Bayesian forward modeling of Galaxy Clustering
Cosmology and Astrophysics inference

Nhat-Minh Nguyen

München 2020

Bayesian forward modeling of Galaxy Clustering
Cosmology and Astrophysics inference

Nhat-Minh Nguyen

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universit"at
München

vorgelegt von
Nhat-Minh Nguyen
aus Saigon, Vietnam

München, den 28. Apr 2020

Erstgutachter: Prof. Dr. Eiichiro Komatsu
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 16. Juni 2020

Contents

List of Figures

List of Tables

Zusammenfassung

Die großräumige Struktur des Universums (large-scale structure, LSS) erlaubt uns, dank neuer, größerer Himmelsdurchmusterungen, immer präzisere Vermessungen der Geschichte des Universums und Tests von kosmologischen Modellen. Die immer größeren und genaueren Datensätze stellen uns allerdings auch vor Herausforderungen in der Analyse und insbesondere Kombination von verschiedenen Datensätzen. Vorwärts-Modellierung (forward modeling) und Inferenz im Rahmen der Bayes'schen Statistik bieten für diesen Zweck einen konsistenten Rahmen, der auch die Einbeziehung von systematischen Effekten erlaubt.

Die vorliegende Dissertation zielt darauf ab, dieses statistische Modell und seine Implementierung im BORG-Code einen Schritt näher an die tatsächliche Anwendung auf aktuelle und zukünftige Datensätze zu bringen, insbesondere für den kosmichen Mikrowellenhintergrund (cosmic microwave background, CMB) und für spektroskopische Galaxiendurchmusterungen (galaxy redshift surveys). Die Inferenz von sowohl kosmologischen als auch astrophysikalischen Parametern wird systematisch untersucht: erstens die Anfangsbedingungen für die Struktur im Universum, das heisst die Verteilung der kleinen Dichtefluktuationen im frühen Universum, und zweitens der kinematische Sunyaev-Zel'dovich (kSZ) Effekt im CMB, der unter anderem von Galaxienhaufen hervorgerufen wird. Im ersteren Fall werden kosmologische N-Körper-Simulationen benutzt, während der kSZ-Effekt auf tatsächlichen Beobachtungsdaten gemessen wird: die CMB-Messungen des Planck-Satelliten, die Galaxienverteilung der Sloan Digital Sky Survey (SDSS), und der Galaxienhaufenkatalog maxBCG. Entscheidend sind hier eine sorgfältige Behandlung der spezifischen und gemeinsamen systematischen Effekte der verschiedenen Datensätze.

In Kapitel 5 zeige ich, dass das Bayes'sche Vorwärts-Modell im BORG-Code in der Lage ist, die Anfangsbedingungen zu ca. 90% korrekt zu rekonstruieren. Dies gilt auf großen Skalen für verschiedene Vorwärts-Modelle für das Dichtefeld sowie für den Galaxienbias. Was kosmologische Parameter angeht, wie die Amplitude des linearen Dichtefeldes σ_8, spielt die Form der Wahrscheinlichkeitsfunktion (likelihood) für das Galaxienfeld eine wichtige Rolle. Dies wird in Kapitel 7 demonstriert, wo ich zeige, dass eine Wahrscheinlichkeitsfunktion im Fourier-Raum eine Inferenz des Parameters σ_8 innerhalb eines systematischen Fehlers von 10% erlaubt. Diese Wahrscheinlichkeitsfunktion kann im Rahmen der effektiven Feldtheorie (effective field theory, EFT) für LSS hergeleitet werden.

In Kapitel 6 benutze ich dann die BORG-Rekonstruktion basierend auf der SDSS/BOSS Galaxiendurchmusterung, um den kSZ-Effekt um Galaxienhaufen aus dem maxBCG-Katalog (der das gleiche Volumen umfasst) zu messen. In diesem Fall benutze ich die Rekonstruk-

tion, um die Pekuliargeschwindigkeiten der Galaxienhaufen zu schätzen, die nicht direkt beobachtbar sind. Ich finde Hinweise auf ein kSZ-Signal mit etwa 2σ Konfidenz, sowohl in der Messung auf bestimmten Winkelskalen wie auch in der Kombination von Messungen auf verschiedenen Skalen. Diese Messung und ihr Fehlerbalken sind die ersten kSZ-Messungen, die die Unsicherheiten in der Schätzung der Geschwindigkeiten berücksichtigen.

Summary

With the future large-scale structure (LSS) surveys being on the horizon, precision cosmology is seeing a unprecedented opportunity to constrain cosmological parameters and differentiate cosmological models. Such opportunities naturally bring also unparalleled challenges – specifically in the form of understanding, examining and, especially, combining various datasets. Bayesian forward modeling and inference, in this context, provides a consistent and transparent framework to extract information from separate datasets while accounting for multiple systematic sources.

This thesis is a dedicated effort to bring this framework one step closer to being ready for the upcoming challenges posed by high-precision Cosmic Microwave Background (CMB) experiments and large-volume galaxy redshift surveys. We systematically examine the constraining power of the Bayesian forward modeling approach to galaxy clustering on both cosmological and astrophysical observables, namely the initial conditions of our Universe, the clustering amplitude of galaxies and the kinematic Sunyaev-Zel'dovich (kSZ) effects of galaxy clusters. While the first two focus only on halo clustering in N-body simulation, the last one brings together observational datasets from separate experiments and surveys: the Planck CMB experiment, the Sloan Digital Sky survey (SDSS) and the maxBCG cluster catalog, which include both common and different sources of systematics.

