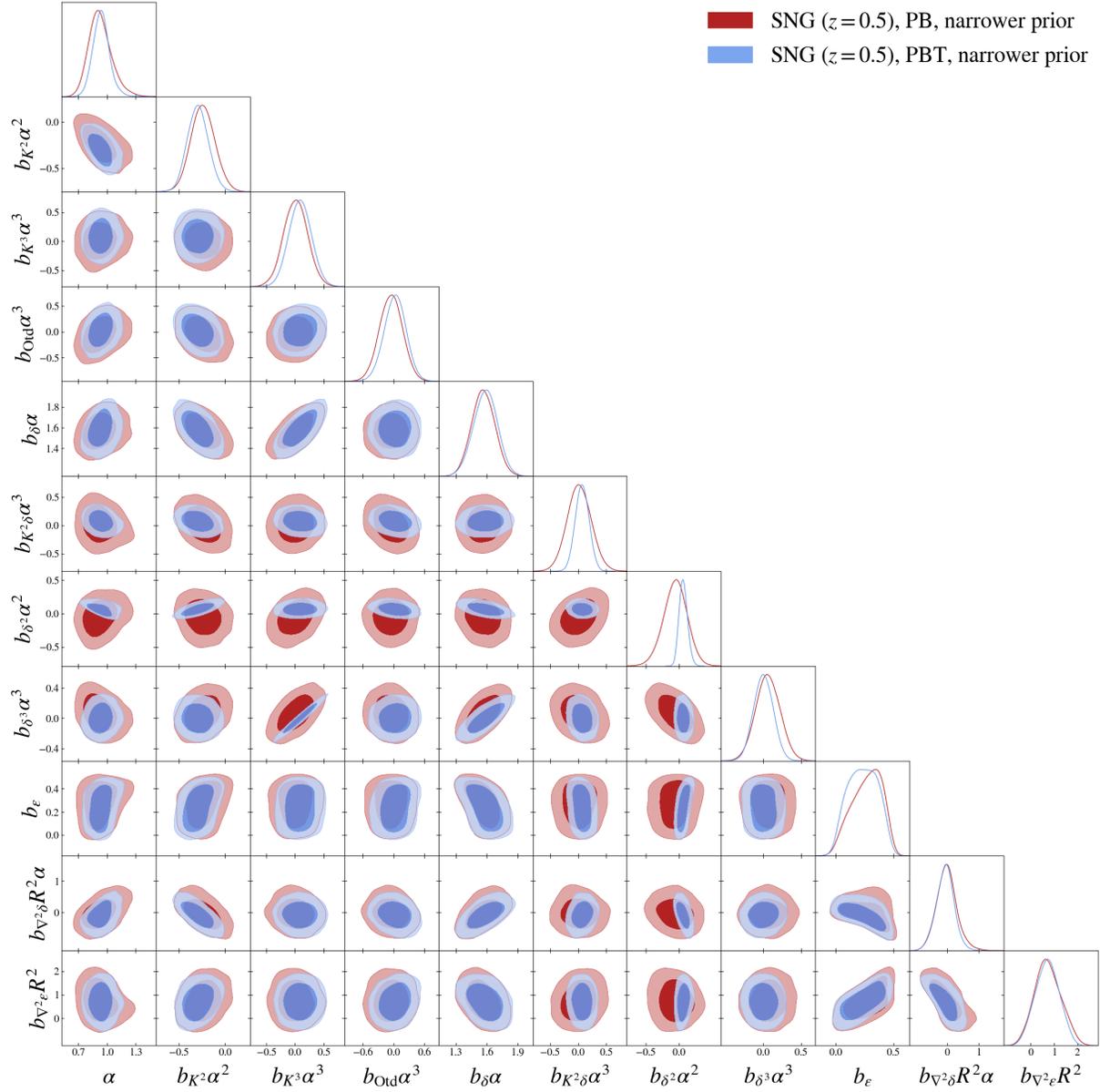


Figure 6.11: Same as Fig. 6.9, but using a simulation budget of  $N_{\text{sim}} = 5 \times 10^5$  and the narrower prior given in Eq. (6.29).

|                      | Mean and 68% CL error on $\alpha$ |                   |                   |                   |
|----------------------|-----------------------------------|-------------------|-------------------|-------------------|
|                      | Baseline prior                    |                   | Narrower prior    |                   |
|                      | PB                                | PBT               | PB                | PBT               |
| SNG ( $z = 0.5$ )    | $0.858 \pm 0.175$                 | $0.880 \pm 0.191$ | $0.929 \pm 0.122$ | $0.938 \pm 0.087$ |
| Uchuu ( $z = 1.03$ ) | $1.175 \pm 0.162$                 | $1.084 \pm 0.139$ | $1.166 \pm 0.131$ | $1.093 \pm 0.091$ |

Table 6.1: Results of inference on the cosmological parameter  $\alpha$  obtained from dark-matter halos, with values for the mean and 68% confidence level (CL) error on  $\alpha$  from by the posterior samples obtained by NPE. We compare results using the baseline and narrower priors, for PB (power-spectrum and bispectrum) and PBT (power-spectrum, bispectrum and trispectrum) data vectors.

inference analysis presented in Chapter 5. Moreover, we emphasize that third-order bias parameters are prior dominated when using the galaxy  $n$ -point functions, which is not the case for field-level inference. It would be interesting however to perform field-level inference on these dark-matter halos using the narrower prior for a fair comparison in case we really do not find any improvement when adding the trispectrum for SNG at  $z = 0.5$  using the baseline prior.

**Reduced parameter space.** Since the parameter space considered for analyzing dark-matter halos with the galaxy trispectrum is considerably large, and we suspect that the lack of information gain at  $z = 0.5$  when using the baseline prior to be connected to the strong degeneracies between bias parameters, we provide here an analysis of the information gain by adding the trispectrum to the power-spectrum and bispectrum when the parameter space is reduced.

First, we consider a linear forward model given by Eq. (6.11) to infer  $\{b_\delta, b_{\delta^2}, b_{\delta^3}\}$ . Note that, since in the linear forward model we do not use LPT, it is not possible to constrain  $\alpha$ . The results are shown in Fig. 6.12 for  $z = 0.5$ , where we analyze mock data generated from `LEFTfield`. The information gain from adding the trispectrum in this case is huge, since the power-spectrum and bispectrum alone are not able to constrain  $b_{\delta^3}$ , which only appears in the trispectrum. Therefore,  $b_{\delta^3}$  is completely prior dominated for the PB case. This is a good confirmation that in a very simple scenario the trispectrum adds considerably information at  $z = 0.5$ .

Next, we use 2LPT to infer  $\{\alpha, b_\delta\alpha, b_{\delta^2}\alpha^2, b_{\delta^3}\alpha^3\}$ . The results are shown in Fig. 6.13 for  $z = 0.5$  and  $z = 1.0$ , where we again analyze mock data generated from `LEFTfield`. The information gain from adding the trispectrum in this case is of order 1.5 for both redshifts, where we have started from a prior much larger than the region of posterior support. This indicates that the trispectrum contains relevant cosmological information and demonstrates the power of breaking bias degeneracies even when the prior is very wide at  $z = 0.5$  when the parameter space is reduced.

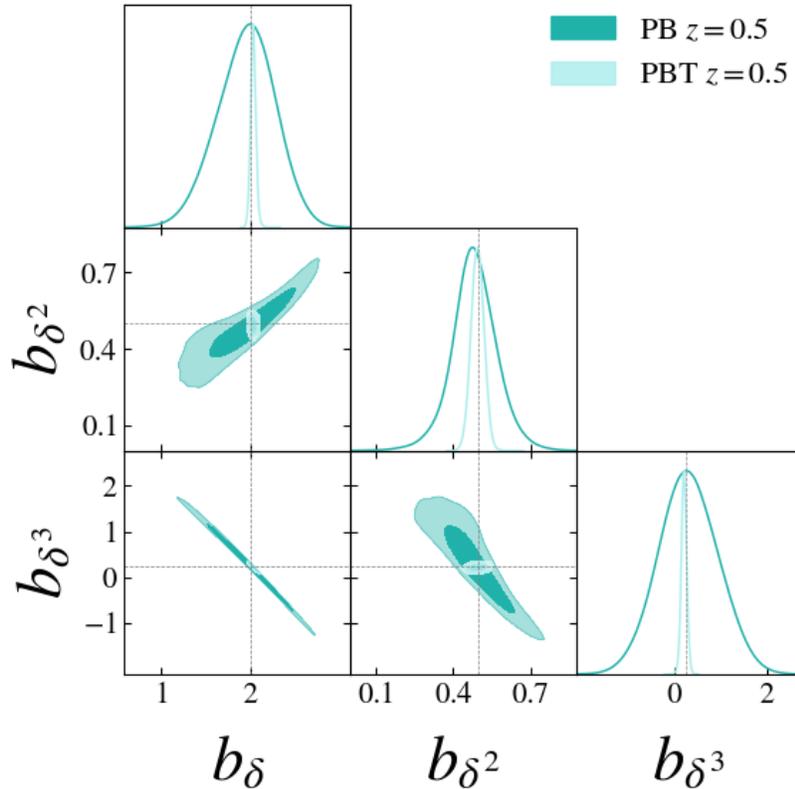


Figure 6.12: Posterior distribution of the model parameters using mock data generated from the same model at redshift  $z = 0.5$  when only sampling  $\{b_\delta, b_{\delta^2}, b_{\delta^3}\}$  and using the trivial forward model. Dark and light blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For both cases, the method SNPE was used using 15 rounds and a simulation budget of  $N_{\text{sim}} = 10^4$  per round, and scale cut of  $k_{\text{max}} = 0.1 h\text{Mpc}^{-1}$ .

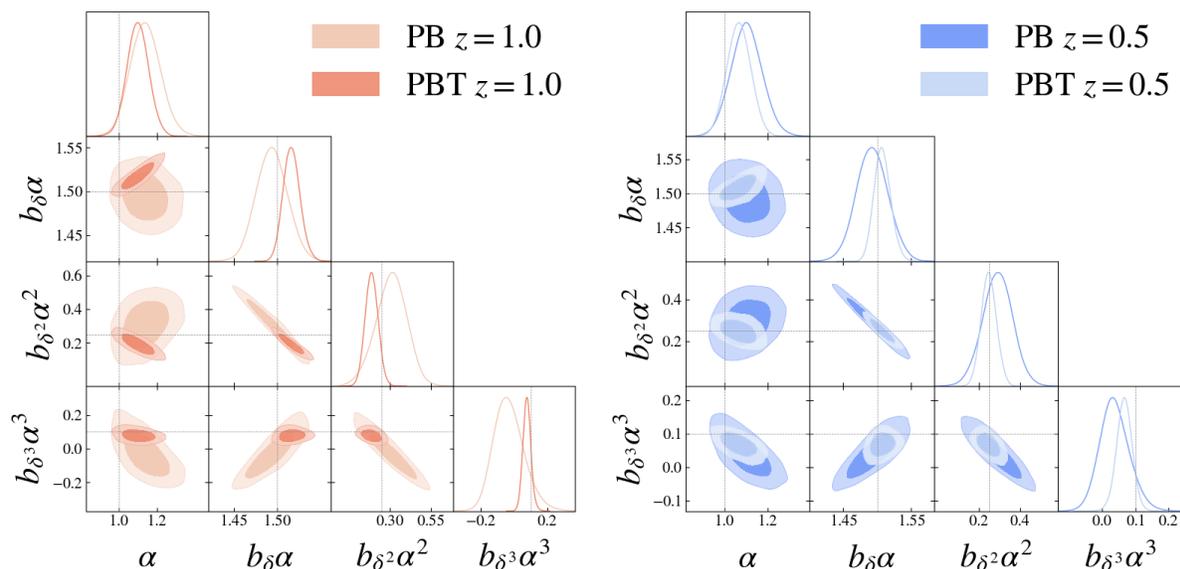


Figure 6.13: Posterior distribution of the model parameters inferred using mock data generated from `LEFTfield` when only sampling  $\{\alpha, b_\delta\alpha, b_{\delta^2}\alpha^2, b_{\delta^3}\alpha^3\}$  and using 2LPT. *Left:* Inference at redshift  $z = 1.0$ . Light and dark pink contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. *Right:* Inference at redshift  $z = 0.5$ . Dark and light blue contours correspond to when we include only power-spectrum and bispectrum (“PB”) or when we combine them with the trispectrum (“PBT”) in the data vector. For all cases, the method SNPE was used using 15 rounds and a simulation budget of  $N_{\text{sim}} = 10^4$  per round, and scale cut of  $k_{\text{max}} = 0.1h\text{Mpc}^{-1}$ .

**Further tests.** We have identified the issue that the trispectrum does not add information for the SNG halo sample at  $z = 0.5$ , but does for Uchuu at  $z = 1.03$ , when using the baseline prior. There is information gain at  $z = 0.5$ , however, when we narrow the prior or reduce the parameter space, what could indicate a failure in the inference procedure when using the baseline prior at  $z = 0.5$ . As discussed in the beginning of this section, an exhaustive list of tests were performed to spot causes for this possible failure. However, the reason why SBI “struggles” more with  $z = 0.5$  as opposed to  $z = 1.03$  for the baseline prior is still unknown. We have observed that there is information gain from adding the trispectrum at  $z = 1.5$  with the baseline priors, but not at  $z = 0$ , which seems to indicate a trend where SBI struggles more with lower redshifts samples.

One hypothesis that was raised is that SBI would tend to struggle more when the data vector is highly correlated, because at lower redshifts (e.g.,  $z = 0.5$ ) the data vector elements exhibit higher cross-correlations than at higher redshifts (e.g.,  $z = 1.03$ ; see Fig. 6.6). To test this, we performed a Gaussian likelihood analysis as described in Sec. 4.3.2, using Eq. 6.23 for the analytical covariance of the trispectrum. When including the trispectrum, this analysis becomes more challenging, as the number of cross-trispectra to be calculated and interpolated over  $\alpha$  increases significantly. Therefore we only include the deterministic bias parameters in this analysis and not higher-derivative or stochastic terms. For a mock data generated with FTP (“fixed true phases”, see Chapter 5) SNG best fit results, we show in Fig. 6.14 the difference between the Gaussian likelihood analysis when including or not the cross-terms in the covariance, i.e., whether the data is sampled from an analytical or sample covariance.

If the issue was indeed the cross-correlations between the data vector at  $z = 0.5$ , we would expect a dramatic difference when comparing the sample and analytical covariance cases, since the analytical one only has diagonal terms. However, Fig. 6.14 suggests that, although the analytical covariance tends to underestimate the errors, as expected, both posteriors are still highly correlated, and the trispectrum does not add information at this redshift for breaking bias parameters degeneracies. If the prior is narrower, we see information gain from adding the trispectrum at  $z = 0.5$ , for both covariances considered. This tests therefore disfavors the hypothesis that the issue at  $z = 0.5$  comes from the higher cross-correlations between the data vector elements at this redshift.

We have also tested decorrelating the bias parameters in the posterior. From a posterior we have obtained with NPE and for a given set of bias parameters that we want to decorrelate, we can calculate the covariance matrix of the posterior samples  $\Sigma$ , which can be decomposed into  $VDV^T$ . We can use the eigenvectors  $V$  to rotate the bias parameters into a basis where they are uncorrelated in the posterior. We perform the transformation

$$\tilde{b}_{O(n)}\alpha^n = V^T(b_{O(n)}\alpha^n - \langle b_{O(n)}\alpha^n \rangle), \quad (6.30)$$

where  $\tilde{b}_{O(n)}\alpha^n$  are the uncorrelated parameters. We have tried this transformation for the set  $\{b_1\alpha, b_{\delta^3}\alpha^3, b_{K^3}\alpha^3\}$  and for all bias parameters. Again, the results were that the trispectrum added information at  $z = 1.03$  when compared to power-spectrum and bispectrum for both the baseline and narrower priors, while at  $z = 0.5$  the information gain only happens for the narrower prior.