We find in Chapter 5 that the Bayesian forward inference approach is able to, on large scales, recover up to $\simeq 90\%$ the input initial conditions of the GADGET-2 simulation using halos identified in the same simulation as tracers. The framework is robust regarding to choices of gravitational forward model for the matter density fields and deterministic bias model for tracers. The LSS likelihood, on the other hand, might play an important role for unbiased inference of not only the initial conditions, but also the cosmological parameters. This is demonstrated in Chapter 7, where we are able to recover the input σ_8 of the same simulation with systematic error under $\simeq 10\%$, using a Fourier-space likelihood derived from the effective field theory (EFT) approach to LSS with rigorously controlled theoretical systematics.

In Chapter 6, we use results from the Bayesian forward reconstruction of the BOSS/SDSS3 volume to measure the large-scale bulk flow and kSZ signal of maxBCG catalog. We find evidence of the kSZ effect at $\simeq 2\sigma$, consistently in individual- as well as multi-scale measurements. Our reported signal-to-noise is the first to include uncertainties from the velocity reconstruction in this type of measurement.

Chapter 1

Introduction

1.1 Overview

The advent of galaxy and cluster of galaxies surveys [1, 2, 3, 4, 5, 6, 7] brought cosmology great power to explore the large-scale structure (LSS) of our Universe and to trace back the history of its formation. Through the now well-established method of summary statistics, i.e. the use of two-point *correlation function* – or its Fourier-space counterpart, the *power spectrum* [8, 9] – high signal-to-noise information was extracted from these surveys which, for the first time, rigorously confronted our theoretical models of the Universe [10, 11, 12, 13]. A prime example of this came in the early 1990s. As mismatches emerged between the two-point angular correlation function of matter predicted by the then favorite flat CDM-dominated Einstein–de Sitter (EdS) Universe model and that of APM galaxies [7, 14], given constraints from COBE measurement of the Cosmic Microwave Background (CMB) [15], the cosmological constant Λ naturally arose as a missing piece to resolve the discrepancies and eventually became a key ingredient in the now standard picture of our Universe – the flat ΛCDM model.

With that early success, it comes as no surprise that recent and upcoming generations of galaxy redshift surveys – e.g. SDSS-III BOSS [16], SDSS-IV eBOSS [17] and DESI [18], Euclid [19], LSST [20], PFS [21] – or those of CMB experiments – e.g. WMAP [22], Planck [23, 24, 25], SPTPol [26], ACTPol [27] and AdvACT [28], CMB-S4 [29] – have been reaching deeper and wider, offering data not only in remarkably larger volumes but also at significantly higher precisions. Indeed, the focus in studies of cosmology from LSS has shifted to galaxy clustering on smaller, *quasi-linear* scales[1] [30, 31] and its correlation with matter clustering – as traced by weak lensing measurements – on these scales [32, 33, 34]. As the clustering of matter on these quasi-linear scales has been extensively processed by gravity as the Universe evolves, it also encodes rich information about gravity, initial matter

[1] In this thesis, we refer to mildly non-linear scales where perturbation theory still converges to correct results when carried out to a sufficiently high order as *quasi-linear*. Readers should note that this regime is referred to also as *quasi-nonlinear* by some authors. On the other hand, we will refer to scales at which higher-order correction terms are no longer smaller than lower-order terms and perturbation theory fails to converge as *non-linear*.

perturbations and content of our Universe, especially at late time. Extracting this wealth of information (without biasing the constraints they impose on cosmological parameters) requires great responsibility in modeling the complex datasets and their cross-correlations, especially in consistently propagating all systematics. In addition, although the initial seeds of matter perturbations are highly *Gaussian*, as predicted by theory due to the nature of quantum fluctuations during inflation and repeatedly confirmed by the observation of the CMB anisotropies [35, 22, 36], the evolved galaxy and matter distributions at late time, are far from Gaussian on quasi-linear and nonlinear scales. Thus, as noted by [37, 38], even an infinite hierarchy of higher-order n-point correlation functions could prove insufficient to completely capture all the information encoded in such distributions.

This thesis documents an attempt to rise to that challenge by taking an alternative approach to the standard use of summary statistics, focusing on modeling the complex 3D clustering of galaxies[2] on quasi-linear scales directly at the *field-level* and extracting information relevant to cosmology and astrophysics. This work is practical only due to the recent development of the BORG (Bayesian Origin Reconstruction from Galaxies) framework [39] wherein a gravity-based *physical forward model* was introduced to evolve the highly Gaussian (see above), initial matter density field at very high redshift, $z \simeq 1000$, to the complex, evolved matter density field at low redshift, which in turn is compared to the observed galaxy density field in order to jointly reconstruct the initial and evolved matter density fields. The gravitational forward model forms a *deterministic* link between a given realization of the initial density field and its corresponding evolved density field while captures the nonlinear effects of gravitational collapse, consistently taking into account all the uncertainties and systematics in the dataset. In other words, given an observed galaxy density field found in a galaxy redshift survey with specific survey geometry, selection effects and galaxy biases, BORG explores the very high dimensional parameter space, $N_{\rm dim} \simeq 256^3$, of all physically compatible initial matter density fields – with the help of a Hamiltonian Monte Carlo (HMC) sampler (see Appendix B) [39, 40]. A bona fide by-product of having a dynamical forward model in the inference machinery is the reconstructed large-scale velocity field which might have useful cosmological and astrophysical applications. Chapter 6 describes such an application.

1.2 Structure

The thesis is structured as follows. In the first part, from Chapter 2 to Chapter 4, we present the theoretical background and statistical framework on which the work in this thesis is constructed.

Specifically, in Chapter 2, we review the growth of perturbations in CDM from the initial Gaussian seeds of inflation to the late-time LSS, within the framework of standard

[2]Note that, in this thesis, for the purpose of investigating our model and testing our implementation, we frequently make use of simulation data in which our tracers are cold Dark Matter (CDM) halos instead of real galaxies. We thus use the words "galaxies" and "halos" interchangeably whenever our arguments or results apply to both.

cosmological perturbation theory (PT). We explicitly re-derive the results of PT in Lagrangian space to first and second order as they are the main choices of our gravitational forward model throughout our work.

Chapter 3 is then devoted to an effective field theory (EFT) prescription of galaxy and halo bias – the key link between theoretical predictions of matter distribution and observation – based on the PT description of the evolved matter density field. We also briefly review observations of the distribution of galaxies in galaxy redshift surveys as well as determination of mean mass-richness relation of galaxy clusters from weak-lensing measurements; we focus on the SDSS3-BOSS DR12[3] [41] and the maxBCG cluster catalog [42] highlighting the important observational effects, as those datasets are directly involved in our analysis.

Chapter 4 describes the Bayesian framework for the joint inference of initial and evolved matter density fields from the biased tracer field. We discuss the main ingredients of the inference, including the *Gaussian* priors on the initial conditions and the LSS likelihood which captures the scatter induced by small-scale modes not included in our gravitational forward model and bias expansion around the predicted *mean* tracer field. We also outline the highly modular BORG framework, focusing on the modules we modified and extended in our work. We then summarize how the algorithm works, using cold Dark Matter (CDM) halos and galaxies as examples of biased tracers.

The second part of the thesis consists of our investigation into the robustness of our clustering models in recovering unbiased initial conditions and cosmological parameters from the reconstructed matter density fields, and our measurement of the *kinematic* Sunyaev-Zel'dovich (kSZ) effect using the reconstructed cosmic velocity field.

In particular, in Chapter 5, we compare the performance of different combinations of bias models and LSS likelihoods in terms of recovering the correct phases and unbiased amplitudes of initial matter density field, using CDM-only, N-body simulations. Additionally, we investigate effects of the tracer density, grid resolution, (gravitational) forward model, and tracer bias model on results of the inference.

Chapter 6 presents an astrophysical application of the reconstructed large-scale velocity field within the SDSS3-BOSS volume [43] – measuring the kSZ effect of selected maxBCG clusters from the Planck `SMICA` CMB map [44]. Our reconstruction method allows for uncertainties in the large-scale velocity reconstruction to be consistently propagated into the final uncertainties on the amplitude of the kSZ signal for the first time. We additionally take into account uncertainties in the small-scale velocity and those in the photometric redshift data. We detect the kSZ signal with a total significance of $\simeq 2\sigma$.

Chapter 7 documents our ongoing attempt to extend BORG to allow for inference of cosmological parameters of the ΛCDM model, in particular, the power spectrum normalization σ_8. We highlight the challenging issues of recovering *unbiased* σ_8 with a *localized*, *real-space* LSS likelihood. We then show that these problems can be alleviated by an EFT-based, *Fourier-space* LSS likelihood given the additional introduction of a smoothing filter with a sharp-k cut-off to ensure the Gaussianity of the error power spectrum. We illustrate

[3] `https://data.sdss.org/sas/dr12/boss/lss/`

the issues with the real-space likelihood and demonstrate the Fourier-space likelihood using the same suite of CDM-only, N-body simulations.

1.3 Notation

For reference, we list in Table 1.1 abbreviations commonly used throughout the thesis. As a guide for readers, mathematical symbols and conventions are summarized in Table 1.2, while notations of frequently encountered physical quantities are highlighted in Table 1.3.

Arguments and computations throughout the thesis are carried out under the two main assumptions about our Universe:

1. Given a constant proper-time slice through 4D spacetime manifold, the statistical distribution of any real scalar cosmic field is homogeneous and isotropic on large scale – in accordance with the *cosmological principle*. That is, the joint n-point probability distribution function $\mathcal{P}(\delta(\boldsymbol{x}_1), ...\delta(\boldsymbol{x}_n))$ of the field $\delta(\boldsymbol{x})$ is invariant under spatial translation and rotation.

2. The *ergodic hypothesis* holds for samples from well-separated regions in our Universe, that is they can be considered as independent realizations of the same underlying physical process.

In line with the cosmological principle, we adopt the perturbed and unperturbed Friedmann-Robertson-Walker (FRW) metrics, assuming a spatially flat geometry,

$$ds^2 = a^2(\tau)[-(1+2\Phi)d\tau^2 + (1-2\Psi)\delta_{ij}dx^i dx^j] \tag{1.1}$$

and

$$ds^2 = a^2(\tau)[-d\tau^2 + \delta_{ij}dx^i dx^j] \tag{1.2}$$

to describe the spacetime of our Universe – whose expansion history is encoded in the dimensionless scale factor $a(\tau)$. In this thesis, unless explicitly stated otherwise, we work with conformal time $d\tau \equiv a^{-1}dt$ and comoving coordinates $\boldsymbol{x} \equiv a^{-1}\boldsymbol{r}(\boldsymbol{x}, \tau)$, in natural units where $c = \hbar = G = k_B = 1$. Note that Eqs. (1.1)–(1.2) also imply that we work in conformal Newtonian gauge and only consider scalar perturbations.

ΛCDM	Λ Cold Dark Matter
EdS	Einstein-de Sitter (flat, matter-dominated Universe)
CMB	Cosmic microwave background
kSZ	kinematic Sunyaev-Zel'dovich effect
tSZ	thermal Sunyaev-Zel'dovich effect
BAO	Baryon acoustic oscillation
EFT	Effective field theory
PT	Perturbation theory
SPT	Eulerian standard perturbation theory
LPT	Lagrangian perturbation theory
LO	Leading order (tree level)
NLO	Next-to-leading order (1-loop)
LIMD	Local in matter density (previously commonly known as "local bias")
RSD	Redshift-space distortions
LOS	Line-of-sight
MCMC	Markov chain Monte Carlo methods
HMC	Hamiltonian Monte Carlo method
MLE	Maximum likelihood estimate
MAP	Maximum a posteriori estimate
SO	Spherical overdensity algorithm (for halo finding)
FoF	Friends-of-friends algorithm (for halo finding)

Table 1.1: List of commonly used abbreviations in the text.