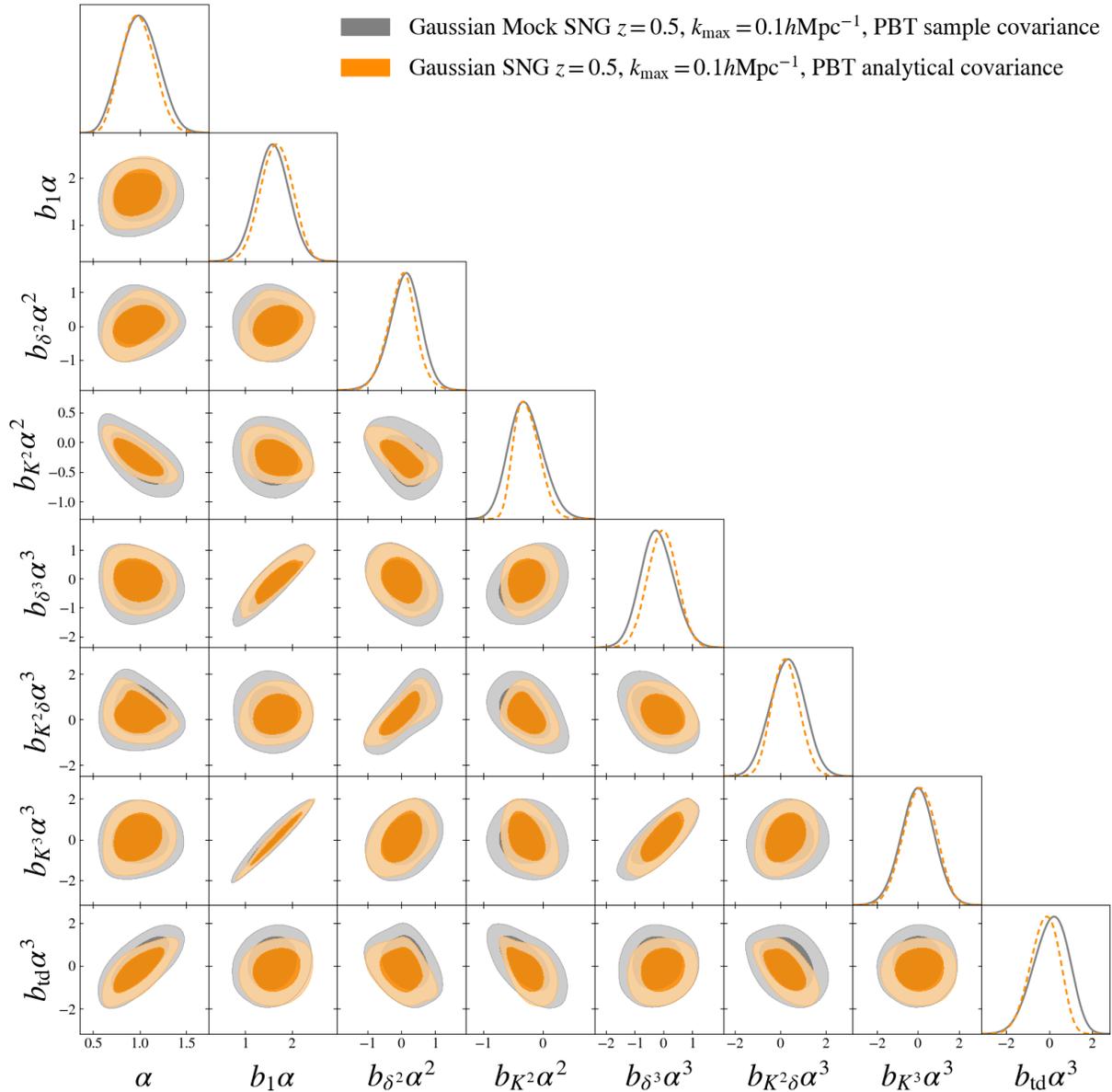


Figure 6.14: Posterior distribution of the model parameters inferred using mock data from SNG FTP parameters at redshift  $z = 0.5$  as the “observed” data vector consisting of the power-spectrum, bispectrum and trispectrum. In this case, the parameter space is reduced to include the cosmological parameter  $\alpha$  and deterministic bias parameters. The data vector generated for training is sampled from a Gaussian likelihood. Grey and orange contours correspond to when we use an analytical or sample covariance. For both cases, the method NPE was used from a simulation budget of  $N_{\text{sim}} = 5 \times 10^5$ , scale cut of  $k_{\max} = 0.1h\text{Mpc}^{-1}$ .

## 6.6 Conclusion & next steps

In this chapter, we presented the first joint SBI analysis using the galaxy power-spectrum, bispectrum, and trispectrum (PBT), applied to dark-matter halos. Using neural posterior estimation (NPE) and  $N_{\text{sim}} = 8 \times 10^5$  simulations from `LEFTfield`, we assessed the impact of including the trispectrum on constraints for the cosmological parameter  $\alpha$ , bias and stochastic parameters. These results are shown in Fig. 6.8 (for Uchuu halos) and in Fig. 6.9 (for SNG halos). Moreover, to the best of our knowledge, this work presents the first validation in the literature of the trispectrum estimator used in this work against the predictions of the tree-level trispectrum contributions  $T_{1122}$  and  $T_{1113}$  (see Sec. 6.3).

We find that the addition of the trispectrum leads to significantly different improvements under the baseline prior for the two halo samples considered: a 20% gain in constraining power for  $\alpha$  in the Uchuu halo sample at  $z = 1.03$ , and no gain for the SNG sample at  $z = 0.5$ . This result appears robust across a wide range of tests, such as variations in binning,  $k_{\text{max}}$ , network architecture, compression scheme, density estimator, and mock data, indicating that the lack of gain in some cases is not due to numerical or architectural limitations.

Instead, we found that high-order bias parameters are highly degenerate with third-order bias being prior dominated, which challenges the inference process. By narrowing the prior (Fig. 6.10 and Fig. 6.11) and reducing the number of parameters (Fig. 6.12 and Fig. 6.13), the constraining power of the trispectrum becomes clearly evident. Using a narrower prior, we observe consistent improvement in  $\alpha$  constraints for both halo samples, with an enhancement factor of 1.4 when adding the trispectrum to the power-spectrum and bispectrum (see Table 6.1).

It is interesting to note the dependence of SBI on the prior when analyzing the combined galaxy power spectrum, bispectrum, and trispectrum at  $z = 0.5$ , in contrast to  $z = 1.03$ . One would expect that sequential SBI would eventually converge to the true posterior, but this does not appear to be the case here. Before publishing the manuscript, we plan to compare the SBI posteriors to those obtained via MCMC to investigate whether the issue stems from inability of SBI to break certain degeneracies. In this comparison, we can use the interpolated cross-spectra from the Gaussian likelihood analysis as predictions for the mean of the data vector, combined with either analytical or sample covariances.

In addition, we plan to investigate the impact of eliminating cross-correlations between bias operators with the orthonormalization scheme proposed in [194], where again the cross-power spectra interpolated in  $\alpha$  can be used in the Gram-Schmidt procedure. This strategy may help reduce parameter degeneracies, as the bias parameters are known to be highly correlated, which poses a challenge for SBI.

We plan to present results using the compression techniques discussed in Appendix C to reduce the dimensionality of the galaxy trispectrum and improve the convergence properties of SBI, especially at higher  $k_{\text{max}}$ . Additionally, we will compare the SBI posteriors to Fisher predictions. We also plan to investigate the impact of higher-order stochastic contributions, specifically those of order 3 (e.g.,  $\epsilon^3$ ), on the galaxy trispectrum in later stages of our

analysis (see Appendix A.2)).

What is clear is that the galaxy trispectrum encodes cosmological information beyond what is accessible through the power spectrum and bispectrum alone. This supports findings such as those of [157] presented in Chapter 5, which demonstrate additional information content in the galaxy density field. While these results are promising, caution is warranted, especially given the novelty of the methods used, such as whether SBI is sufficiently robust in this context and whether the strong bias degeneracies in `LEFTfield` for the  $n$ -point functions can be mitigated. To trust trispectrum-based inference independently of prior assumptions, further validation is essential. In particular, we aim to avoid relying on artificially narrow priors, which could ultimately bias the inferred parameters.



# Chapter 7

## Conclusion

In this thesis, we have explored how to use simulation-based inference (SBI) in the context of galaxy clustering. Our goal was to obtain more robust error bars, since SBI accounts for the true distribution of the data vector, and to achieve higher precision on cosmological parameters by analyzing more informative summary statistics, such as the galaxy trispectrum. Moreover, SBI provided a way to perform an apples-to-apples comparison with field-level inference, revealing the true information content encoded in the galaxy density field for the first time.

In Chapter 2, we reviewed the main perturbative techniques used to construct the galaxy bias expansion, which lies at the core of `LEFTfield`. By the end of this section, it should be clear that `LEFTfield` constitutes a reliable model for galaxy clustering. Despite being valid only on large scales due to its perturbative foundation, it remains agnostic to the complexities of galaxy formation. Furthermore, it is a fast simulator suitable for SBI, enabling convergence tests for the posterior with respect to the number of simulations.

Chapter 3 introduced SBI in detail. Its power and flexibility should be evident: SBI is a sophisticated and promising inference framework that may represent the future of cosmological analysis, especially as we aim to model more complex summary statistics and adopt more reliable simulations. We also described field-level inference, which is enabled by `LEFTfield`. After understanding these inference methods in theory, we moved to deriving cosmological constraints from simulations, an essential first step in testing both SBI and `LEFTfield`.

In Chapter 4, we demonstrated how to combine SBI and `LEFTfield` using measurements of the power spectrum and bispectrum on mock data. This was the first step in our inference pipeline and revealed the need for  $\sim 10^5$  simulations for convergence. Appendix C discusses several techniques that could help reduce this number, with compression via VMIM being the most promising. This compression method is also being tested for inference with the galaxy trispectrum, which is highly dimensional.

In Chapter 5, we discussed the application of SBI to dark matter halos with power-spectrum and bispectrum. Remarkably, SBI performs well even in the presence of model mismatch. We also compared this to field-level inference using `LEFTfield`, showing that the galaxy density field can contain up to five times more information on  $\sigma_8$ . If so much

information lies in the field, we should expect to find signals in the trispectrum. SBI is, in principle, well-suited to handle it, as it avoids issues related to non-Gaussian likelihoods or the estimation and inversion of covariance matrices. However, challenges remain. Our trispectrum results show improvement factors that depend on the prior, reflecting the novelty and complexity of using the trispectrum, SBI, and `LEFTfield`. Despite the difficulty, preliminary results are promising, indicating that the galaxy trispectrum indeed encodes cosmological information.

In the next stages of this research, I plan to use the reconstruction procedure described in Appendix E to combine galaxy  $n$ -point functions before and after reconstruction. SBI is ideal for this, as estimating the combined covariance and theoretical prediction of the data vector is difficult. With SBI and `LEFTfield`, we can use power spectrum and bispectrum measurements both pre- and post-reconstruction to extract tighter cosmological constraints.

Another interesting direction, although still in early stages (and with too few results to include here), is to perform SBI directly at the field level, without requiring an explicit likelihood. This can be achieved by compressing the field using techniques discussed in Appendix C, rather than working with summary statistics. Otherwise, more sophisticated methods than normalizing flows are needed to handle the high dimensionality for sampling the initial conditions. This is particularly interesting in our case, as it allows for a rare comparison with the explicit likelihood approach, which is typically avoided due to its complexity.

# Appendix A

## Bispectrum and trispectrum stochasticity

### A.1 Leading order stochastic bispectrum

The leading order stochastic contribution for the galaxy bispectrum in `LEFTfield` can be written as

$$\begin{aligned} \left\langle \delta_g(\mathbf{k}_1)\delta_g(\mathbf{k}_2)\delta_g(\mathbf{k}_3) \right\rangle \Big|_{\text{stoch.}}^{\text{LO}} &= c_{\varepsilon^2} \left\langle \varepsilon(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)[\varepsilon^2](\mathbf{k}_3) + 2 \text{ perm.} \right\rangle \\ &+ c_{\varepsilon\delta} \left\langle \delta_{g,\text{det}}(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)[\varepsilon\delta](\mathbf{k}_3) + 5 \text{ perm.} \right\rangle, \end{aligned} \quad (\text{A.1})$$

where the brackets denote multiplication in real space. Higher-order contributions are controlled by  $c_{\varepsilon^2}c_{\varepsilon\delta}$  or more than one power of  $c_{\varepsilon^2}$  or  $c_{\varepsilon\delta}$ , which are expected to be small (see below). Note that terms with only one power of  $\delta_{g,\text{det}} \sim \delta$ ,  $\delta$ ,  $\varepsilon$  or  $\varepsilon^2$  vanish due to the zero mean of these operators. Now we calculate the first contribution of Eq. (A.1) as

$$\begin{aligned} \left\langle \varepsilon(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)[\varepsilon^2](\mathbf{k}_3) \right\rangle &= \int_{\mathbf{q}} \left\langle \varepsilon(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)\varepsilon(\mathbf{q})\varepsilon(\mathbf{k}_3 - \mathbf{q}) \right\rangle \\ &= \int_{\mathbf{q}} \left[ \delta_D(\mathbf{k}_1 + \mathbf{q})P_{\varepsilon}(\mathbf{k}_1)\delta_D(\mathbf{k}_2 + \mathbf{k}_3 - \mathbf{q})P_{\varepsilon}(\mathbf{k}_2) \right. \\ &\quad \left. + \delta_D(\mathbf{k}_1 + \mathbf{k}_3 - \mathbf{q})P_{\varepsilon}(\mathbf{k}_1)\delta_D(\mathbf{k}_2 + \mathbf{q})P_{\varepsilon}(\mathbf{k}_2) \right] \\ &= 2\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)P_{\varepsilon}(\mathbf{k}_1)P_{\varepsilon}(\mathbf{k}_2) \\ &= 2\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)P_{\varepsilon}^2, \end{aligned} \quad (\text{A.2})$$

where  $\int_{\mathbf{q}} \equiv \int d\mathbf{q}/(2\pi)^3$  and on the last line we assumed that  $P_\varepsilon(k) = P_\varepsilon, \forall k$ . The second contribution in turn reads

$$\begin{aligned} \langle \delta_{g,\text{det}}(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)[\varepsilon\delta](\mathbf{k}_3) \rangle &= \int_{\mathbf{q}} \langle \delta_{g,\text{det}}(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)\varepsilon(\mathbf{q})\delta(\mathbf{k}_3 - \mathbf{q}) \rangle \\ &= \int_{\mathbf{q}} \left[ \delta_D(\mathbf{k}_1 + \mathbf{k}_3 - \mathbf{q}) P_{\delta_{g,\text{det}},\delta}(\mathbf{k}_1) \delta_D(\mathbf{k}_2 + \mathbf{q}) P_\varepsilon(\mathbf{k}_2) \right] \\ &= \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) P_{\delta_{g,\text{det}},\delta}(\mathbf{k}_1) P_\varepsilon(\mathbf{k}_2) \\ &= \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) b_1 P_\varepsilon P_m(\mathbf{k}_1), \end{aligned} \quad (\text{A.3})$$

where on the last line we additionally assumed that, at leading order,  $\delta_{g,\text{det}}(\mathbf{k}) = b_1\delta(\mathbf{k})$ , and  $P_m$  is the evolved matter power spectrum. Collecting all terms, Eq. (A.1) can be written as

$$\left\langle \delta_g(\mathbf{k}_1)\delta_g(\mathbf{k}_2)\delta_g(\mathbf{k}_3) \right\rangle'_{\text{stoch}} \Big|_{\text{LO}} = 6 c_{\varepsilon^2} P_\varepsilon^2 + 2 b_1 c_{\varepsilon\delta} P_\varepsilon (P_m(\mathbf{k}_1) + 2 \text{ perm.}), \quad (\text{A.4})$$

where the prime denotes the Dirac delta function. We can now do the matching between the above equation and Eq. (2.83) to obtain that

$$6 c_{\varepsilon^2} P_\varepsilon^2 = B_\varepsilon, \quad c_{\varepsilon\delta} P_\varepsilon = P_{\varepsilon\varepsilon\delta}. \quad (\text{A.5})$$

In the Poisson limit, it is expected that  $P_\varepsilon = n_g^{-1}$ ,  $B_\varepsilon = n_g^{-2}$  and  $P_{\varepsilon\varepsilon\delta} = b_1 n_g^{-1}/2$  [189], where  $n_g$  is the galaxy comoving number density. We therefore arrive at the values

$$c_{\varepsilon^2} \stackrel{\text{Poisson}}{=} \frac{1}{6}, \quad c_{\varepsilon\delta} \stackrel{\text{Poisson}}{=} \frac{b_1}{2}, \quad (\text{A.6})$$

which will be used for our fiducial values, as in Eq. (4.10). Note that the leading order stochastic spectra are only used here in order to determine the fiducial value of the stochastic parameters, and the forward model takes into account all loop contributions for cosmological inference.

## A.2 Leading order stochastic trispectrum

From [87, 221], the leading-order stochastic galaxy trispectrum,

$$\langle \delta_g(\mathbf{k}_1)\delta_g(\mathbf{k}_2)\delta_g(\mathbf{k}_3)\delta_g(\mathbf{k}_4) \rangle_{\text{stoch}}^{\text{L.O.}} \equiv (2\pi)^3 \delta_D(\mathbf{k}_{1234}) T_{\text{stoch}}^{\text{L.O.}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \quad (\text{A.7})$$

has Poisson expectation

$$\begin{aligned} T_{\text{stoch}}^{\text{L.O.}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= \frac{1}{\bar{n}} [B(\mathbf{k}_1 + \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) + 5 \text{ perm.}] \\ &+ \frac{1}{\bar{n}^2} [P(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) + 3 \text{ perm.}] \\ &+ \frac{1}{\bar{n}^2} [2P(\mathbf{k}_1 + \mathbf{k}_2) + 2 \text{ perm.}] \\ &+ \frac{1}{\bar{n}^3}. \end{aligned} \quad (\text{A.8})$$

Now, we shall look at the trispectrum generated by LEFTfield, where  $\varepsilon$  is Gaussian, and identify which higher-order stochastic terms will generate. At third order and ignoring higher-derivatives corrections, we have

$$\delta_g = \delta_{g,\text{det}} + \varepsilon + b_{\varepsilon^2} [\varepsilon^2] + b_{\varepsilon^3} [\varepsilon^3] + \sum_{\varepsilon^n \mathcal{O}^m}^{n+m \leq 3} b_{\varepsilon^n \mathcal{O}^m} [\varepsilon^n \mathcal{O}^m] \quad (\text{A.9})$$

More specifically,

$$\varepsilon_{\text{tot}} = c_\varepsilon \left( b_\varepsilon \varepsilon + b_{\varepsilon^2} [\varepsilon^2] + b_{\varepsilon^3} [\varepsilon^3] + \sum_{\varepsilon^n \mathcal{O}^m}^{n+m \leq 3} b_{\varepsilon^n \mathcal{O}^m} [\varepsilon^n \mathcal{O}^m] \right), \quad (\text{A.10})$$

where for simplicity we set  $c_\varepsilon = 1$  and  $b_\varepsilon^2 \langle \varepsilon \varepsilon \rangle \equiv P_\varepsilon$ . This general term can be explicitly written as

$$\sum_{\varepsilon^n \mathcal{O}^m}^{n+m \leq 3} b_{\varepsilon^n \mathcal{O}^m} [\varepsilon^n \mathcal{O}^m] = b_{\varepsilon \delta} [\varepsilon \delta] + b_{\varepsilon \delta^2} [\varepsilon \delta^2] + b_{\varepsilon K^2} [\varepsilon K^2] + b_{\varepsilon \delta^3} [\varepsilon^2 \delta]. \quad (\text{A.11})$$

Now, our task is to identify the momentum structure appearing from the trispectrum in LEFTfield. Then we can simply match the bias parameters to the Poisson expectations.