Probability density function (PDF)	$\mathcal{P}(x)$
Fourier, spatial vector	$\boldsymbol{k}$, $\boldsymbol{x}$
Grid data vector	$\vec{\delta}$
Spatial derivatives	$\nabla \equiv \frac{\partial}{\partial \boldsymbol{x}}$, $\partial_i \equiv \partial/\partial x^i$, $\mathcal{D}_{ij} \equiv \frac{\partial_i \partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}$
Laplace operator	$\nabla^2 \equiv \delta^{ij}\partial_i \partial_j$
Spatial integral	$\int_{\boldsymbol{x}} \equiv \int d^3x$
Momentum integral	$\int_{\boldsymbol{k}} \equiv \int \frac{d^3k}{(2\pi)^3}$
Fourier transform	$f(\boldsymbol{k}) \equiv \int d^3x\, f(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \equiv \int_{\boldsymbol{x}} f(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}$
Inverse Fourier transform	$f(\boldsymbol{x}) \equiv \int \frac{d^3k}{(2\pi)^3} f(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \equiv \int_{\boldsymbol{k}} f(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$
Dirac delta function[1]	$\delta_D(\boldsymbol{x}) = \int_{\boldsymbol{p}} e^{i\boldsymbol{p}\cdot\boldsymbol{x}}$
Kronecker symbol	δ_{ij}
Heaviside step function	$\Theta_{\rm H}(x) = 1$ for $x > 0$ and 0 otherwise
Complementary error function	$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty du\, e^{-u^2}$
Hermite polynomials	$H_N(x)$

[1] This implies $(2\pi)^3 \int_{\boldsymbol{k}} \delta_D(\boldsymbol{k} - \boldsymbol{k}_0) f(\boldsymbol{k}) = f(\boldsymbol{k}_0)$.

Table 1.2: List of mathematical symbols and conventions adopted in this thesis.

Quantity	Symbol	Defining relation
Conformal time	τ	$d\tau \equiv a^{-1}dt$, Eqs. (1.1)–(1.2)
Eulerian comoving coordinate	$\boldsymbol{x}$	Eqs. (1.1)–(1.2)
Scale factor	$a(\tau)$	$a = \frac{1}{1+z}$
Hubble rate	$H(\tau)$	$H \equiv \dot{a}/a$
Conformal Hubble rate	$\mathcal{H}(\tau)$	$\mathcal{H} \equiv a^{-1}da/d\tau = aH$
Critical density	$\rho_{\rm cr,0}(\tau)$	$\rho_{\rm cr,0} = \frac{3H^2}{8\pi G}$
Mean comoving matter density	$\overline{\rho}_m(\tau)$	$\overline{\rho}_m \equiv a^3(\tau)\overline{\varrho}_m(\tau)$
Linear growth factor[1]	$D(\tau)$	$D(\tau) = D^{(+)}(\tau) = \frac{D^{(+)}(\tau)}{D^{(+)}(\tau_0)}$, Eq. (2.19)
Logarithmic growth rate	$f(\tau)$	$f \equiv d\ln D/d\ln a$
Gravitational potential	Φ	Eq. (1.1)
Lagrangian comoving coordinate	$\boldsymbol{q}$	$\boldsymbol{q} = \lim_{\tau\to 0}\boldsymbol{x}(\boldsymbol{q},\tau)$
Lagrangian displacement	$\boldsymbol{s}(\boldsymbol{q},\tau)$	Eq. (2.41)
Peculiar velocity	$\boldsymbol{v}$	$\boldsymbol{v} \equiv a\,\dot{\boldsymbol{x}} = d\boldsymbol{x}/d\tau$
Matter density contrast	δ_m	Eq. (2.1)
Galaxy, halo density contrast	δ_g, δ_h	$\delta_g(\boldsymbol{x},\tau) \equiv n_g(\boldsymbol{x},\tau)/\overline{n}_g(\tau) - 1$, Eq. (3.1)
Tidal field	K_{ij}	$K_{ij} \equiv (\partial_i\partial_j/\nabla^2 - \delta_{ij}/3)\delta$
Operator constructed out of density field	O	e.g., $O(\boldsymbol{x},\tau) = [\delta(\boldsymbol{x},\tau)]^2$
Renormalized operator	$[O]$	Appendix J
Linear matter power spectrum	$P_{\rm L}(k,\tau)$	$P_{\rm L}(k,\tau) \equiv \left\langle \delta^{(1)}(\boldsymbol{k},\tau)\delta^{(1)}(\boldsymbol{k}',\tau)\right\rangle'$
Variance of linear density field on scale R	$\sigma^2(R)$	$\sigma^2(R) \equiv \int_{\boldsymbol{k}} P_{\rm L}(k)W_R^2(k)$
Critical density (collapse threshold)	$\delta_{\rm cr} \simeq 1.686$	Figure 3.1
Peak significance	ν_c	$\nu_c \equiv \delta_{\rm cr}/\sigma(R)$, Eq. (3.6)
Bias parameter[3] with respect to operator O	b_O	$\delta_h(\boldsymbol{x},\tau) = \sum_O b_O(\tau)[O](\boldsymbol{x},\tau)$
N-th order LIMD bias parameter	b_N	$b_N \equiv N!\,b_{\delta^N}$, Eq. (3.11)
Lagrangian bias parameter	b_O^L	$\delta_h^L(\boldsymbol{q},\tau_0) = \sum_O b_O^L(\tau_0)[O^L](\boldsymbol{q},\tau_0)$
Filter function[4] on scale R	$W_R(x)$, $W_R(k)$	See Appendix A.

[1] This implies that $D(a=1) = 1$.
[2] This is the physical, renormalized bias, see Appendix J.
[3] Filter functions are normalized such that $\int d^3x\, W_R(x) = 1$ and $\lim_{k\to 0} W_R(k) = 1$.

Table 1.3: List of symbols and notations of frequently discussed physical quantities.

Chapter 2

Large-scale structure formation

This chapter aims to provide readers a brief overview of

1. the statistical description of initial and evolved matter density fields, and

2. the physically motivated mathematical framework for linking the two fields together, that is perturbation theory.

We refer readers to [45, 46, 47, 8] for much more thorough reviews of the subject.

2.1 Gaussian initial conditions

The notion of Gaussian initial conditions has not only arisen from theory in a few different ways but also stood the test of numerous observations. From the theoretical standpoint, given the general assumption that the initial conditions originated from quantum fluctuations of the scalar inflaton field(s) during inflationary phase of our Universe, the fact that quantum averages of products of such fluctuations are Gaussian themselves implies a Gaussian nature of the initial conditions. This is a robust prediction of the inflationary paradigm, not just limited to single-field models of inflation [46, 47]. From the observational side, results obtained from analysis of the CMB temperature anisotropies have been found to be consistent with adiabatic, Gaussian initial conditions generated by the standard single-field, slow-roll inflationary paradigm [1] [49, 50]. In addition, the *central-limit theorem* establishes that, the linear superposition of a large number of independent or uncorrelated random variables, all drawn from the same distribution, asymptotes to a Gaussian distribution. This can often be applied for primordial matter density fluctuations which are generated by the same physical mechanism [51]. Here and throughout this thesis, we assume that primordial matter density fluctuations are Gaussian random fields. We review below the summary statistics of such fields.

[1] See [48] and references therein for a full review of constraining primordial non-Gaussianity with CMB observables.

Let us first consider small matter fluctuations $\delta_m(\boldsymbol{x})$ on the homogeneous comoving background density field $\bar{\rho}_m = \langle \rho_m(\boldsymbol{x}) \rangle$, i.e.

$$\delta_m(\boldsymbol{x}) = \frac{\rho_m(\boldsymbol{x})}{\bar{\rho}_m} - 1, \tag{2.1}$$

such that $\langle \delta_m(\boldsymbol{x}) \rangle = 0$ by construction. Its n-point correlation function is then defined as

$$\begin{aligned} \xi_m^{(n)}(\delta_m(\boldsymbol{x}_1) \dots \delta_m(\boldsymbol{x}_n)) &\equiv \langle \delta_m(\boldsymbol{x}_1) \dots \delta_m(\boldsymbol{x}_n) \rangle \\ &= \int d\delta_m(\boldsymbol{x}_1) \dots d\delta_m(\boldsymbol{x}_n) \mathcal{P}(\delta_m(\boldsymbol{x}_1) \dots \delta_m(\boldsymbol{x}_n)) \delta_m(\boldsymbol{x}_1) \dots \delta_m(\boldsymbol{x}_n), \end{aligned} \tag{2.2}$$

where $\mathcal{P}(\delta_m(\boldsymbol{x}_1) \dots \delta_m(\boldsymbol{x}_n))$ denotes the joint distribution function of local fluctuations $\delta_m(\boldsymbol{x}_1), \dots, \delta_m(\boldsymbol{x}_n)$.

For a zero-mean Gaussian field $\delta_m(\boldsymbol{x})$, as in the case of the primordial fluctuations discussed in the previous paragraph, $\mathcal{P}(\delta_m(\boldsymbol{x}_1) \dots \delta_m(\boldsymbol{x}_n))$ is a zero-mean multivariate Gaussian distribution

$$\mathcal{P}(\delta_m(\boldsymbol{x}_1) \dots \delta_m(\boldsymbol{x}_n)) = \frac{1}{(2\pi)^{n/2}} |\mathcal{C}|^{-1/2} \exp\left[-\frac{1}{2} \left(\vec{\delta_m} \right)^{\intercal} \mathcal{C}^{-1} \vec{\delta_m} \right] \tag{2.3}$$

with the random vector $\vec{\delta}_m \equiv (\delta_m(\boldsymbol{x}_1), \dots, \delta_m(\boldsymbol{x}_n))$ and a positive semi-definite covariance matrix $\mathcal{C}_{ij} \equiv \langle \delta_m(\boldsymbol{x}_i) \delta_m(\boldsymbol{x}_j) \rangle$. This implies, as a consequence of Isserlis' theorem [52] – much more familiar with physicists as *Wick's theorem* [53] – that any higher-order correlation function can be written in terms of different combinations of the two-point correlation function $\xi_m^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2)$

$$\xi_m^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) \equiv \langle \delta_m(\boldsymbol{x}_1)\, \delta_m(\boldsymbol{x}_2) \rangle\,, \tag{2.4}$$

which, following statistical homogeneity and isotropy,

$$\xi_m^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \xi_m^{(2)}(\boldsymbol{x}_1 - \boldsymbol{x}_2) = \xi_m^{(2)}(r = |\boldsymbol{x}_1 - \boldsymbol{x}_2|) = \langle \delta_m(\boldsymbol{x})\, \delta_m(\boldsymbol{x} + \boldsymbol{r}) \rangle\,. \tag{2.5}$$

All statistical properties of $\delta_m(\boldsymbol{x})$ are thus completely captured in the two-point correlation function $\xi_m(r)$ [37, 45].