We consider the PT hierarchy

- 1)  $b_\varepsilon, b_1$
- 2)  $b_{\varepsilon^2}, b_{\varepsilon \delta}, b_2, b_{K^2}$
- 3)  $b_{\varepsilon^3}, b_{\varepsilon \delta^2}, b_{\varepsilon K^2}, b_{\mathcal{O}^{(3)}}$

We are going to consider, first, the L.O. terms appearing up to 2. We generate the expected Poisson momenta structure already at this order.

The leading order galaxy trispectrum then contains, up to permutations,

$$\begin{aligned} T_{\text{stoch}}^{\text{L.O.}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= b_{\varepsilon \delta} b_1^2 \langle \varepsilon(\mathbf{k}_1) [\varepsilon \delta](\mathbf{k}_2) \delta(\mathbf{k}_3) \delta(\mathbf{k}_4) \rangle + 5 \text{ perm.} \\ &+ b_{\varepsilon^2} b_{\varepsilon \delta} b_1 \langle \varepsilon(\mathbf{k}_1) \varepsilon^2(\mathbf{k}_2) [\varepsilon \delta](\mathbf{k}_3) \delta(\mathbf{k}_4) \rangle + 3 \text{ perm.} \\ &+ (b_{\varepsilon \delta})^2 \langle \varepsilon(\mathbf{k}_1) \varepsilon(\mathbf{k}_2) [\varepsilon \delta](\mathbf{k}_3) [\varepsilon \delta](\mathbf{k}_4) \rangle + 5 \text{ perm.} \\ &+ (b_{\varepsilon^2})^2 \langle \varepsilon(\mathbf{k}_1) \varepsilon(\mathbf{k}_2) \varepsilon^2(\mathbf{k}_3) \varepsilon^2(\mathbf{k}_4) \rangle + 5 \text{ perm.} \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned} T_{\text{stoch}}^{\text{L.O.}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= 4P_\varepsilon b_{\varepsilon \delta} b_1^2 [B(|\mathbf{k}_1 + \mathbf{k}_2|, k_3, k_4) + 5 \text{ perm.}] \\ &+ 8P_\varepsilon^2 b_{\varepsilon^2} b_{\varepsilon \delta} b_1 [P(|\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3|) + 3 \text{ perm.}] \\ &+ 4P_\varepsilon^2 b_{\varepsilon \delta}^2 [2P(|\mathbf{k}_1 + \mathbf{k}_2|) + 2 \text{ perm.}] \\ &+ 24P_\varepsilon^3 b_{\varepsilon^2}^2. \end{aligned} \quad (\text{A.13})$$

$$\begin{aligned}
\langle \varepsilon(\mathbf{k}_1)[\varepsilon\delta](\mathbf{k}_2)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4) \rangle &= \int_{\mathbf{q}} \langle \varepsilon(\mathbf{k}_1) \varepsilon(\mathbf{q}) \delta(\mathbf{k}_2 - \mathbf{q}) \delta(\mathbf{k}_3) \delta(\mathbf{k}_4) \rangle \\
&= \int_{\mathbf{q}} \delta_D(\mathbf{k}_2 - \mathbf{q} + \mathbf{k}_3 + \mathbf{k}_4) B_m(|\mathbf{k}_2 - \mathbf{q}|, k_3, k_4) \delta_D(\mathbf{k}_1 + \mathbf{q}) P_\varepsilon(k_1) \\
&= \delta_D(\mathbf{k}_{1234}) P_\varepsilon B_m(|\mathbf{k}_1 + \mathbf{k}_2|, k_3, k_4)
\end{aligned} \tag{A.14}$$

$$\begin{aligned}
\langle \varepsilon(\mathbf{k}_1)\varepsilon^2(\mathbf{k}_2)[\varepsilon\delta](\mathbf{k}_3)\delta(\mathbf{k}_4) \rangle &= \int_{\mathbf{q}_1, \mathbf{q}_2} \langle \varepsilon(\mathbf{k}_1) \varepsilon(\mathbf{q}_1) \varepsilon(\mathbf{k}_2 - \mathbf{q}_1) \varepsilon(\mathbf{q}_2) \delta(\mathbf{k}_3 - \mathbf{q}_2) \delta(\mathbf{k}_4) \rangle \\
&= \int_{\mathbf{q}_1, \mathbf{q}_2} \delta_D(\mathbf{k}_3 - \mathbf{q}_2 + \mathbf{k}_4) P_m(\mathbf{k}_3 - \mathbf{q}_2) \\
&\quad \times [\delta_D(\mathbf{k}_1 + \mathbf{q}_1) P_\varepsilon(k_1) \delta_D(\mathbf{k}_2 - \mathbf{q}_1 + \mathbf{q}_2) P_\varepsilon(|\mathbf{k}_2 - \mathbf{q}_1|) \\
&\quad + \delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1) P_\varepsilon(k_1) \delta_D(\mathbf{q}_1 + \mathbf{q}_2) P_\varepsilon(q_1)] \\
&= 2\delta_D(\mathbf{k}_{1234}) P_m(|\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3|) P_\varepsilon^2
\end{aligned} \tag{A.15}$$

$$\begin{aligned}
\langle \varepsilon(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)[\varepsilon\delta](\mathbf{k}_3)[\varepsilon\delta](\mathbf{k}_4) \rangle &= \int_{\mathbf{q}_1, \mathbf{q}_2} \langle \varepsilon(\mathbf{k}_1) \varepsilon(\mathbf{k}_2) \varepsilon(\mathbf{q}_1) \delta(\mathbf{k}_3 - \mathbf{q}_1) \varepsilon(\mathbf{q}_2) \delta(\mathbf{k}_4 - \mathbf{q}_2) \rangle \\
&= \int_{\mathbf{q}_1, \mathbf{q}_2} \delta_D(\mathbf{k}_3 - \mathbf{q}_1 + \mathbf{k}_4 - \mathbf{q}_2) P_m(\mathbf{k}_3 - \mathbf{q}_1) \\
&\quad \times [\delta_D(\mathbf{k}_1 + \mathbf{q}_1) P_\varepsilon(k_1) \delta_D(\mathbf{k}_2 + \mathbf{q}_2) P_\varepsilon(k_2) \\
&\quad + \delta_D(\mathbf{k}_1 + \mathbf{q}_2) P_\varepsilon(k_1) \delta_D(\mathbf{k}_2 + \mathbf{q}_1) P_\varepsilon(k_2)] \\
&= \delta_D(\mathbf{k}_{1234}) P_\varepsilon^2 [P_m(|\mathbf{k}_1 + \mathbf{k}_3|) + P_m(|\mathbf{k}_1 + \mathbf{k}_4|)]
\end{aligned} \tag{A.16}$$

The last line follows from  $T_{1122}$ ,

$$\langle \varepsilon(\mathbf{k}_1)\varepsilon(\mathbf{k}_2)\varepsilon^2(\mathbf{k}_3)\varepsilon^2(\mathbf{k}_4) \rangle = 4P_\varepsilon^3 \tag{A.17}$$

## Appendix B

## Appendix C: Bias relations

The bias expansion is constructed upon a sum of operators at given order in perturbation theory at a certain given time, where one can always relate a given basis of operators one to another. From the fiducial value of the linear bias  $b_1$ , which is related to the Eulerian basis  $\mathbf{O}_E = \{\delta, \delta^2, K^2\}$ , where  $K^2 \equiv (K_{ij})^2$  and

$$K_{ij}(\mathbf{k}) \equiv \left[ \frac{k_i k_j}{k^2} - \frac{1}{3} \delta_{ij} \right] \delta(\mathbf{k}), \quad (\text{B.1})$$

we wish to determine a physically-motivated fiducial value for the second-order bias parameters associated to the Lagrangian operators  $\mathbf{O}_L = \{\text{tr}[\mathbf{M}^{(1)}], \text{tr}[\mathbf{M}^{(1)}]^2, \text{tr}[\mathbf{M}^{(1)}\mathbf{M}^{(1)}]\}$ . At leading order, these operators can be related as

$$\sigma^2 \equiv \left( \text{tr}[\mathbf{M}^{(1)}] \right)^2 = \delta^2, \quad \text{tr}[\mathbf{M}^{(1)}\mathbf{M}^{(1)}] = K^2 + \frac{1}{3}\delta^2. \quad (\text{B.2})$$

and assuming a local Lagrangian bias model with vanishing Lagrangian tidal bias, we can use the co-evolution relations [66] to obtain that the conversion between both sets of bias parameters reads

$$\begin{aligned} b_{\sigma\sigma} &= b_2 - \frac{4}{21}(b_1 - 1), \\ b_{\text{tr}[\mathbf{M}^{(1)}\mathbf{M}^{(1)}]} &= \frac{1}{3} \left( b_2 - \frac{4}{21}(b_1 - 1) \right), \end{aligned} \quad (\text{B.3})$$

where the bias terms with no subscripts are the Eulerian bias. By setting  $b_1 = 1.5$ ,  $b_2$  is determined by the fitting formula  $b_2(b_1)$  presented in Ref. [123] as  $b_2 = -0.69$ . Our second order bias terms are then set to the fiducial values  $\bar{b}_{\sigma\sigma} = -0.79$  and  $\bar{b}_{\text{tr}[\mathbf{M}^{(1)}\mathbf{M}^{(1)}]} = -0.26$ , as in Eq. (4.10). Note that the fiducial linear bias is  $\bar{b}_1 = 1.5$ , since it is the Eulerian linear bias by construction.



# Appendix C

## Optimizing SBI

In this section, we analyze whether we can improve on the number of simulations needed for convergence in SBI. In light of Fig. 4.12 and Fig. 5.9, the number of simulations needed for convergence when using power-spectrum and bispectrum to constrain  $\alpha$  is of order  $10^5$ . In general, sequential algorithms tend to decrease the simulation budget needed, as evident in Fig. 4.13. However, posteriors obtained by sequential methods are not amortized, and we are restrained from performing SBC tests on those. It is therefore desirable to improve the convergence properties of SBI, i.e., look for strategies that make the convergence of the posterior as a number of simulation budget to flatten at lower values. By having `LEFTfield` at hand, in principle we can generate as many simulations as we want for estimating the posterior with SBI. However, as we increase the parameter space and the dimensionality of the data vector, generating orders of magnitude of more simulations becomes increasingly more challenging. Note that, if we artificially lower the prior, the simulation budget for convergence decreases. However, we always intend to use uninformative priors to avoid biased inference.

### C.1 Convergence with simulation budget

We start our discussion by noting the convergence properties and validation loss for a simple case, the linear forward model considered in [217] (Sec. 4.3.1). As we can see in Fig. C.1, since we only sample two parameters and use the power-spectrum for analysis, we note that in this case both the standard deviation of  $b_1^2$  and the validation loss of NPE tend to converge pretty quickly, at around  $N_{\text{sim}} \sim 10^3$ .

We now consider the *Uchuu* P+B *lower cutoff* case of [157] where we use both power-spectrum and bispectrum (P+B) at  $k_{\text{max}} = 0.12h\text{Mpc}^{-1}$ , with total dimensionality  $D = 443$ , to infer 11 parameters ( $\alpha$ , third order bias, Gaussian stochasticity and leading order higher-derivative terms). Further details on the data and model can be found in Sec. 5.2 and of SBI in Sec. 5.5. As we can see in Fig. C.2, validation convergence in this case is achieved at least for  $N_{\text{sim}} = 2 \times 10^5$ . One can ask therefore what can be done to improve the convergence properties in this case, as we discuss in the following sections.

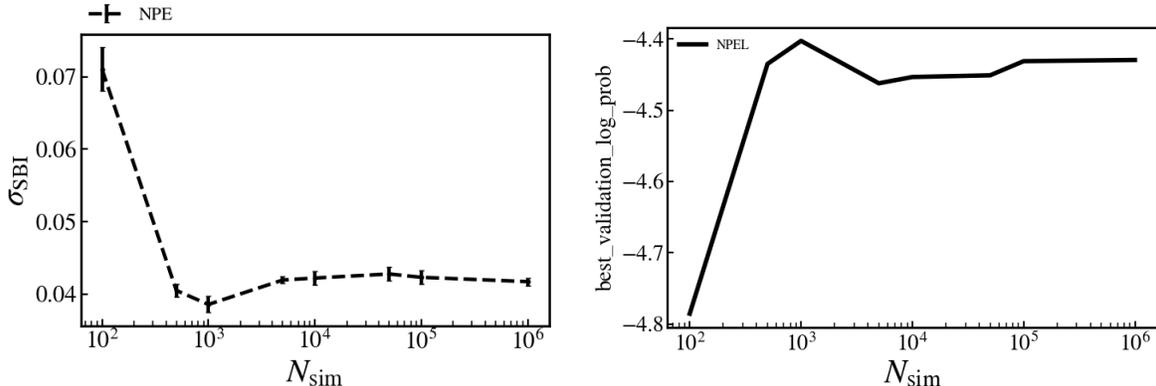


Figure C.1: *Left*: standard deviation of  $b_1^2$  obtained by NPE as a function of simulation budget used in training. *Right*: logarithm of the best validation loss over all epochs obtained by NPE as a function of simulation budget used in training.

## C.2 Hyperparameter optimization

Since in Fig. C.2 the hyperparameters of SBI were fixed, here we try to check how hyperparameter optimization can improve the validation loss at fixed simulation budget using OPTUNA [14]. The idea discussed in the literature is that, if you train multiple models (for example, 100), and do an ensemble of a fraction of them with best objective function to be optimized (for example, 10 models with best validation loss), you can minimize the neural network error and obtain better convergence properties. We use again the Uchuu P+B for lower cutoff, where we use NPE as the neural density estimator and MAFs for constructing the normalizing flow. The hyperparameters changed by OPTUNA are, with baseline values in paranthesis:

- “nt”: number of transforms in the flow (10);
- “lr”: learning rate ( $10^{-4}$ );
- “bs”: batch size (50);
- “hf”: number of hidden units (100).

In Fig. C.3, Fig. C.4 and Fig. C.5, we keep the simulation budget fixed ( $N_{\text{sim}} = 5 \times 10^4$ ,  $N_{\text{sim}} = 10^5$  and  $N_{\text{sim}} = 5 \times 10^5$ , respectively), and change the hyperparameters listed above for training NPE. We check how hyperparameter optimization changes both the standard deviation of  $\alpha$  (as a proxy for convergence) and best validation losses (as a proxy for the quality of the trained model). From Fig. C.2, we saw that, with the baseline hyperparameters,  $N_{\text{sim}} = 5 \times 10^4$  is far from convergence,  $N_{\text{sim}} = 10^5$  is close to convergence and  $N_{\text{sim}} = 5 \times 10^5$  is already converged. We want to investigate here whether hyperparameter optimization can help with convergence, and we would expect,

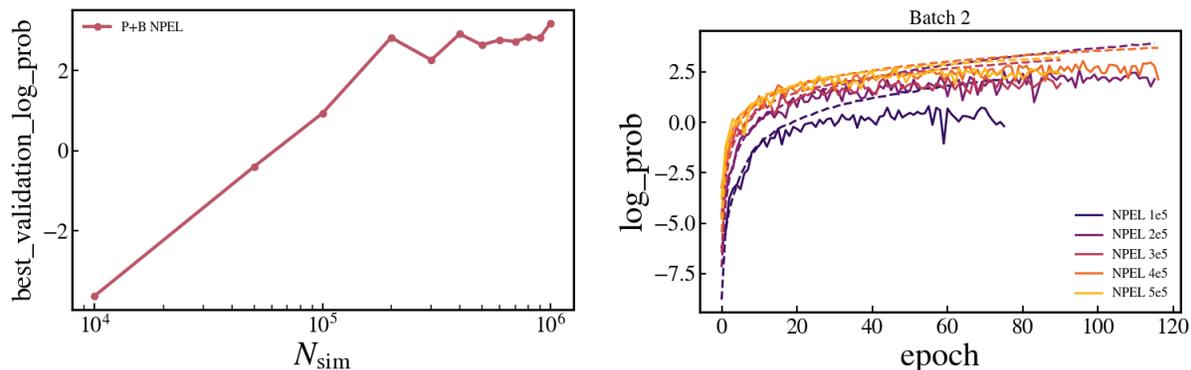


Figure C.2: *Left*: logarithm of the best validation loss over all epochs obtained by NPE as a function of simulation budget used in training. *Right*: logarithm of the validation (training) loss over batches obtained by NPE in continuous (dashed) lines as a function of training epoch for different simulation budgets.

for a the cases with limited simulation budget, the models with best validation loss to be correlated with a lower error on  $\alpha$ , since usually the trend is that the error decreases until convergence. We have to be careful about our conclusions, which can change depending on the dimensionality of parameter and data vector spaces, type of summary statistics being used, neural density estimator considered, etc. We draw our conclusions for the specific case considered here.