Let us consider the Fourier-space representation of the r.h.s. of Eq. (2.5),

$$\begin{aligned} \langle \delta(\boldsymbol{k}) \delta^*(\boldsymbol{k}') \rangle = \langle \delta(\boldsymbol{k}) \delta(-\boldsymbol{k}') \rangle &= \int d^3\boldsymbol{x} \int d^3\boldsymbol{r}\, \langle \delta(\boldsymbol{x})\, \delta(\boldsymbol{x} + \boldsymbol{r}) \rangle\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} e^{i\boldsymbol{k}'\cdot(\boldsymbol{x}+\boldsymbol{r})} \\ &= \int_{\boldsymbol{x}} \int_{\boldsymbol{r}} \langle \delta_m(\boldsymbol{x})\, \delta_m(\boldsymbol{x} + \boldsymbol{r}) \rangle\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} e^{i\boldsymbol{k}'\cdot(\boldsymbol{x}+\boldsymbol{r})} \\ &= \int_{\boldsymbol{x}} e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}} \int_{\boldsymbol{r}} \xi(r) e^{-i\boldsymbol{k}'\cdot\boldsymbol{r}} \\ &= (2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}') P(k), \end{aligned} \tag{2.6}$$

where we have used the fact that $\delta_m(\boldsymbol{x})$ is real, thus $\delta_m^*(\boldsymbol{k}') = \delta_m(-\boldsymbol{k}')$. It can be noted from Eq. (2.6) that all information about $\delta_m(\boldsymbol{x})$ is equivalently encoded in the Fourier-space counterpart of the two-point correlation function, the power spectrum $P(k)$,

$$P(k) = \int_{\boldsymbol{r}} \xi^{(2)}(r) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{2.7}$$

Note that, inversely,

$$\xi^{(2)}(r) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} P(k) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = \int_{\boldsymbol{k}} P(k) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{2.8}$$

2.2 The perturbative theory of matter fluctuations: from initial conditions to large-scale structure

To mathematically describe how tiny matter density fluctuations in the initial conditions, amplified by gravitational collapse, grow to become the non-linear LSS now observed in our Universe, is the goal of perturbation theory (PT). In this section, assuming that LSS formation is driven by non-relativistic[2] *collisionless* CDM [54, 55, 56, 57], i.e. the dynamics of $\delta_m(\boldsymbol{x})$ can be well approximated by that of a Newtonian self-gravitating, pressureless perfect fluid, we summarize the key results of PT in Eulerian and Lagrangian description, up to second-order.

The PT framework relies on the notion of the phase-space distribution function $f(\boldsymbol{x}, \boldsymbol{p}, \tau)$ of a single CDM particle with mass m and momentum $\boldsymbol{p} = m\boldsymbol{v}$ $f(\boldsymbol{x}, \boldsymbol{p}, \tau)$ obeys the *collisionless Boltzmann equation* for massive particles [45, 47] in Newtonian limit

$$\begin{aligned}\frac{df(\boldsymbol{x}, \boldsymbol{p}, \tau)}{d\tau} &= \frac{\partial f(\boldsymbol{x}, \boldsymbol{p}, \tau)}{\partial \tau} + \frac{d\boldsymbol{x}}{d\tau} \cdot \nabla f(\boldsymbol{x}, \boldsymbol{p}, \tau) + \frac{d\boldsymbol{p}}{d\tau} \frac{\partial f(\boldsymbol{x}, \boldsymbol{p}, \tau)}{\partial \boldsymbol{p}} \\ &= \frac{\partial f(\boldsymbol{x}, \boldsymbol{p}, \tau)}{\partial \tau} + \frac{\boldsymbol{p}}{ma(\tau)} \cdot \nabla f(\boldsymbol{x}, \boldsymbol{p}, \tau) - ma(\tau) \nabla \Phi(\boldsymbol{x}, \tau) \frac{\partial f(\boldsymbol{x}, \boldsymbol{p}, \tau)}{\partial \boldsymbol{p}} = 0. \end{aligned} \tag{2.9}$$

$\nabla \equiv \frac{\partial}{\partial \boldsymbol{x}}$ denotes spatial derivatives; $\Phi(\boldsymbol{x}, \tau)$ denotes the peculiar gravitational potential sourced by matter density fluctuations $\delta_m(\boldsymbol{x}, \tau)$, as described by the Poisson equation

$$\nabla^2 \Phi(\boldsymbol{x}, \tau) = \frac{3}{2} \mathcal{H}(\tau)^2 \Omega_{\rm m}(\tau) \delta_m(\boldsymbol{x}, \tau). \tag{2.10}$$

The evolution of spatial distribution of CDM is then obtained by taking momentum moments of the system of collisionless Boltzmann-Poisson equations Eqs. (2.9)–(2.10). The hierarchy of equations of motion is then closed at second-order with some certain ansatz. For example, a common choice is to set the stress tensor σ_{ij} to null, corresponding to the

[2]This implies $\boldsymbol{v}(\boldsymbol{x}, \tau) \ll c$, which is usually a good approximation on the scales of interest for LSS formation.

notion of a single-stream or pressureless fluid – which is valid for the case of collisionless CDM (as discussed above), up until *shell crossing* [45].