In Fig. C.3, it is interesting to notice how with a limited simulation budget ( $N_{sim} = 5 \times 10^4$ ), you never achieve the validation loss even of the baseline trained with  $N_{sim} = 10^5$  (indicated in black). This seems to indicate that adding more simulations is more important than hyperparameter tuning for achieving a better model. The error on  $\alpha$  presents a huge scatter when we change the hyperparameters, going above and below the baseline models with both  $N_{sim} = 10^5$  and  $N_{sim} = 5 \times 10^5$ . However, the few posteriors that yield a smaller standard deviations for  $\alpha$  do not have a significantly better validation loss, as we can see from the correlation plot. These results also indicate that there is no correlation between convergence and improvement of the model when doing hyperparameter optimization for a limited simulation budget.

In Fig. C.4 and Fig. C.5, we can see that the error on  $\alpha$  tends to show a smaller scatter after convergence ( $N_{sim} = 5 \times 10^5$ ) when we change the hyperparameters. However, we do not see any evident correlation between the best validation loss and a smaller error on  $\alpha$  for a fixed simulation budget.

We conclude that, although hyperparameter optimization is important for reducing the intrinsic neural network error on the parameters posterior, it does not significantly help with convergence. Adding more simulations to the training data will always yield to a better performance, both in terms of validation loss as well as (reliable) convergence properties.

### C.3 Data compression

In Sec. 3.5, we discussed a few data compression strategies that can be used for reducing the dimensionality of the data vector. Here, we investigate whether these techniques can improve convergence in SBI, again for the Uchuu P+B lower cutoff case.

**MOPED.** MOPED is an optimal compression scheme when the data vector has a perfect Gaussian likelihood and no parameter dependence, which is not exactly true for our case considered here. However, MOPED has the advantage of not having additional free parameters as in the case of network-based compression schemes, and is (arguably) more interpretable. By inspecting Eq. 3.23, we need to estimate the covariance and the derivative of the data vector with respect to the model parameters. This is done exactly as for Fisher analysis, described in Sec. 5.5. In essence, we fix the fiducial parameters to the MAP of the posterior; the sample covariance is estimated over  $10^5$  realizations of initial conditions, while the derivatives are obtained by finite differences over  $10^3$  realizations.

**VMIM.** VMIM requires training additional parameters  $\varphi$  corresponding to the compression network  $F_\varphi$  besides  $\phi$  of the network  $q_\phi$ . This is implemented in a straightforward way, as the loss function is the same as NPE; the difference is that we embed a compression network  $F_\varphi$  to the SBI algorithm. For the architecture, given that our data vector is one dimensional, we have tried both a fully connected network (“FCE”) with layers and hidden features and a 1D CNN (“CNN”) with . We have experimented with the hyperparameters in this case, although it is not shown here. Interestingly, in [201] it was found that, independently of the architecture used, VMIM yielded better cosmological constraints when compared to other loss functions used.

**MSE.** We have also tried MSE using a fully connected network, although the results are not shown here due to its poor performance in our case. In particular, the validation metric (inferred versus true parameters, for example) for higher-order bias parameters was pretty low. Further investigation is necessary in terms of changing the architecture and hyperparameter optimization in this case. This is however somewhat consistent to what was found in [201], where VMIM showed significantly better performance than MSE.

**Normalization.** We experiment also normalizing the data vector input in two ways: “PBnorm” corresponds to dividing the power-spectrum and bispectrum by their tree-level expectation values, as described in Sec. 4.4, while “PBstd” we normalize each bin by subtracting the mean and dividing by the standard deviation obtained in the simulations used for the covariance calculation.

**Results.** In Fig. C.6, we show both the validation loss and error on  $\alpha$  as a function of simulation budget, using the baseline hyperparameter values for NPE (“NPEL”). The best performance was achieved by VMIM; particularly “NPEL PBstd FCE VMIM” was

the best model, which has the “PBstd” normalization and VMIM using a fully connected network as the compression scheme. In this case, we can expect convergence at half to one order of magnitude before the baseline case with no compression. MOPED is also expected to help with the convergence properties, although VMIM overcomes it. We conclude that data compression can help with SBI convergence, as expected.

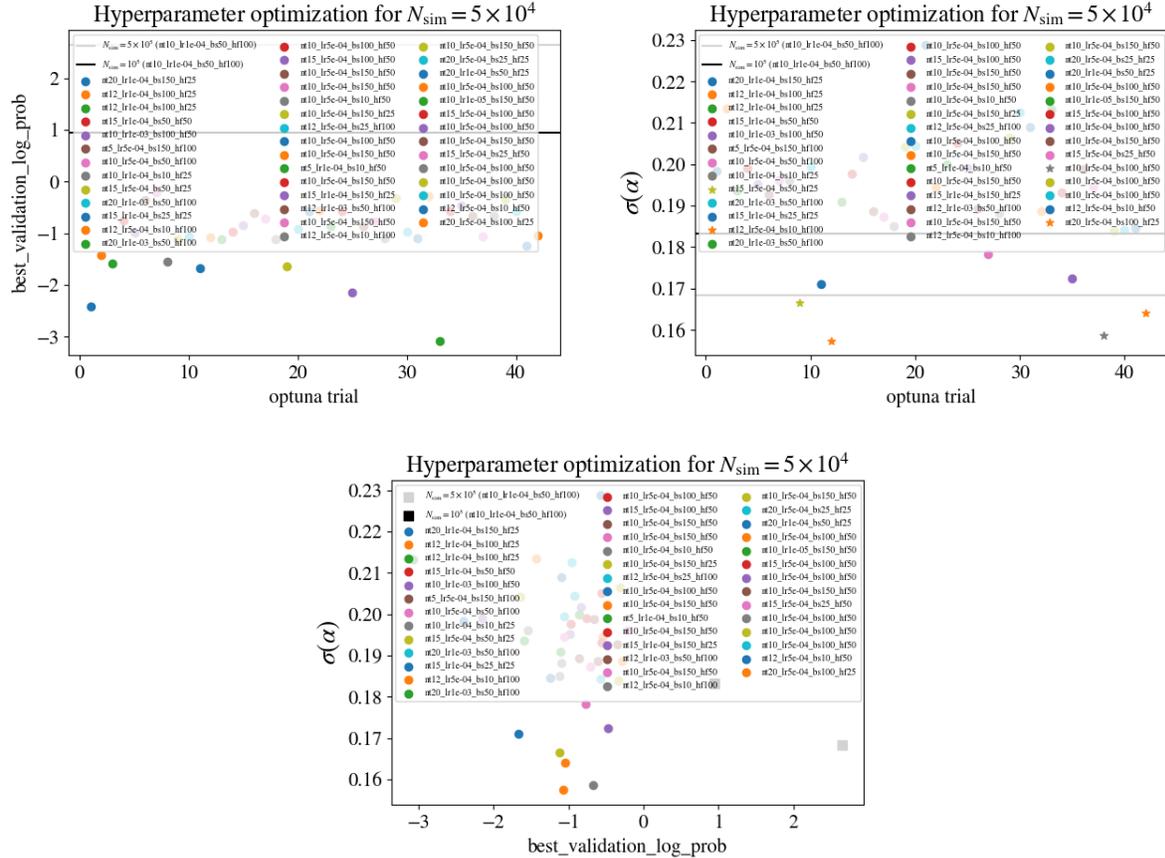


Figure C.3: Hyperparameter optimization for NPE when using  $N_{\text{sim}} = 5 \times 10^4$ . Grey (black) horizontal lines indicates NPE results using baseline hyperparameters and  $N_{\text{sim}} = 10^5$  ( $N_{\text{sim}} = 5 \times 10^5$ ). See the hyperparameters list C.2 for a description of specific hyperparameters chosen by OPTUNA in each label. *Top left:* logarithm of the best validation loss over all epochs obtained by NPE as a function of OPTUNA trial. No model presents a better validation loss than the baselines with more simulations. *Top right:* standard deviation of  $\alpha$  obtained by NPE as a function of OPTUNA trial. The models with lower error on  $\alpha$  than the baseline with  $N_{\text{sim}} = 5 \times 10^5$  (in grey) are displayed with a star mark. *Bottom:* correlation of standard deviation of  $\alpha$  and best validation loss over all epochs obtained by NPE for each OPTUNA trial.

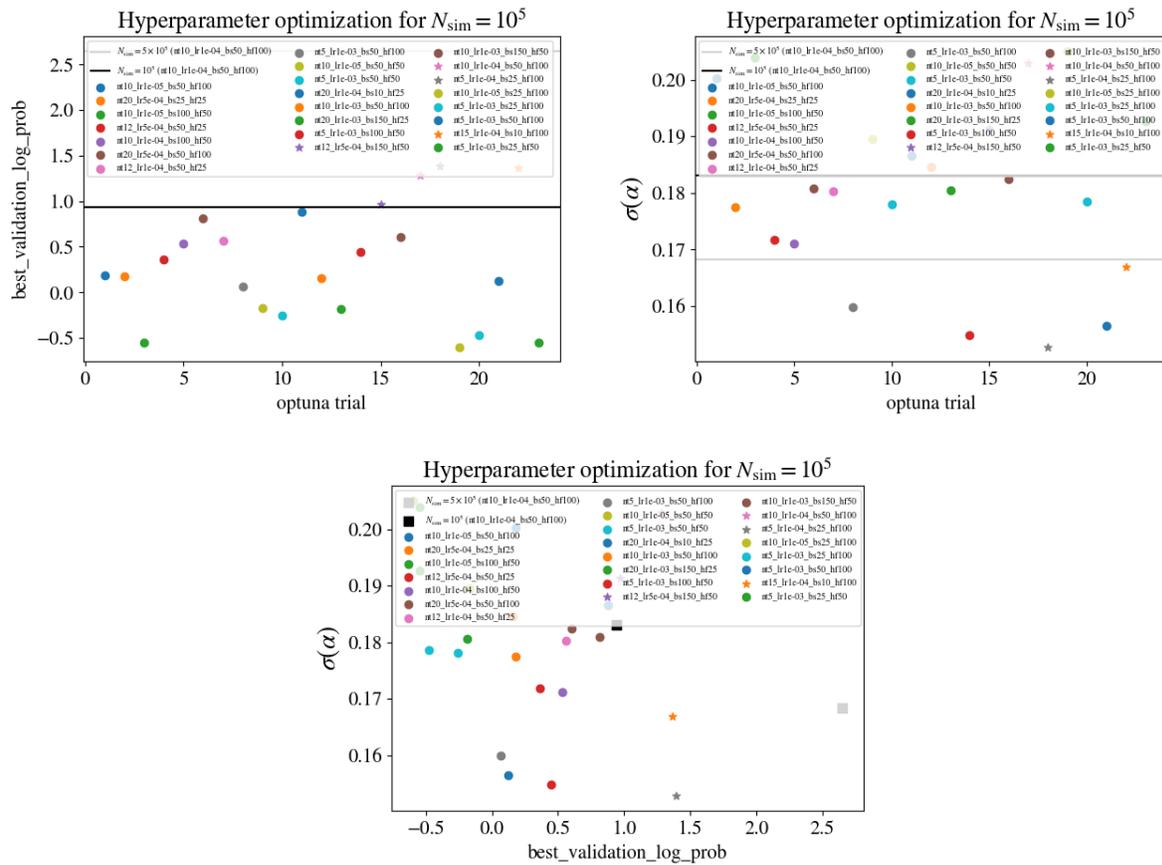


Figure C.4: Same as Fig. C.3, but the hyperparameter optimization for NPE is done when using  $N_{\text{sim}} = 10^5$ . The models with higher validation loss than the baseline with  $N_{\text{sim}} = 10^5$  (in black) are displayed with a star mark.

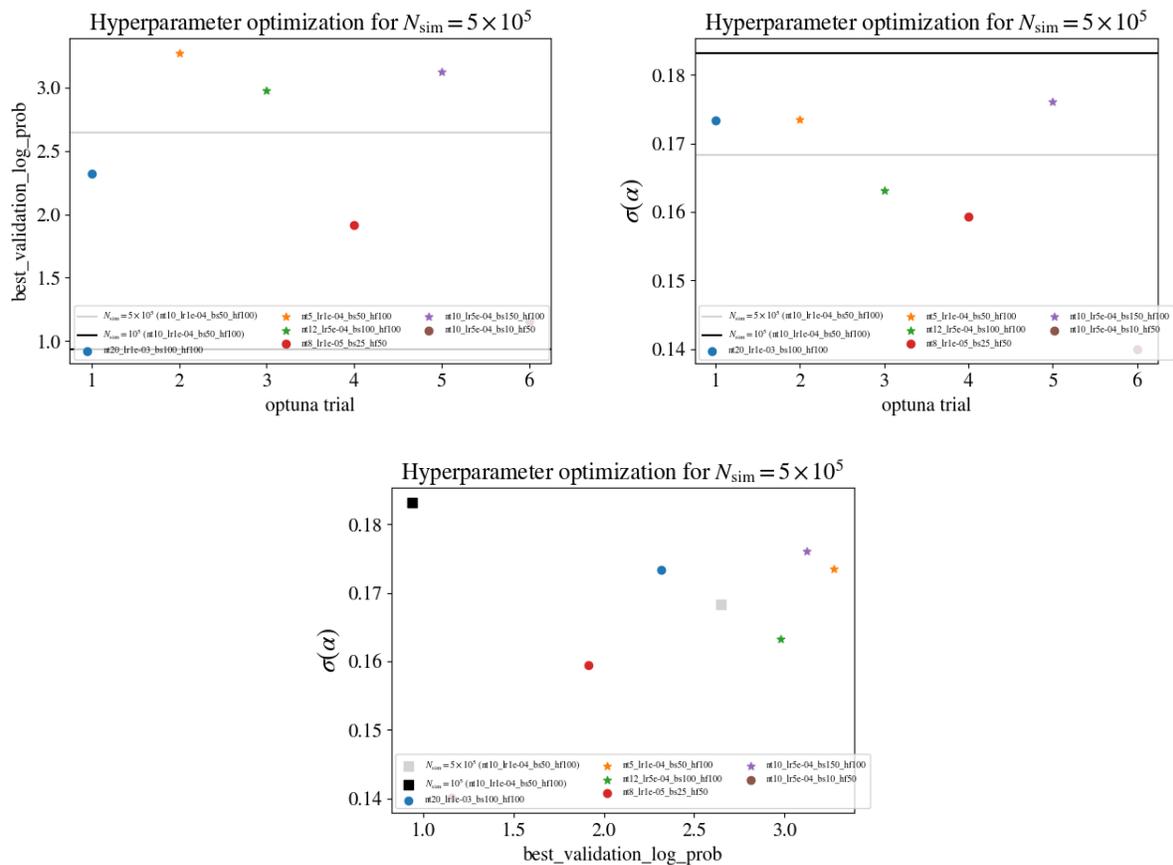


Figure C.5: Same as Fig. C.3, but the hyperparameter optimization for NPE is done when using  $N_{\text{sim}} = 5 \times 10^5$ . The models with higher validation loss than the baseline with  $N_{\text{sim}} = 5 \times 10^5$  (in grey) are displayed with a star mark.

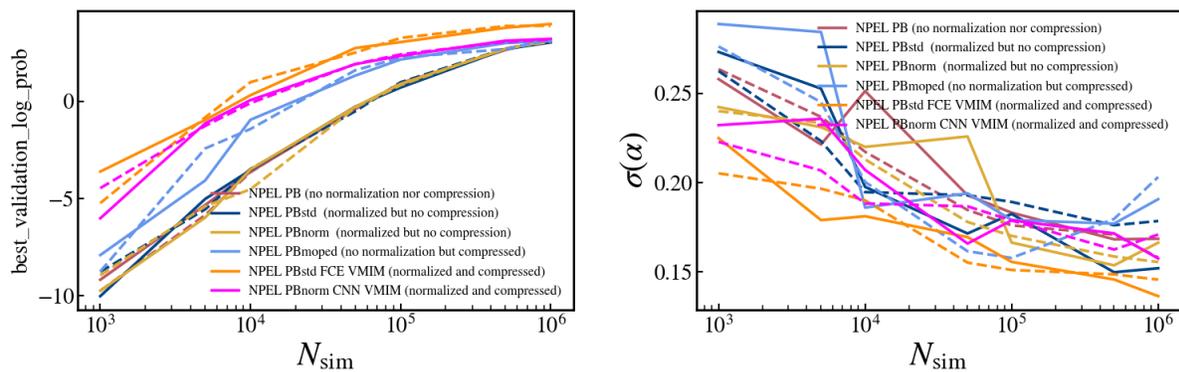


Figure C.6: Data compression impact on convergence properties of NPE. Solid and dashed lines correspond to different runs with exact same configurations and indicate the intrinsic network error. *Left*: logarithm of the best validation loss over all epochs obtained by each method as a function of simulation budget used in training. *Right*: standard deviation of  $\alpha$  obtained by NPE as a function of simulation budget used in training.



# Appendix D

## On the non-Gaussianity of the $n$ -point functions

The main premise of [217] is analyzing the impact of the non-Gaussianity of the galaxy  $n$ -point functions at low- $k$  in cosmological inference. To illustrate the problem, we first consider  $10^5$  realizations of the power spectrum,  $P(k) = \{P(k_1), P(k_2), \dots, P(k_{N_{\text{bin}}})\}$ , that we measure from LEFTfield. In practice, we are measuring the true distribution of the data vector, which is expected to be non-Gaussian on large scales due to the break of central limit theorem (see discussion in Chapter 4). We keep the parameters fixed to the MAP of the posterior, while we only change the initial conditions realization, as is done for calculating the data vector covariance. From this set of simulations, we can calculate the mean  $\langle P(k) \rangle$  and the standard deviation  $\sigma[P(k)]$  over the realizations for each bin and define two distributions,

$$\mathcal{P} = (P(k) - \langle P(k) \rangle) / \sigma[P(k)], \quad (\text{D.1})$$

and

$$\mathcal{P}_{\text{Gaussian}} = (P_G(k) - \langle P_G(k) \rangle) / \sigma[P_G(k)], \quad P_G(k) \sim \mathcal{N}(\langle P(k) \rangle, \sigma^2[P(k)]). \quad (\text{D.2})$$

If the full distribution of measured spectra  $\mathcal{P}$  is Gaussian, we expect it to match perfectly that of a perfectly Gaussian distribution,  $\mathcal{P}_{\text{Gaussian}}$ . This exercise can be repeated also for the bispectrum and integrated trispectrum over multiple bins. We show in Fig. D.1 the results for each of the  $n$ -point functions, where we divide both histograms by  $\max[\mathcal{P}]$  so that the maximum of the histogram is unity. As expected, we see larger deviations from Gaussianity at larger scales, and the deviations appear to be more significant for higher-order  $n$ -point functions. This is confirmed in Fig. D.2, where we calculate the KL divergence between these two distributions for each  $n$ -point function.

In standard cosmological analysis, by assuming that the likelihood of the  $n$ -point functions is Gaussian, we are assuming that they are distributed as the grey lines in Fig. D.1. Although we have shown in [217] that this does not impact  $\alpha$  inference for the power-spectrum and bispectrum, the advantage of SBI is that we will always use the true data vector distribution (for example, the colored lines), without having to assume a specific form for their distribution and worrying whether it will impact cosmological inference.

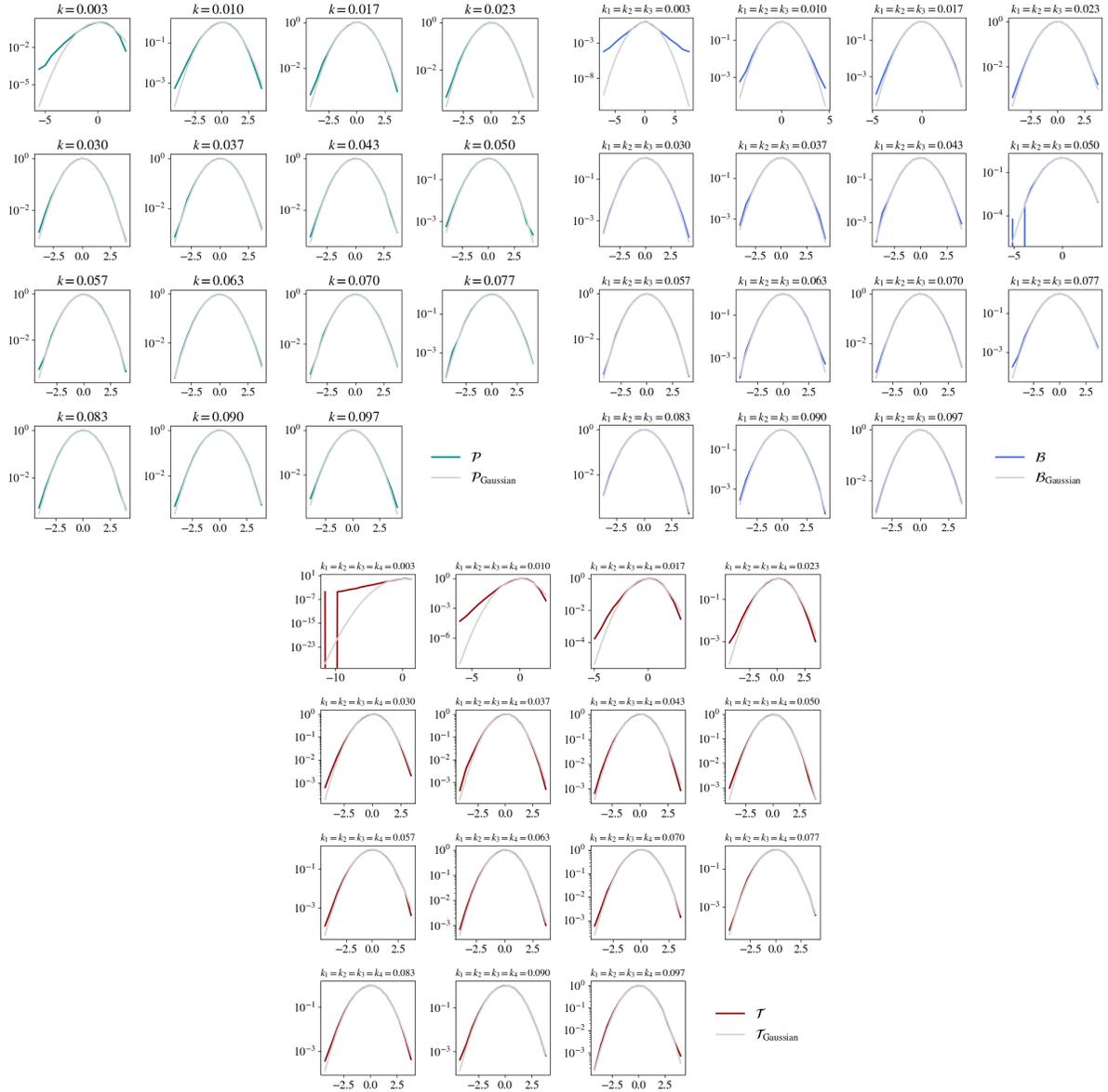


Figure D.1: Distribution of galaxy  $n$ -point functions over initial conditions realizations measured in LEFTfield (in colored lines) compared to a Gaussian distribution (in grey). *Top left:* measured power-spectrum distribution  $\mathcal{P}$  (in green) and Gaussian distribution  $\mathcal{P}_{\text{Gaussian}}$  (in grey) for different  $k$ -bins. *Top right:* measured bispectrum distribution  $\mathcal{B}$  (in blue) and Gaussian distribution  $\mathcal{B}_{\text{Gaussian}}$  (in grey) for different triangle bins. We only select equilateral bins in this illustration ( $k_1 = k_2 = k_3$ ). *Bottom:* measured integrated trispectrum distribution  $\mathcal{T}$  (in red) and Gaussian distribution  $\mathcal{T}_{\text{Gaussian}}$  (in grey) for different quadrilateral bins. We only select equilateral bins in this illustration ( $k_1 = k_2 = k_3 = k_4$ ). The bump observed in the first panel is due to discreteness effects (i.e, it would not be present if we used more simulations).

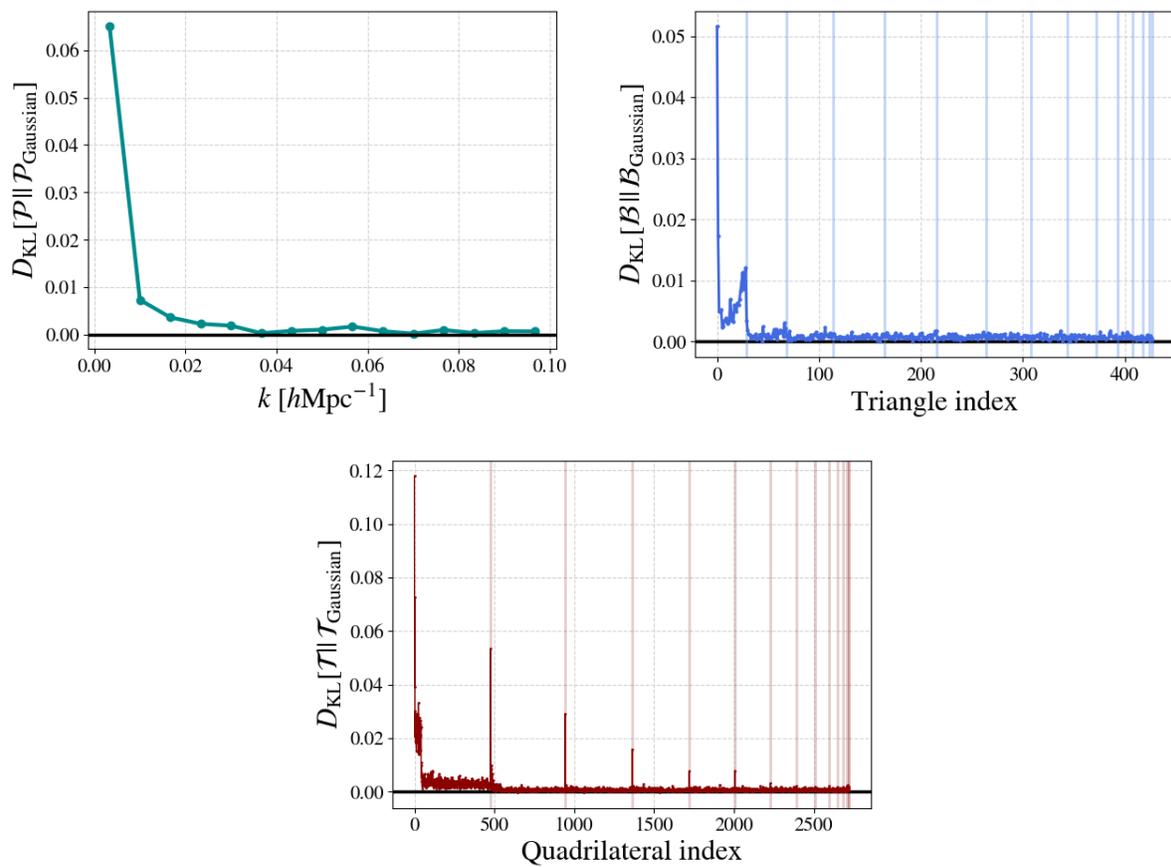


Figure D.2: KL divergence between the full and Gaussian distribution of each  $n$ -point function. The vertical lines in the bispectrum and trispectrum figures correspond to when the innermost index starts running again, i.e., when we go to large scales at the left parts of the plots.



# Appendix E

## Field-level inference of the BAO scale

*This chapter contains a brief summary of [23] (published in JCAP) and [24] (submitted to JCAP). Authors: Ivana Babić, Fabian Schmidt & Beatriz Tucci.*

The purpose of these works is constraining the BAO scale at the field-level using `LEFTfield`. In the first paper [23] (Sec. E.1), we keep the initial conditions fixed, sampling the BAO scale together with bias and stochastic parameters. We present a comparison of field-level constraints on the BAO scale against power-spectrum analysis. In the second paper [24] (Sec. E.2), we infer all parameters at the field-level jointly with the initial conditions. We then present a comparison with pre- and post-reconstruction power-spectrum analysis.

My main contributions were helping the first author, Ivana Babić, with the development of the fixed-phases power spectrum covariance for the first paper and the reconstruction algorithm for the second paper (see Sec. E.2). All analysis presented here were primarily conducted by her.

### E.1 Fixed initial conditions

This section presents a summary of [23]. We compared field-level inference of the BAO scale from halo catalogs using a forward modeling approach based on `LEFTfield` with the traditional method relying on power spectrum compression. In this work, we have used 3LPT and third-order Lagrangian bias expansion, and Gaussian noise with the leading order higher/derivative stochastic operators.

### Modeling the BAO at the field-level

The standard method for extracting the BAO signal involves comparing the observed scale of the BAO feature with predictions, allowing one to infer the angular diameter distance to a given redshift under an assumed cosmology. However, this strategy is not directly applicable in our context. Since we are working with simulations in a cubic box with periodic boundary conditions, altering distances would require the introduction of a window

function and would prevent us from maintaining fixed initial conditions corresponding to the true cosmology. Modifying the assumed distance scale also changes the comoving volume, complicating the comparison with our fixed-volume simulations. To avoid these issues, we take a different route: we effectively rescale the predicted comoving sound horizon instead.

Our goal is to extract the BAO scale  $r_s$  using only the oscillatory component of the power spectrum, excluding its broad-band part, which is influenced by other cosmological parameters. A straightforward method might be to vary the baryon density  $\omega_b$  and find which value best reproduces the observed oscillations. However, adjusting  $\omega_b$  modifies both the oscillatory and smooth components of the power spectrum, making this approach unsuitable.

We instead approximate the linear matter power spectrum with the form

$$P_L(k, \beta) = P_{L,\text{sm}}(k)[1 + A \sin(k\beta r_{\text{fid}}) \exp(-k/k_D)], \quad (\text{E.1})$$

where  $A$  and  $k_D$  are fixed constants, and  $r_{\text{fid}}$  is the fiducial value of the BAO scale. This formulation separates the broad-band component, given by  $P_{L,\text{sm}}(k)$ , from the oscillatory part of the spectrum. The sinusoidal term,  $\sin(k\beta r_{\text{fid}})$ , encodes the BAO signal, while the exponential factor represents Silk damping due to photon diffusion before recombination, which is not captured in the fluid approximation.

We define the scaling factor

$$\beta = r_s/r_{\text{fid}}, \quad (\text{E.2})$$

so that varying  $\beta$  corresponds to adjusting the BAO scale  $r_s$  while keeping the overall distance scale fixed. Importantly, since the BAO signal was imprinted during the early Universe, altering its scale in the initial linear power spectrum is physically justified. This modification changes only the oscillatory structure of the spectrum without affecting its global shape.

The smooth power spectrum  $P_{L,\text{sm}}(k)$  takes the form depending on the smooth transfer function of [72] and with free parameters fitted by CLASS code [130].

Given the known fiducial power spectrum—used to generate the initial conditions in our simulations—we can derive the power spectrum corresponding to a different BAO scale using Eq. (E.1). We define a modulation factor  $f(k, \beta)$  such that

$$f^2(k, \beta) = \frac{P_L(k, \beta)}{P_{\text{fid}}(k)} = \frac{1 + A \sin(k\beta r_{\text{fid}}) \exp(-k/k_D)}{1 + A \sin(kr_{\text{fid}}) \exp(-k/k_D)}. \quad (\text{E.3})$$

This function satisfies  $f(k, 1) = 1$  and allows us to relate the fiducial and rescaled linear density fields via

$$\delta_\beta(k, \beta) = f(k, \beta)\delta_{\text{fid}}(k). \quad (\text{E.4})$$

Thus,  $\delta_\beta$  represents a linear density field with the same phases as the fiducial one but with a modified BAO scale  $r_s = \beta r_{\text{fid}}$ . These rescaled fields  $\delta_\beta$  are used as the initial conditions in our forward modeling throughout this work.

## Results

In this work, we have used 3LPT and third-order Lagrangian bias expansion, while we assume gaussian stochasticity with leading-order higher-derivative term. For the data, we consider a box with size  $L = 2000h^{-1}\text{Mpc}$ . We use four halo mass bins in the mass range  $10^{12.5}h^{-1}M_{\odot} - 10^{14.5}h^{-1}M_{\odot}$ , with number densities spanning  $\bar{n}_h \sim 10^{-5} - 10^{-4} h^3/\text{Mpc}^3$ , depending on the mass and redshift.

The field-level inference results show that systematic biases in  $\hat{\beta}$  are well controlled, remaining below  $\sim 2\%$  across all samples. To benchmark the EFT likelihood, we compared it against a Gaussian likelihood constructed from the power spectrum using the same underlying forward model and bias parameter fits from the field-level analysis, where we use the fixed-phases covariance derived in the paper.

A direct comparison of the two approaches (Fig. E.1) reveals that while both methods perform similarly at small cutoffs, the EFT likelihood significantly outperforms the power spectrum-based inference at  $\Lambda > 0.12 h \text{Mpc}^{-1}$ . At  $\Lambda = 0.2 h \text{Mpc}^{-1}$ , the BAO scale uncertainties from the power spectrum are up to 3.3 times larger, depending on the halo sample. This improvement is expected since the EFT likelihood incorporates higher-order correlations and bulk-flow information not accessible to the power spectrum alone.

## E.2 Free initial conditions

This section presents a summary of [24]. In this work, we have reported the results of field-level inferences of the BAO scale, performed using `LEFTfield` when simultaneously inferring the BAO scale, initial conditions, bias, and stochastic parameters, and has been applied to mock datasets in real space (i.e., neglecting redshift-space distortions). We also provide a comparison to the reconstructed power spectrum.

### BAO reconstruction

After generating the mock catalog, we proceed with the reconstruction process. To ensure an unbiased comparison with the field-level method and maintain consistency in the accessible  $k$ -modes, we carefully match the smoothing scale and grid sizes. The smoothing choice is particularly crucial since the field-level method uses a sharp- $k$  filter, whereas standard reconstruction applies a Gaussian filter.

For standard reconstruction, we set the Gaussian smoothing scale to  $R = \Lambda^{-1}$ . This choice is conservative, as it permits contributions from modes with  $k > \Lambda$  in standard reconstruction while excluding them in the field-level approach. The reconstruction process for a given cutoff  $\Lambda < \Lambda_0$  follows these steps (see Fig. E.2 for a flowchart summarizing the procedure):

1. The galaxy tracers are assigned to a grid of size  $N_{\text{assign}} = 2N^{\Lambda}$  using a NUFFT-based assignment, yielding the initial density field  $\delta_g$ .