In the standard approach, the linear matter density fluctuations $\delta_m^{(1)}(\boldsymbol{x},\tau)$ and linear peculiar velocity $\boldsymbol{v}^{(1)}(\boldsymbol{x},\tau)$ are obtained by solving the linearized equations of motion for CDM. Higher-order solutions can then be expressed as perturbative expansions[3] around those linear solutions [58, 59, 60, 45]

$$\delta_m(\boldsymbol{x},\tau)=\sum_n^{\infty}\delta_m^{(n)}(\boldsymbol{x},\tau),\quad \theta(\boldsymbol{x},\tau)=\sum_n^{\infty}\theta^{(n)}(\boldsymbol{x},\tau). \tag{2.11}$$

Let us next briefly review this approach in two different but equivalent descriptions of the CDM fluid, highlighting the key results later employed in our work. Readers who are interested in this specific topic can find more details in [61] and references therein.

2.2.1 Eulerian Perturbation Theory

The Eulerian description of a fluid fixates on a particular position in space, such that our CDM pressure-less fluid are depicted by two dynamical variables, the matter density contrast δ_m and the peculiar velocity $\boldsymbol{v}$. Both are functions of comoving coordinates and conformal time $(\boldsymbol{x},\tau)$. Their evolution is governed by the continuity, Euler and Poisson equations

$$\frac{\partial\delta_m(\boldsymbol{x},\tau)}{\partial\tau}+\nabla\cdot\{[1+\delta_m(\boldsymbol{x},\tau)]\,\boldsymbol{v}(\boldsymbol{x},\tau)\}=0, \tag{2.12}$$

$$\frac{\partial\boldsymbol{v}(\boldsymbol{x},\tau)}{\partial\tau}+[\boldsymbol{v}(\boldsymbol{x},\tau)\cdot\nabla]\,\boldsymbol{v}(\boldsymbol{x},\tau)+\mathcal{H}(\tau)\boldsymbol{v}(\boldsymbol{x},\tau)=-\nabla\Phi(\boldsymbol{x},\tau), \tag{2.13}$$

wherein $\mathcal{H}\equiv a^{-1}da/d\tau=aH$ denotes the conformal Hubble rate. Note that we have used the single-stream approximation and set $\sigma_{ij}=0$ on the r.h.s. of Eq. (2.13).

Linear Eulerian Perturbation Theory

In the linear regime of gravitational evolution, where $\delta_m(\boldsymbol{x},\tau)=\delta_m^{(1)}(\boldsymbol{x},\tau)\ll 1$, we can linearize Eqs. (2.12)–(2.13) by keeping only first-order terms

$$\frac{\partial\delta_m^{(1)}(\boldsymbol{x},\tau)}{\partial\tau}+\theta^{(1)}(\boldsymbol{x},\tau)=0, \tag{2.14}$$

$$\frac{\partial\boldsymbol{v}^{(1)}(\boldsymbol{x},\tau)}{\partial\tau}+\mathcal{H}(\tau)\boldsymbol{v}^{(1)}(\boldsymbol{x},\tau)=-\nabla\Phi(\boldsymbol{x},\tau), \tag{2.15}$$

where we have introduced the divergence of the peculiar velocity field $\theta(\boldsymbol{x},\tau)\equiv\nabla\cdot\boldsymbol{v}(\boldsymbol{x},\tau)$. Taking the divergence on both sides of Eq. (2.15) and plugging Eq. (2.10) in, we have

$$\frac{\partial\theta^{(1)}(\boldsymbol{x},\tau)}{\partial\tau}+\mathcal{H}(\tau)\theta^{(1)}(\boldsymbol{x},\tau)+\frac{3}{2}\Omega_{\rm m}(\tau)\mathcal{H}^2(\tau)\delta_m^{(1)}(\boldsymbol{x},\tau)=0. \tag{2.16}$$

[3] This ansatz is valid so long as we restrict ourselves to the quasi-linear scales.

A single second-order ordinary differential equation can then be obtained by combining Eq. (2.14) and Eq. (2.16)

$$\frac{\partial^2\delta_m^{(1)}(\boldsymbol{x},\tau)}{\partial\tau^2}+\mathcal{H}(\tau)\frac{\partial\delta_m^{(1)}(\boldsymbol{x},\tau)}{\partial\tau}-\frac{3}{2}\Omega_{\rm m}(\tau)\mathcal{H}^2(\tau)\delta_m^{(1)}(\boldsymbol{x},\tau)=0, \tag{2.17}$$

or

$$\frac{d^2D_1(\tau)}{d\tau^2}+\mathcal{H}(\tau)\frac{dD_1(\tau)}{d\tau}-\frac{3}{2}\Omega_{\rm m}(\tau)\mathcal{H}^2(\tau)D_1(\tau)=0, \tag{2.18}$$

where, in the last step, we have factored out the linear growth factor $D_1(\tau)$ from the linear solution $\delta_m^{(1)}(\boldsymbol{x},\tau)$

$$\delta_m^{(1)}(\boldsymbol{x},\tau)=\frac{D_1(\tau)}{D_1(\tau_0)}\delta_m^{(1)}(\boldsymbol{x},\tau_0). \tag{2.19}$$

The evolution of the linear matter density field $\delta_m^{(1)}(\boldsymbol{x},\tau)$ can then be written as a linear superposition of the two independent solutions of Eq. (2.18)

$$\delta_m^{(1)}(\boldsymbol{x},\tau)=\left[D_1^{(+)}(\tau)+D_1^{(-)}(\tau)\right]\delta_m^{(1)}(\boldsymbol{x},\tau_0), \tag{2.20}$$

or, equivalently,

$$\delta_m^{(1)}(\boldsymbol{k},\tau)=\left[D_1^{(+)}(\tau)+D_1^{(-)}(\tau)\right]\delta_m^{(1)}(\boldsymbol{k},\tau_0), \tag{2.21}$$

where the $(+)$ and $(-)$ signs denote the growing and decaying mode, respectively. It is worth emphasizing that, in the linear regime, all Fourier modes of the matter density fluctuations independently evolve at a universal rate.