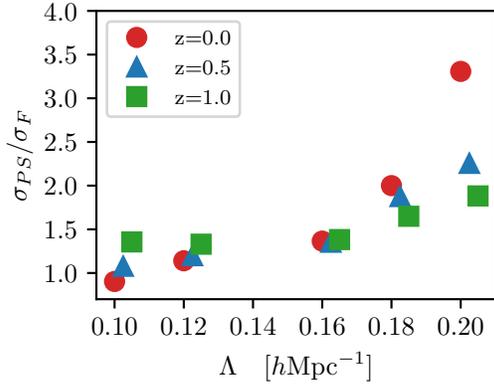
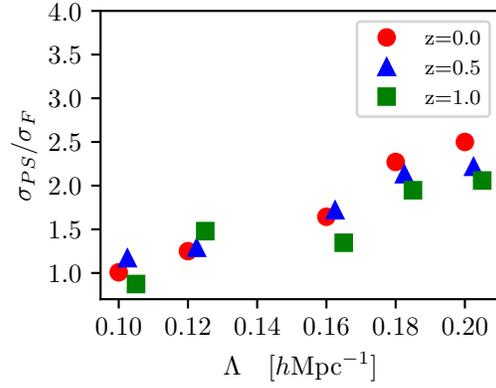
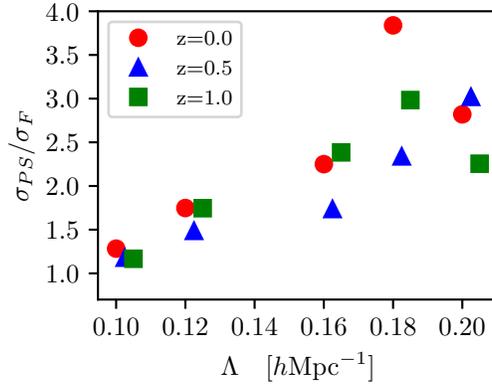
(a)  $\log_{10}(M/h^{-1}M_{\odot}) = 12.5 - 13.0$ (b)  $\log_{10}(M/h^{-1}M_{\odot}) = 13.0 - 13.5$ (c)  $\log_{10}(M/h^{-1}M_{\odot}) = 13.5 - 14.0$ 

Figure E.1: Ratio of the uncertainty on the BAO scale inferred from the power spectrum likelihood,  $\sigma_{PS}(\hat{\beta})$ , to that from the field-level likelihood,  $\sigma_F(\hat{\beta})$ , as a function of cutoff for different redshifts. Each panel corresponds to a different halo mass range. From From [23].

2. The density field  $\delta_g$  is then smoothed using a Gaussian filter  $W_G(kR)$  with scale  $R = 1/\Lambda$ :

$$\delta_g(\mathbf{k}) \rightarrow W_G(k/\Lambda)\delta_g(\mathbf{k}). \quad (\text{E.5})$$

3. The estimated displacement field  $\boldsymbol{\psi}$  is obtained from the smoothed density field as:

$$\boldsymbol{\psi}(\mathbf{k}) = -i\frac{\mathbf{k}}{k^2}W_G(k/\Lambda)\frac{\delta_g(\mathbf{k})}{b_\delta}, \quad (\text{E.6})$$

where the bias parameter  $b_\delta$  is set to the true value used in generating the mock catalog.

4. The displacement field  $\boldsymbol{\psi}$  is interpolated to the tracer positions and used to shift them accordingly.
5. The displaced density field  $\delta_d$  is constructed by reassigning the shifted tracers to a grid of size  $2N^\Lambda$  using NUFFT. This step effectively cancels out a significant portion of the large-scale perturbations in  $\delta_g$ .
6. A uniform grid of particles is generated and displaced using  $\boldsymbol{\psi}$  to form the “shifted” field  $\delta_s$ . The assignment to the grid again follows the NUFFT method, with the same resolution  $2N^\Lambda$ .
7. The reconstructed density field is computed as:

$$\delta_g^{\text{rec}} = \delta_d - \delta_s. \quad (\text{E.7})$$

Here, subtracting  $\delta_s$  restores the large-scale modes removed during the displacement step.

8. Finally, the power spectrum of the reconstructed field,  $P_g^{\text{rec}}$ , is measured, imposing  $k_{\text{max}} = \Lambda$ .

Post-reconstruction, the BAO wiggles appear significantly sharper, highlighting the method’s success in reducing nonlinear damping.

We extract the BAO scale parameter  $\beta$  and its associated uncertainty from both pre- and post-reconstruction datasets by performing a Markov Chain Monte Carlo (MCMC) analysis with the `emcee` sampler [76]. In this process, we simultaneously constrain  $\beta$  and the parameters describing the broadband shape of the power spectrum using the following fitting template:

$$P_{\text{model}}(k, \beta) = (B_1 + B_2k^2)P_m(k, \beta) + A(k), \quad (\text{E.8})$$

where  $B_1$  and  $B_2$  are nuisance parameters, and

$$P_m(k, \beta) = P_{\text{fid}}(k)f^2(\beta, k) \quad (\text{E.9})$$

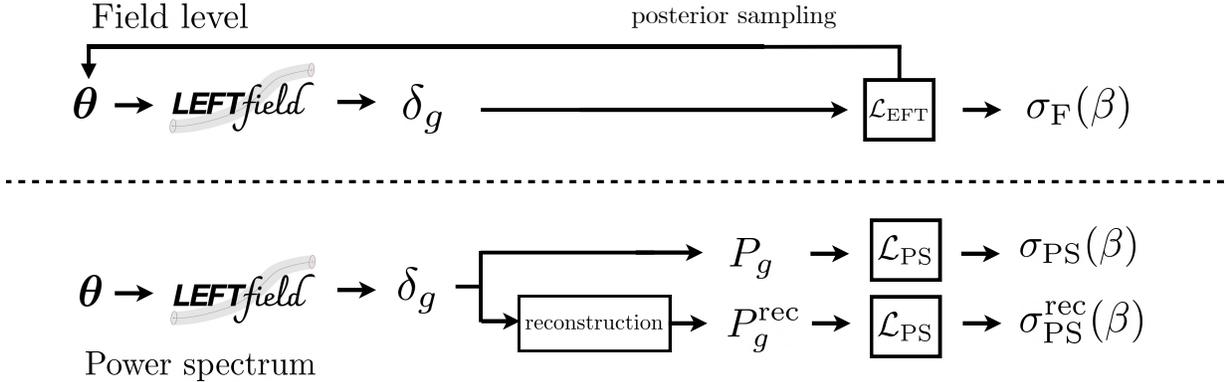


Figure E.2: Flowchart of the two methods for BAO inference employed in this work: field-level inference (top) and power spectrum-based inference with and without reconstruction (bottom). The reconstruction step includes the generation of a discrete tracer catalog from  $\delta_g$ , as described in the text. From [24].

denotes the linear power spectrum modified to include the rescaled BAO signal, as defined in Eq. (E.3). The term  $A(k)$  captures smooth contributions and is modeled by a cubic polynomial:

$$A(k) = a_0 + a_2 k^2 + a_3 k^3. \quad (\text{E.10})$$

We compute the power spectrum using linear binning with a spacing of  $\Delta k = 0.0028 h \text{ Mpc}^{-1}$ , and restrict the fit to the wavenumber range  $0.03 h \text{ Mpc}^{-1} < k < \Lambda$ .

We assume Gaussian likelihood and covariance for the power spectrum analysis.

## Results

In this work, we use 2LPT and second-order Lagrangian bias with operators and second order Eulerian bias, and Gaussian noise.

Two mocks were generated using bias parameters extracted from fixed-initial-condition analyses of halo catalogs, but at significantly higher resolution (or cutoff) than used in our inference procedure. This deliberate model mismatch aims to better mimic realistic scenarios. The essential difference between the two types of mocks lies in the bias model used to generate them: Mock A was built using second-order Lagrangian bias, while Mock B relied on second-order Eulerian bias.

For Mock A, we only used a second-order Lagrangian bias expansion during inference. The resulting systematic bias in  $\beta$  remained below 1% and was negligible across all  $\Lambda$  values, except at the highest cutoff, where it rose to roughly 1.8%. We observe that switching to a third-order bias expansion in inference likely mitigates this bias.

In the case of Mock B, based on Eulerian bias, we examined both second-order Eulerian and Lagrangian bias expansions in inference. The Eulerian approach yielded residual bias consistent with zero, with decreasing trend as  $\Lambda$  increased. Conversely, inference using

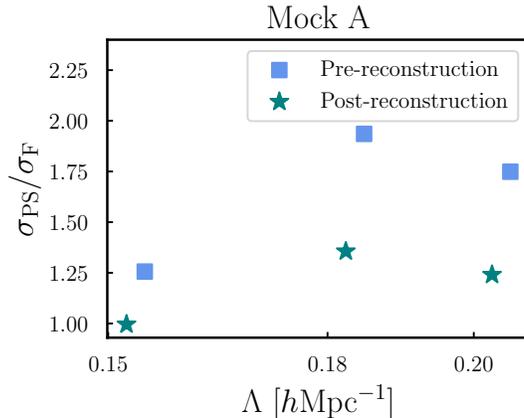


Figure E.3: Inferred error bar  $\sigma_{\text{PS}}(\beta)$  on the BAO scale using pre- and post-reconstruction power spectrum, relative to that in the field-level inference  $\sigma_{\text{F}}$ , in the case of Mock A. This mock was generated (and sampled, in case of field-level inference) using Lagrangian bias. Pre-reconstruction results are depicted using squares, whereas post-reconstruction results are represented by stars. It is evident that, even comparing to power spectrum after BAO reconstruction, the field-level BAO scale inference is more precise, by up to a factor of 1.35. From [24].

the Lagrangian model showed slightly elevated residual bias. Once again, increasing the order of the bias expansion is expected to reduce this effect. Additionally, the Lagrangian inference yielded narrower error bars compared to the Eulerian case, likely due to the additional filtering used when constructing Eulerian bias operators. This filtering omits certain mode-coupling contributions that are retained in the Lagrangian formulation.

Standard BAO inference typically involves reconstructing the density field from the tracer catalog and extracting the BAO scale from the post-reconstruction power spectrum. In Sec. E.2, we described the application of this reconstruction method to our mocks. As is standard practice, we fixed the linear bias parameter during reconstruction, favoring the reconstruction-based method over the field-level approach, where all bias parameters are inferred simultaneously with the BAO scale. Figures E.3 provide a comparison between the field-level error bars,  $\sigma_{\text{F}}$ , and those from the power spectrum,  $\sigma_{\text{PS}}$ . Depending on the mock and  $\Lambda$ ,  $\sigma_{\text{F}}$  can be up to twice as small (pre-reconstruction) or about 1.5 times smaller (post-reconstruction) than  $\sigma_{\text{PS}}$ . Nevertheless, field-level inference presents a more coherent framework, as it allows joint inference of all relevant parameters and avoids the need for fitting functions like Eq. (E.8) used in power spectrum analyses.

### E.3 Conclusions and next steps

In the future, we plan to infer the BAO scale via SBI and `LEFTfield`. This will allow for consistent marginalization over bias, noise, and cosmological parameters, instead of fixing

them as done in traditional analysis.

Moreover, combining BAO reconstruction with the full-shape of the galaxy power-spectrum has been shown to achieve even tighter cosmological constraints [170], but modeling the covariance and the reconstructed spectra remains challenging. Since we developed a reconstruction algorithm in `LEFTfield`, we plan to use SBI to do a joint analysis of pre- and post-reconstructed galaxy power-spectrum and bispectrum in order to obtain more cosmological information. SBI circumvents the issues of covariance modeling of these combined statistics, while `LEFTfield` naturally provides a model for the reconstructed spectra by means of measuring the n-point functions on the reconstructed field.

In this context, we would be pioneer to include the bispectrum into the analysis, besides analyzing how much information is transferred between pre- and post-reconstructed higher-order statistics, and comparing the information content to field-level inference.

# Bibliography

- [1] T.M.C. Abbott et al. , *Phys. Rev. D* **105** (2022), 023520.
- [2] T.M.C. Abbott et al. , *Open J. Astrophys.* **6** (2023), 2305.17173.
- [3] M.M. Abidi und T. Baldauf, *JCAP* **2018** (2018), 029.
- [4] A.G. Adame et al. , (2024).
- [5] A.G. Adame et al. , *JCAP* **02** (2025), 021.
- [6] R.J. Adler, B. Casey und O.C. Jacob, *American Journal of Physics* **63** (1995), 620.
- [7] N. Agarwal, V. Desjacques, D. Jeong und F. Schmidt, *JCAP* **03** (2021), 021.
- [8] A. Aghamousa et al. , (2016). ArXiv:1611.00036.
- [9] N. Aghanim et al. , *Astron. Astrophys.* **641** (2020), A6. [Erratum: *Astron. Astrophys.* 652, C4 (2021)].
- [10] S. Aiola et al. , *JCAP* **12** (2020), 047.
- [11] A. Akbar Abolhasani, M. Mirbabayi und E. Pajer, *JCAP* **2016** (2016), 063.
- [12] J. Akeret, A. Refregier, A. Amara, S. Seehars und C. Hasner, *JCAP* **2015** (2015), 043.
- [13] A. Akhmetzhanova, S. Mishra-Sharma und C. Dvorkin, *arXiv e-prints* (2023), arXiv:2308.09751.
- [14] T. Akiba, S. Sano, T. Yanase, T. Ohta und M. Koyama: *Optuna: A Next-generation Hyperparameter Optimization Framework. Optuna: A Next-generation Hyperparameter Optimization Framework*, In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* (2019).
- [15] J. Alsing, T. Charnock, S. Feeney und B. Wandelt, *MNRAS* **488** (2019), 4440.
- [16] J. Alsing und B. Wandelt, *MNRAS* **488** (2019), 5093.
- [17] J. Alsing, B. Wandelt und S. Feeney, *MNRAS* **477** (2018), 2874.

- [18] L. Amendola, S. Appleby, D. Bacon, T. Baker, M. Baldi, N. Bartolo, A. Blanchard, C. Bonvin, S. Borgani, E. Branchini, C. Burrage, S. Camera, C. Carbone, L. Casarini, M. Cropper, C. de Rham, C. Di Porto, A. Ealet, P.G. Ferreira, F. Finelli, J. García-Bellido, T. Giannantonio, L. Guzzo, A. Heavens, L. Heisenberg, C. Heymans, H. Hoekstra, L. Hollenstein, R. Holmes, O. Horst, K. Jahnke, T.D. Kitching, T. Koivisto, M. Kunz, G. La Vacca, M. March, E. Majerotto, K. Markovic, D. Marsh, F. Marulli, R. Massey, Y. Mellier, D.F. Mota, N.J. Nunes, W. Percival, V. Pettorino, C. Porciani, C. Quercellini, J. Read, M. Rinaldi, D. Sapone, R. Scaramella, C. Skordis, F. Simpson, A. Taylor, S. Thomas, R. Trotta, L. Verde, F. Vernizzi, A. Vollmer, Y. Wang, J. Weller und T. Zlosnik, *Living Reviews in Relativity* **16** (2013), 6.
- [19] A. Amon und G. Efstathiou, *Mon. Not. Roy. Astron. Soc.* (2022).
- [20] R.E. Angulo, M. Zennaro, S. Contreras, G. Aricò, M. Pellejero-Ibañez und J. Stücker, *MNRAS* **507** (2021), 5869.
- [21] V.I. Arnold: *Mathematical Methods of Classical Mechanics*. Graduate Texts in Mathematics. Springer, 1989.
- [22] V. Assassi, D. Baumann, D. Green und M. Zaldarriaga, *JCAP* **2014** (2014), 056.
- [23] I. Babić, F. Schmidt und B. Tucci, *JCAP* **2022** (2022), 007.
- [24] I. Babić, F. Schmidt und B. Tucci, *arXiv e-prints* (2024), arXiv:2407.01524.
- [25] T. Baldauf, L. Mercolli, M. Mirbabayi und E. Pajer, *JCAP* **2015** (2015), 007.
- [26] L. Balkenhol et al. , *Phys. Rev. D* **108** (2023), 023510.
- [27] D. Barber und F. Agakov: *The IM algorithm: a variational approach to Information Maximization*. *The IM algorithm: a variational approach to Information Maximization*, In *Proceedings of the 17th International Conference on Neural Information Processing Systems NIPS'03*. MIT Press, Cambridge, MA, USA (2003) Seite 201–208.
- [28] J.M. Bardeen, J.R. Bond, N. Kaiser und A.S. Szalay, *Astrophys. J.* **304** (1986), 15.
- [29] A.H. Barnett, J. Magland und L. af Klinteberg, *SIAM Journal on Scientific Computing* **41** (2019), C479.
- [30] A. Barreira, T. Lazeyras und F. Schmidt, *JCAP* **2021** (2021), 029.
- [31] D. Baumann, A. Nicolis, L. Senatore und M. Zaldarriaga, *JCAP* **2012** (2012), 051.
- [32] D. Baumann, A. Nicolis, L. Senatore und M. Zaldarriaga, *JCAP* **07** (2012), 051.
- [33] P.S. Behroozi, R.H. Wechsler und H.Y. Wu, *Astrophys. J.* **762** (2013), 109.
- [34] A.A. Berlind und D.H. Weinberg, *Astrophys. J.* **575** (2002), 587.

- [35] F. Bernardeau, M. Crocce und R. Scoccimarro, *PhRvD* **85** (2012), 123519.
- [36] R.C. Bernardo, D. Grandón, J. Levi Said und V.H. Cárdenas, *Physics of the Dark Universe* **40** (2023), 101213.
- [37] M. Betancourt. *A Conceptual Introduction to Hamiltonian Monte Carlo*, 2018.
- [38] Beyond-2pt Collaboration, E. Krause, Y. Kobayashi, A.N. Salcedo, M.M. Ivanov, T. Abel, K. Akitsu, R.E. Angulo, G. Cabass, S. Contarini, C. Cuesta-Lazaro, C. Hahn, N. Hamaus, D. Jeong, C. Modi, N.M. Nguyen, T. Nishimichi, E. Pailas, M. Pellejero Ibañez, O.H.E. Philcox, A. Pisani, F. Schmidt, S. Tanaka, G. Verza, S. Yuan und M. Zennaro, *arXiv e-prints* (2024), arXiv:2405.02252.
- [39] C.M. Bishop. *Mixture density networks* (1994).
- [40] A. Blanchard et al. , *Astron. Astrophys.* **642** (2020), A191.
- [41] L. Blot, M. Crocce, E. Sefusatti, M. Lippich, A.G. Sánchez, M. Colavincenzo, P. Monaco, M.A. Alvarez, A. Agrawal, S. Avila, A. Balaguera-Antolínez, R. Bond, S. Codis, C. Dalla Vecchia, A. Dorta, P. Fosalba, A. Izard, F.S. Kitaura, M. Pellejero-Ibanez, G. Stein, M. Vakili und G. Yepes, *MNRAS* **485** (2019), 2806.
- [42] V. Bogachev: *Measure Theory*. Nummer Bd. 1 in Measure Theory. Springer Berlin Heidelberg, 2007.
- [43] B. Bose et al. , (2023).
- [44] T. Buchert, *MNRAS* **254** (1992), 729.
- [45] G. Cabass, *JCAP* **01** (2021), 067.
- [46] G. Cabass und F. Schmidt, *JCAP* **04** (2020), 042.
- [47] G. Cabass und F. Schmidt, *JCAP* **07** (2020), 051.
- [48] G. Cabass, M. Simonović und M. Zaldarriaga, *Phys. Rev. D* **109** (2024), 043526.
- [49] J.J.M. Carrasco, M.P. Hertzberg und L. Senatore, *JHEP* **09** (2012), 082.
- [50] S.M. Carroll: *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press, 7 2019.
- [51] S.M. Carroll, S. Leichenauer und J. Pollack, *PhRvD* **90** (2014), 023518.
- [52] J.F. Chen, Y.C. Wang, T. Zhang und T.J. Zhang, *PhRvD* **107** (2023), 063517.
- [53] A. Chudaykin, M.M. Ivanov, O.H.E. Philcox und M. Simonović, *PhRvD* **102** (2020), 063533.

- [54] M. Colavincenzo, E. Sefusatti, P. Monaco, L. Blot, M. Crocce, M. Lippich, A.G. Sánchez, M.A. Alvarez, A. Agrawal, S. Avila, A. Balaguera-Antolínez, R. Bond, S. Codis, C. Dalla Vecchia, A. Dorta, P. Fosalba, A. Izard, F.S. Kitaura, M. Pellejero-Ibanez, G. Stein, M. Vakili und G. Yepes, *MNRAS* **482** (2019), 4883.
- [55] A. Cole, B.K. Miller, S.J. Witte, M.X. Cai, M.W. Grootes, F. Nattino und C. Weniger, *JCAP* **2022** (2022), 004.
- [56] W.R. Coulton, O.H.E. Philcox und F. Villaescusa-Navarro, *arXiv e-prints* (2024), arXiv:2406.15546.
- [57] K. Cranmer, J. Brehmer und G. Louppe, *Proceedings of the National Academy of Science* **117** (2020), 30055.
- [58] B. Dai und U. Seljak, *MNRAS* **516** (2022), 2363.
- [59] G. D’Amico, Y. Donath, M. Lewandowski, L. Senatore und P. Zhang, *arXiv e-prints* (2022), arXiv:2206.08327.
- [60] G. d’Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang, F. Beutler und H. Gil-Marín, *JCAP* **2020** (2020), 005.
- [61] G. D’Amico, L. Senatore und P. Zhang, *JCAP* **01** (2021), 006.
- [62] R. Daviet. *Inference with Hamiltonian Sequential Monte Carlo Simulators*, 2018.
- [63] K.S. Dawson, D.J. Schlegel, C.P. Ahn, S.F. Anderson, É. Aubourg, S. Bailey, R.H. Barkhouser, J.E. Bautista, A. Beifiori, A.A. Berlind, V. Bhardwaj, D. Bizyaev, C.H. Blake, M.R. Blanton, M. Blomqvist, A.S. Bolton, A. Borde, J. Bovy, W.N. Brandt, H. Brewington, J. Brinkmann, P.J. Brown, J.R. Brownstein, K. Bundy, N.G. Busca, W. Carithers, A.R. Carnero, M.A. Carr, Y. Chen, J. Comparat, N. Connolly, F. Cope, R.A.C. Croft, A.J. Cuesta, L.N. da Costa, J.R.A. Davenport, T. Delubac, R. de Putter, S. Dhital, A. Ealet, G.L. Ebelke, D.J. Eisenstein, S. Escoffier, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, R. Génova-Santos, J.E. Gunn, H. Guo, D. Haggard, P.B. Hall, J.C. Hamilton, B. Harris, D.W. Harris, S. Ho, D.W. Hogg, D. Holder, K. Honscheid, J. Huehnerhoff, B. Jordan, W.P. Jordan, G. Kauffmann, E.A. Kazin, D. Kirkby, M.A. Klaene, J.P. Kneib, J.M. Le Goff, K.G. Lee, D.C. Long, C.P. Loomis, B. Lundgren, R.H. Lupton, M.A.G. Maia, M. Makler, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, D. Margala, K.L. Masters, C.K. McBride, P. McDonald, I.D. McGreer, R.G. McMahon, O. Mena, J. Miralda-Escudé, A.D. Montero-Dorta, F. Montesano, D. Muna, A.D. Myers, T. Naugle, R.C. Nichol, P. Noterdaeme, S.E. Nuza, M.D. Olmstead, A. Oravetz, D.J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-Desabrouille, K. Pan, J.K. Parejko, I. Pâris, W.J. Percival, I. Pérez-Fournon, I. Pérez-Ràfols, P. Petitjean, R. Pfaffenberger, J. Pforr, M.M. Pieri, F. Prada, A.M. Price-Whelan, M.J. Raddick, R. Rebolo, J. Rich, G.T. Richards,

- C.M. Rockosi, N.A. Roe, A.J. Ross, N.P. Ross, G. Rossi, J.A. Rubiño-Martin, L. Samushia, A.G. Sánchez, C. Sayres, S.J. Schmidt, D.P. Schneider, C.G. Scóccola, H.J. Seo, A. Sheldon, E. Sheldon, Y. Shen, Y. Shu, A. Slosar, S.A. Smee, S.A. Snedden, F. Stauffer, O. Steele, M.A. Strauss, A. Streblyanska, N. Suzuki, M.E.C. Swanson, T. Tal, M. Tanaka, D. Thomas, J.L. Tinker, R. Tojeiro, C.A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, D.A. Wake, M. Watson, B.A. Weaver, D.H. Weinberg, B.J. Weiner, A.A. West, M. White, W.M. Wood-Vasey, C. Yeche, I. Zehavi, G.B. Zhao und Z. Zheng, *AJ* **145** (2013), 10.
- [64] N.S.M. de Santi, H. Shao, F. Villaescusa-Navarro, L.R. Abramo, R. Teyssier, P. Villanueva-Domingo, Y. Ni, D. Anglés-Alcázar, S. Genel, E. Hernandez-Martinez, U.P. Steinwandel, C.C. Lovell, K. Dolag, T. Castro und M. Vogelsberger, *arXiv e-prints* (2023), arXiv:2302.14101.
- [65] A. Delaunoy, J. Hermans, F. Rozet, A. Wehenkel und G. Louppe, *arXiv e-prints* (2022), arXiv:2208.13624.
- [66] V. Desjacques, D. Jeong und F. Schmidt, *PhysRep* **733** (2018), 1.
- [67] L. Dinh, J. Sohl-Dickstein und S. Bengio, *arXiv e-prints* (2016), arXiv:1605.08803.
- [68] N.A. Dmitriev und Y.B. Zel'dovich, *JETP* **18** (1964), 793.
- [69] S. Dodelson und F. Schmidt: *Modern Cosmology*. Academic Press, 2020.
- [70] C. Durkan, I. Murray und G. Papamakarios, *arXiv e-prints* (2020), arXiv:2002.03712.
- [71] J. Ehlers und T. Buchert, *General Relativity and Gravitation* **29** (1997), 733.
- [72] D.J. Eisenstein und W. Hu, *The Astrophysical Journal* **496** (1998), 605–614.
- [73] F. Elsner, F. Schmidt, J. Jasche, G. Lavaux und N.M. Nguyen, *JCAP* **01** (2020), 029.
- [74] Euclid Collaboration, V. Ajani, M. Baldi, A. Barthelemy, A. Boyle, P. Burger, V.F. Cardone, S. Cheng, S. Codis, C. Giocoli, J. Harnois-Déraps, S. Heydenreich, V. Kansal, M. Kilbinger, L. Linke, C. Llinares, N. Martinet, C. Parroni, A. Peel, S. Pires, L. Porth, I. Tereno, C. Uhlemann, M. Vicinanza, S. Vinciguerra, N. Aghanim, N. Aurichio, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, S. Camera, V. Capobianco, C. Carbone, J. Carretero, F.J. Castander, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, C.J. Conselice, L. Conversi, L. Corcione, F. Courbin, M. Cropper, A. Da Silva, H. Degaudenzi, A.M. Di Giorgio, J. Dinis, M. Douspis, F. Dubath, X. Dupac, S. Farrens, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, S. Galeotta, B. Garilli, B. Gillis, A. Grazian, F. Grupp, H. Hoekstra, W. Holmes, A. Hornstrup, P. Hudelot, K. Jahnke, M. Jhabvala,

- M. Kümmel, T. Kitching, M. Kunz, H. Kurki-Suonio, P.B. Lilje, I. Lloro, E. Maiorano, O. Mansutti, O. Marggraf, K. Markovic, F. Marulli, R. Massey, S. Mei, Y. Mellier, M. Meneghetti, M. Moresco, L. Moscardini, S.M. Niemi, J. Nightingale, T. Nutma, C. Padilla, S. Paltani, K. Pedersen, V. Pettorino, G. Polenta, M. Poncet, L.A. Popa, F. Raison, A. Renzi, J. Rhodes, G. Riccio, E. Romelli, M. Roncarelli, E. Rossetti, R. Saglia, D. Sapone, B. Sartoris, P. Schneider, T. Schrabback, A. Secroun, G. Seidel, S. Serrano, C. Sirignano, L. Stanco, J.L. Starck, P. Tallada-Crespí, A.N. Taylor, R. Toledo-Moreo, F. Torradeflot, I. Tutusaus, E.A. Valentijn, L. Valenziano, T. Vassallo, Y. Wang, J. Weller, G. Zamorani, J. Zoubian, S. Andreon, S. Bardelli, A. Boucaud, E. Bozzo, C. Colodro-Conde, D. Di Ferdinando, G. Fabbian, M. Farina, J. Graciá-Carpio, E. Keihänen, V. Lindholm, D. Maino, N. Mauri, C. Neissner, M. Schirmer, V. Scottez, E. Zucca, Y. Akrami, C. Baccigalupi, A. Balaguera-Antolínez, M. Ballardini, F. Bernardeau, A. Biviano, A. Blanchard, S. Borgani, A.S. Borlaff, C. Burigana, R. Cabanac, A. Cappi, C.S. Carvalho, S. Casas, G. Castignani, T. Castro, K.C. Chambers, A.R. Cooray, J. Coupon, H.M. Courtois, S. Davini, S. de la Torre, G. De Lucia, G. Desprez, H. Dole, J.A. Escartin, S. Escoffier, I. Ferrero, F. Finelli, K. Ganga, J. Garcia-Bellido, K. George, F. Giacomini, G. Gozaliasl, H. Hildebrandt, A. JIMENEZ MU\{N\}OZ, B. Joachimi, J.J.E. Kajava, C.C. Kirkpatrick, L. Legrand, A. Loureiro, M. Magliocchetti, R. Maoli, S. Marcin, M. Martinelli, C.J.A.P. Martins, S. Matthew, L. Maurin, R.B. Metcalf, P. Monaco, G. Morgante, S. Nadathur, A.A. Nucita, V. Popa, D. Potter, A. Pourtsidou, M. Pöntinen, P. Reimberg, A.G. Sánchez, Z. Sakr, A. Schneider, E. Seifusatti, M. Sereno, A. Shulevski, A. Spurio Mancini, J. Steinwagner, R. Teyssier, J. Valiviita, A. Veropalumbo, M. Viel und I.A. Zinchenko, *arXiv e-prints* (2023), arXiv:2301.12890.
- [75] J. Fluri, T. Kacprzak, A. Refregier, A. Lucchi und T. Hofmann, *PhRvD* **104** (2021), 123526.
- [76] D. Foreman-Mackey, D.W. Hogg, D. Lang und J. Goodman, *Publications of the Astronomical Society of the Pacific* **125** (2013), 306–312.
- [77] J.A. Frieman und E. Gaztanaga, *ApJ* **425** (1994), 392.
- [78] J.N. Fry, *Phys. Rev. Lett.* **73** (1994), 215.
- [79] S. Gagnon-Hartman, J. Ruan und D. Haggard, *MNRAS* **520** (2023), 1.
- [80] F. Gerardi, S.M. Feeney und J. Alsing, *PhRvD* **104** (2021), 083531.
- [81] M. Germain, K. Gregor, I. Murray und H. Larochele, *arXiv e-prints* (2015), arXiv:1502.03509.
- [82] S. Goldstein, M. Park, M. Raveri, B. Jain und L. Samushia, *Phys. Rev. D* **107** (2023), 063530.

- [83] S. Goldstein, O.H.E. Philcox, J.C. Hill und L. Hui, *arXiv e-prints* (2024), arXiv:2407.08731.
- [84] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville und Y. Bengio, *arXiv e-prints* (2014), arXiv:1406.2661.
- [85] D.S. Greenberg, M. Nonnenmacher und J.H. Macke, *arXiv e-prints* (2019), arXiv:1905.07488.
- [86] D. Gualdi, H. Gil-Marín und L. Verde, *JCAP* **2021** (2021), 008.
- [87] D. Gualdi, S. Novell, H. Gil-Marín und L. Verde, *JCAP* **2021** (2021), 015.
- [88] D. Gualdi und L. Verde, *JCAP* **2022** (2022), 050.
- [89] C. Hahn, F. Beutler, M. Sinha, A. Berlind, S. Ho und D.W. Hogg, *MNRAS* **485** (2019), 2956.
- [90] C. Hahn, M. Eickenberg, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. Régaldo-Saint Blancard und M.M. Abidi, *arXiv e-prints* (2022), arXiv:2211.00723.
- [91] C. Hahn, M. Eickenberg, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. Régaldo-Saint Blancard und M.M. Abidi, *JCAP* **2023** (2023), 010.
- [92] N. Hamaus, U. Seljak, V. Desjacques, R.E. Smith und T. Baldauf, *PhRvD* **82** (2010), 043515.
- [93] A. Heavens, R. Jimenez und O. Lahav, *Mon. Not. Roy. Astron. Soc.* **317** (2000), 965.
- [94] A. Heavens, E. Sellentin und A. Jaffe, *Mon. Not. Roy. Astron. Soc.* **498** (2020), 3440.
- [95] J. Hermans, V. Begy und G. Louppe, *arXiv e-prints* (2019), arXiv:1903.04057.
- [96] J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, V. Begy und G. Louppe, *arXiv e-prints* (2021), arXiv:2110.06581.
- [97] C. Heymans et al. , *Astron. Astrophys.* **646** (2021), A140.
- [98] C.M. Hirata, *MNRAS* **399** (2009), 1074.
- [99] C.W. Huang, D. Krueger, A. Lacoste und A. Courville, *arXiv e-prints* (2018), arXiv:1804.00779.
- [100] H.Y. Ip und F. Schmidt, *JCAP* **2017** (2017), 025.
- [101] T. Ishiyama et al. , *Mon. Not. Roy. Astron. Soc.* **506** (2021), 4210.

- [102] M.M. Ivanov, O.H.E. Philcox, G. Cabass, T. Nishimichi, M. Simonović und M. Zaldarriaga, *PhRvD* **107** (2023), 083515.
- [103] M.M. Ivanov, O.H.E. Philcox, T. Nishimichi, M. Simonović, M. Takada und M. Zaldarriaga, *PhRvD* **105** (2022), 063512.
- [104] M.M. Ivanov, M. Simonović und M. Zaldarriaga, *JCAP* **2020** (2020), 042.
- [105] J. Jasche und B.D. Wandelt, *MNRAS* **432** (2013), 894.
- [106] N. Jeffrey, J. Alsing und F. Lanusse, *MNRAS* **501** (2021), 954.
- [107] D. Jeong und E. Komatsu, *ApJ* **703** (2009), 1230.
- [108] G. Jung, A. Ravenni, M. Baldi, W.R. Coulton, D. Jamieson, D. Karagiannis, M. Liguori, H. Shao, L. Verde, F. Villaescusa-Navarro und B.D. Wandelt, *ApJ* **957** (2023), 50.
- [109] T. Kacprzak und J. Fluri, *Physical Review X* **12** (2022), 031029.
- [110] N. Kaiser, *ApJL* **284** (1984), L9.
- [111] K. Karchev, R. Trotta und C. Weniger, *MNRAS* **520** (2023), 1056.
- [112] D.P. Kingma und J. Ba, *arXiv e-prints* (2014), arXiv:1412.6980.
- [113] D.P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever und M. Welling, *arXiv e-prints* (2016), arXiv:1606.04934.
- [114] D.P. Kingma und M. Welling, *arXiv e-prints* (2013), arXiv:1312.6114.
- [115] F.S. Kitaura, M. Ata, S.A. Rodríguez-Torres, M. Hernández-Sánchez, A. Balaguera-Antolínez und G. Yepes, *MNRAS* **502** (2021), 3456.
- [116] N. Kokron, J. DeRose, S.F. Chen, M. White und R.H. Wechsler, *MNRAS* **505** (2021), 1422.
- [117] E. Komatsu und D.N. Spergel, *PhRvD* **63** (2001), 063002.
- [118] A. Kostić, N.M. Nguyen, F. Schmidt und M. Reinecke, *arXiv e-prints* (2022), arXiv:2212.07875.
- [119] J.U. Lange, A.P. Hearin, A. Leauthaud, F.C. van den Bosch, E. Khakaj, H. Guo, R.H. Wechsler und J. DeRose, *Mon. Not. Roy. Astron. Soc.* (2023).
- [120] G. Lavaux, J. Jasche und F. Leclercq, *arXiv e-prints* (2019), arXiv:1909.06396.
- [121] T. Lazeyras, A. Barreira und F. Schmidt, *JCAP* **2021** (2021), 063.

- [122] T. Lazeyras und F. Schmidt, *JCAP* **2019** (2019), 041.
- [123] T. Lazeyras, C. Wagner, T. Baldauf und F. Schmidt, *JCAP* **2016** (2016), 018.
- [124] F. Leclercq, *PhRvD* **98** (2018), 063511.
- [125] F. Leclercq und A. Heavens, *MNRAS* **506** (2021), L85.
- [126] F. Leclercq und A. Heavens, *MNRAS* **506** (2021), L85.
- [127] P. Lemos, A. Coogan, Y. Hezaveh und L. Perreault-Levasseur, *arXiv e-prints* (2023), arXiv:2302.03026.
- [128] P. Lemos, M. Cranmer, M. Abidi, C. Hahn, M. Eickenberg, E. Massara, D. Yallup und S. Ho, *Machine Learning: Science and Technology* **4** (2023), 01LT01.
- [129] P. Lemos, L.H. Parker, C. Hahn, S. Ho, M. Eickenberg, J. Hou, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. Régaldó-Saint Blancard und D. Spergel, (2023), 18.
- [130] J. Lesgourgues. *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, 2011.
- [131] K. Lin, M. von Wietersheim-Kramsta, B. Joachimi und S. Feeney, *arXiv e-prints* (2022), arXiv:2212.04521.
- [132] M.X. Lin, B. Jain, M. Raveri, E.J. Baxter, C. Chang, M. Gatti, S. Lee und J. Muir, (2023).
- [133] E.V. Linder, *Phys. Rev.* **D72** (2005), 043529.
- [134] E.V. Linder, *PhRvD* **79** (2009), 063519.
- [135] E.V. Linder und R.N. Cahn, *Astropart. Phys.* **28** (2007), 481.
- [136] J. Linhart, A. Gramfort und P.L.C. Rodrigues, *arXiv e-prints* (2022), arXiv:2211.09602.
- [137] J.M. Lueckmann, J. Boelts, D.S. Greenberg, P.J. Gonçalves und J.H. Macke, *arXiv e-prints* (2021), arXiv:2101.04653.
- [138] J.M. Lueckmann, P.J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher und J.H. Macke, *arXiv e-prints* (2017), arXiv:1711.01861.
- [139] M.S. Madhavacheril et al. , *Astrophys. J.* **962** (2024), 113.
- [140] T.L. Makinen, T. Charnock, J. Alsing und B.D. Wandelt, *JCAP* **2021** (2021), 049.
- [141] T.L. Makinen, T. Charnock, P. Lemos, N. Porqueres, A.F. Heavens und B.D. Wandelt, *The Open Journal of Astrophysics* **5** (2022), 18.

- [142] S. Matarrese, L. Verde und A.F. Heavens, *MNRAS* **290** (1997), 651.
- [143] T. Matsubara, *PhRvD* **92** (2015), 023534.
- [144] P. McDonald, *PhRvD* **75** (2007), 043514.
- [145] B. Miller, A. Cole, P. Forré, G. Louppe und C. Weniger, *Advances in Neural Information Processing Systems* **34** (2021), 129.
- [146] B.K. Miller, C. Weniger und P. Forré, *arXiv e-prints* (2022), arXiv:2210.06170.
- [147] J. Milnor und D. Weaver: *Topology from the Differentiable Viewpoint*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1997.
- [148] M. Mirbabayi, F. Schmidt und M. Zaldarriaga, *JCAP* **2015** (2015), 030.
- [149] C. Modi, S. Pandey, M. Ho, C. Hahn, B.R.S. Blancard und B. Wandelt, (2023).
- [150] C. Modi und O.H.E. Philcox, *arXiv e-prints* (2023), arXiv:2309.10270.
- [151] R. Neal. In *Handbook of Markov Chain Monte Carlo* (2011), Seiten 113–162.
- [152] R.M. Neal, *arXiv e-prints* (2000), physics/0009028.
- [153] R.M. Neal: *Handbook of Markov Chain Monte Carlo*, 5 2011.
- [154] D. Nelson et al. , *Comput. Astrophys. Cosmol.* **6** (2019), 2.
- [155] N.M. Nguyen, D. Huterer und Y. Wen, *Phys. Rev. Lett.* **131** (2023), 111001.
- [156] N.M. Nguyen, F. Schmidt, G. Lavaux und J. Jasche, *JCAP* **2021** (2021), 058.
- [157] N.M. Nguyen, F. Schmidt, B. Tucci, M. Reinecke und A. Kostić, *Phys. Rev. Lett.* **133** (2024), 221006.
- [158] C. Nikolis, H. Rubira und F. Schmidt, *JCAP* **08** (2024), 017.
- [159] A. Oddo, E. Sefusatti, C. Porciani, P. Monaco und A.G. Sánchez, *JCAP* **2020** (2020), 056.
- [160] R. Pakmor et al. , *Mon. Not. Roy. Astron. Soc.* **524** (2023), 2539.
- [161] G. Papamakarios, *arXiv e-prints* (2019), arXiv:1910.13233.
- [162] G. Papamakarios und I. Murray, *arXiv e-prints* (2016), arXiv:1605.06376.
- [163] G. Papamakarios, E. Nalisnick, D. Jimenez Rezende, S. Mohamed und B. Lakshminarayanan, *arXiv e-prints* (2019), arXiv:1912.02762.

- [164] G. Papamakarios, T. Pavlakou und I. Murray, *arXiv e-prints* (2017), arXiv:1705.07057.
- [165] G. Papamakarios, D.C. Sterratt und I. Murray, *arXiv e-prints* (2018), arXiv:1805.07226.
- [166] C.F. Park, E. Allys, F. Villaescusa-Navarro und D. Finkbeiner, *ApJ* **946** (2023), 107.
- [167] M.E. Peskin und D.V. Schroeder: *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- [168] O.H.E. Philcox und M.M. Ivanov, *PhRvD* **105** (2022), 043517.
- [169] O.H.E. Philcox, M.M. Ivanov, G. Cabass, M. Simonović, M. Zaldarriaga und T. Nishimichi, *PhRvD* **106** (2022), 043530.
- [170] O.H.E. Philcox, M.M. Ivanov, M. Simonović und M. Zaldarriaga, *JCAP* **2020** (2020), 032.
- [171] Planck Collaboration et al. , *A&A* **594** (2016), A16.
- [172] J. Polchinski, *Nucl. Phys. B* **231** (1984), 269.
- [173] N. Porqueres, A. Heavens, D. Mortlock, G. Lavaux und T.L. Makinen, *arXiv e-prints* (2023), arXiv:2304.04785.
- [174] N. Porqueres, D. Kodi Ramanah, J. Jasche und G. Lavaux, *A&A* **624** (2019), A115.
- [175] R.A. Porto, *PhysRep* **633** (2016), 1.
- [176] V. Poulin, J.L. Bernal, E.D. Kovetz und M. Kamionkowski, *Phys. Rev. D* **107** (2023), 123538.
- [177] D. Prangle, *arXiv e-prints* (2015), arXiv:1507.00874.
- [178] C. Preston, A. Amon und G. Efstathiou, *Mon. Not. Roy. Astron. Soc.* **525** (2023), 5554.
- [179] J.K. Pritchard, M.T. Seielstad, A. Perez-Lezaun und M.W. Feldman, *Molecular Biology and Evolution* **16** (1999), 1791.
- [180] C. Rampf, *JCAP* **2012** (2012), 004.
- [181] C. Rampf und T. Buchert, *JCAP* **2012** (2012), 021.
- [182] A. Reeves, A. Nicola, A. Refregier, T. Kacprzak und L.F.M.P. Valle, *JCAP* **01** (2024), 042.

- [183] M. Reza, Y. Zhang, B. Nord, J. Poh, A. Ciprijanovic und L. Strigari: *Estimating Cosmological Constraints from Galaxy Cluster Abundance using Simulation-Based Inference. Estimating Cosmological Constraints from Galaxy Cluster Abundance using Simulation-Based Inference*, In *Machine Learning for Astrophysics*. (Juli 2022) Seite 20.
- [184] K.K. Rogers, R. Hložek, A. Laguë, M.M. Ivanov, O.H.E. Philcox, G. Cabass, K. Akitsu und D.J.E. Marsh, *JCAP* **06** (2023), 023.
- [185] H. Rubira und F. Schmidt, (2023).
- [186] H. Rubira und F. Schmidt, *JCAP* **01** (2024), 031.
- [187] H. Rubira und F. Schmidt, *JCAP* **10** (2024), 092.
- [188] W. Rudin: *Real and Complex Analysis*. McGraw-Hill series in higher mathematics. Tata McGraw-Hill, 2006.
- [189] F. Schmidt, *PhRvD* **93** (2016), 063512.
- [190] F. Schmidt, *JCAP* **2021** (2021), 033.
- [191] F. Schmidt, *JCAP* **04** (2021), 032.
- [192] F. Schmidt, G. Cabass, J. Jasche und G. Lavaux, *JCAP* **2020** (2020), 008.
- [193] F. Schmidt, F. Elsner, J. Jasche, N.M. Nguyen und G. Lavaux, *JCAP* **01** (2019), 042.
- [194] M. Schmittfull, M. Simonović, V. Assassi und M. Zaldarriaga, *PhRvD* **100** (2019), 043514.
- [195] R. Scoccimarro, S. Colombi, J.N. Fry, J.A. Frieman, E. Hivon und A. Melott, *ApJ* **496** (1998), 586.
- [196] E. Sefusatti, M. Crocce, S. Pueblas und R. Scoccimarro, *PhRvD* **74** (2006), 023522.
- [197] E. Sellentin und A.F. Heavens, *MNRAS* **473** (2018), 2355.
- [198] E. Sellentin, C. Heymans und J. Harnois-Déraps, *MNRAS* **477** (2018), 4879.
- [199] L. Senatore, *JCAP* **2015** (2015), 007.
- [200] H. Shao, F. Villaescusa-Navarro, P. Villanueva-Domingo, R. Teyssier, L.H. Garrison, M. Gatti, D. Inman, Y. Ni, U.P. Steinwandel, M. Kulkarni, E. Visbal, G.L. Bryan, D. Anglés-Alcázar, T. Castro, E. Hernández-Martínez und K. Dolag, *ApJ* **944** (2023), 27.
- [201] D. Sharma, B. Dai und U. Seljak, *JCAP* **08** (2024), 010.

- 
- [202] I. Sobol', *USSR Computational Mathematics and Mathematical Physics* **7** (1967), 86.
- [203] V. Springel, R. Pakmor, O. Zier und M. Reinecke, *Mon. Not. Roy. Astron. Soc.* **506** (2021), 2871.
- [204] M. Srednicki: *Quantum field theory*. Cambridge University Press, 1 2007.
- [205] J. Stadler, F. Schmidt und M. Reinecke, *arXiv e-prints* (2023), arXiv:2303.09876.
- [206] J. Stadler, F. Schmidt und M. Reinecke, (2024).
- [207] J. Stadler, F. Schmidt, M. Reinecke und M. Esposito, (2024).
- [208] S. Sugiyama et al. , *Phys. Rev. D* **108** (2023), 123521.
- [209] M. Takada et al. , *PASJ* **66** (2014), R1.
- [210] S. Talts, M. Betancourt, D. Simpson, A. Vehtari und A. Gelman, *arXiv e-prints* (2018), arXiv:1804.06788.
- [211] A. Tejero-Cantero, J. Boelts, M. Deistler, J.M. Lueckmann, C. Durkan, P. Gonçalves, D. Greenberg und J. Macke, *The Journal of Open Source Software* **5** (2020), 2505.
- [212] L. Thiele, E. Massara, A. Pisani, C. Hahn, D.N. Spergel, S. Ho und B. Wandelt, *arXiv e-prints* (2023), arXiv:2307.07555.
- [213] J. Tinker, A.V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber und D.E. Holz, *ApJ* **688** (2008), 709.
- [214] J.L. Tinker, B.E. Robertson, A.V. Kravtsov, A. Klypin, M.S. Warren, G. Yepes und S. Gottlöber, *ApJ* **724** (2010), 878.
- [215] R. Trotta, *Contemporary Physics* **49** (2008), 71.
- [216] E. Tsaprazi, J. Jasche, G. Lavaux und F. Leclercq, *arXiv e-prints* (2023), arXiv:2301.03581.
- [217] B. Tucci und F. Schmidt, *JCAP* **2024** (2024), 063.
- [218] B. Uria, M.A. Côté, K. Gregor, I. Murray und H. Larochele, *arXiv e-prints* (2016), arXiv:1605.02226.
- [219] A. Vale und J.P. Ostriker, *MNRAS* **353** (2004), 189.
- [220] L. Verde. In *Lecture Notes in Physics, Berlin Springer Verlag*, herausgegeben von G. Wolschin, Band 800 (2010), Seiten 147–177.
- [221] L. Verde und A.F. Heavens, *ApJ* **553** (2001), 14.

- 
- [222] F. Villaescusa-Navarro et al. , *Astrophys. J.* **915** (2021), 71.
- [223] P. Villanueva-Domingo und F. Villaescusa-Navarro, *ApJ* **937** (2022), 115.
- [224] H. Wang, H.J. Mo, X. Yang, Y.P. Jing und W.P. Lin, *ApJ* **794** (2014), 94.
- [225] Y. Wen, N.M. Nguyen und D. Huterer, *JCAP* **09** (2023), 028.
- [226] K.G. Wilson, *Phys. Rev. B* **4** (1971), 3174.
- [227] S. Yuan, B. Hadzhiyska und T. Abel, *MNRAS* **520** (2023), 6283.
- [228] M. Zennaro, R.E. Angulo, M. Pellejero-Ibáñez, J. Stücker, S. Contreras und G. Aricò. *The BACCO simulation project: biased tracers in real space*, Januar 2021.
- [229] P. Zhang, G. D’Amico, L. Senatore, C. Zhao und Y. Cai, *JCAP* **2022** (2022), 036.
- [230] D. Zhao, N. Dalmasso, R. Izbicki und A.B. Lee, *arXiv e-prints* (2021), arXiv:2102.10473.
- [231] V. Zheligovsky und U. Frisch, *Journal of Fluid Mechanics* **749** (2014), 404.
- [232] Z. Zheng, R. Cen, H. Trac und J. Miralda-Escudé, *ApJ* **726** (2011), 38.

# Danksagung

First, I would like to thank my supervisor, Fabian Schmidt, for his continuous support throughout my PhD. The work presented here was developed with his invaluable guidance. I am truly grateful for everything I have learned from our discussions and will carry those lessons with me. Eiichiro Komatsu, thank you for the inspiring conversations on so many different physics topics. Every time we talk, I come away having learned a lot.

I would also like to thank all my friends with whom I have shared unforgettable moments during my PhD. A special thank you to Ivana Babić, my professional manager and the one who always gives me the best advice. I am especially grateful to Ivana Nikolac and Şafak Çelik, who have been by my side not only throughout the PhD, but also during the intense process of writing this thesis. Ivana, thank you for your immeasurable support in my day-to-day life. Şafak, thank you for making my afternoons at MPA full of thoughtful discussions. Andrija Kostić, thank you for the excitement of learning cosmology together. Minh Nguyen, thank you for the collaborative times we have.

The many conversations I had at MPA greatly enriched my research. I would like to thank Luisa Lucie-Smith for her insightful advice on both research and career. I also appreciated the coding sessions with Julia Stadler, the enlightening discussions on the trispectrum with Drew Jamieson, and the inspiring exchanges on simulation-based inference with Noemi Anau Montel, Adri Duivenvoorden, and Jiamin Hou, and EFTofLSS discussions with Henrique Rubira. I am grateful for the EFTofLSS group meetings with Sha Azyzy and Jeongin Moon, and for the exciting LEFTfield meetings and hack/cake weeks with my co-workers.

I am deeply thankful that my path at MPA crossed with Catarina Aydar and Franciele Kruczkiewicz. To my friends across the ocean, Sam Goldstein, thank you for all the long physics discussions, and Noah Sailer, who is always up for teaching me a bit of CMB. Thank you also Nick Kokron and Gabi Satie for the unvaluable support during my postdoc applications. I am also thankful for the insightful conference conversations that shaped parts of this work, particularly with Oliver Philcox and ChangHoon Hahn.

Lastly, I want to thank my family, who remained in Brazil while I moved to Germany to pursue this PhD. Lisange, who is always there for me, and Clara – thank you for your being my sister, exactly the way you are. Yuri, thank you for encouraging my pursuit of science. Livio, I wish I could tell you all about my research. I am sure you would appreciate it.