Plugging Eq. (2.20) into Eq. (2.14), we arrive at the expression for the linear velocity divergence

$$\begin{aligned}\theta^{(1)}(\boldsymbol{x},\tau)&=-\frac{\partial D_1(\tau)}{\partial\tau}\frac{\delta_m^{(1)}(\boldsymbol{x},\tau_0)}{D_1(\tau_0)}\\&=-\frac{d\ln D_1(\tau)}{d\tau}\delta_m^{(1)}(\boldsymbol{x},\tau)=-\mathcal{H}(\tau)\left[f(\Omega_{\rm m},\Omega_\Lambda)+g(\Omega_{\rm m},\Omega_\Lambda)\right]\delta_m^{(1)}(\boldsymbol{x},\tau)\end{aligned} \tag{2.22}$$

in which we have introduced the logarithmic growth rate

$$f(\Omega_{\rm m},\Omega_\Lambda)\equiv\frac{d\ln D_1^{(+)}}{d\ln a}, \tag{2.23}$$

and the logarithmic decay rate

$$g(\Omega_{\rm m},\Omega_\Lambda)\equiv\frac{d\ln D_1^{(-)}}{d\ln a}. \tag{2.24}$$

The forms of Eqs. (2.20)–(2.22) reflects a key feature of linear gravitational evolution: it preserves *Gaussianity* of the initial conditions described in Section 2.1.

Since we are interested only in the growth of structures from primordial density fluctuations while the decaying modes decay away as the Universe expands, i.e. $D_1^{(-)} \propto a^{-n}$ with $n > 0$, we will only focus on the growing mode henceforth and adopt the notation

$$D_1(\tau) = \frac{D_1^{(+)}(\tau)}{D_1^{(+)}(\tau_0)}. \tag{2.25}$$

A general solution of $D_1(\tau)$ for Universe with matter and a cosmological constant Λ can be written in integral form [62]

$$D_1(\tau) = \frac{\mathcal{H}(\tau)}{a(\tau)} \frac{5\Omega_{\rm m}}{2} \int_0^a \frac{da}{a^2(\tau)\mathcal{H}(\tau)}, \tag{2.26}$$

wherein

$$\mathcal{H}(a) = \left[\Omega_\Lambda a^2 + (1 - \Omega_{\rm m} - \Omega_\Lambda) + \Omega_{\rm m} a^{-1}\right]^{-\frac{1}{2}}. \tag{2.27}$$

For convenience, it is common practice to express $D_1(\tau)$ as a function of the scale factor $a(\tau)$.

Second-order Eulerian Perturbation Theory

It is often more convenient to solve for second- and higher-order solutions of Eqs. (2.10)–(2.13) in Fourier-space, where we can combine Eqs. (2.10)–(2.13) into two equations of motion, keeping the higher-order terms

$$\frac{\partial \delta_m(\boldsymbol{k},\tau)}{\partial \tau} + \theta(\boldsymbol{k},\tau) = -\int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2} (2\pi)^3 \delta_D(\boldsymbol{k}-\boldsymbol{k}_{12}) \alpha(\boldsymbol{k}_1,\boldsymbol{k}_2)\theta(\boldsymbol{k}_1,\tau)\delta_m(\boldsymbol{k}_2,\tau), \tag{2.28}$$

$$\frac{\partial \theta(\boldsymbol{k},\tau)}{\partial \tau} + \mathcal{H}(\tau)\theta(\boldsymbol{k},\tau) + \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau)\delta_m(\boldsymbol{k},\tau) = -\int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2} (2\pi)^3 \delta_D(\boldsymbol{k}-\boldsymbol{k}_{12}) \beta(\boldsymbol{k}_1,\boldsymbol{k}_2)\theta(\boldsymbol{k}_1,\tau)\theta(\boldsymbol{k}_2,\tau). \tag{2.29}$$

where $\boldsymbol{k}_{12} = \boldsymbol{k}_1 + \boldsymbol{k}_2$ and the non-linear terms in Eqs. (2.12)–(2.13) are encoded in

$$\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{\boldsymbol{k}_{12}\cdot\boldsymbol{k}_1}{k_1^2}, \quad \beta(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{k_{12}^2(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)}{2k_1^2k_2^2}. \tag{2.30}$$

Note that, as opposed to linear order, non-linearity in gravitational evolution introduces *mode-coupling*, i.e. each Fourier mode $\delta_m(\boldsymbol{k},\tau)$ in Eqs. (2.28)–(2.29) no longer evolves independently but receives contributions from the coupling between all pairs of $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ such that $\boldsymbol{k} = \boldsymbol{k}_1 + \boldsymbol{k}_2$. Similar to Eq. (2.11), given the notion of

$$\delta_m(\boldsymbol{k},\tau) = \sum_n^\infty \delta_m^{(n)}(\boldsymbol{k},\tau), \quad \theta(\boldsymbol{k},\tau) = \sum_n^\infty \theta^{(n)}(\boldsymbol{k},\tau), \tag{2.31}$$

we can write the n-th order solution of Eqs. (2.28)–(2.29) as [58, 59, 60]

$$\delta_m^{(n)}(\boldsymbol{k},\tau) = \int_{\boldsymbol{k}_1}\cdots\int_{\boldsymbol{k}_n}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12\ldots n})F_n(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_n,\tau)\delta_m^{(1)}(\boldsymbol{k}_1,\tau)\ldots\delta_m^{(1)}(\boldsymbol{k}_n,\tau), \quad (2.32)$$

$$\theta^{(n)}(\boldsymbol{k},\tau) = -\mathcal{H}(\tau)f(\tau)\int_{\boldsymbol{k}_1}\cdots\int_{\boldsymbol{k}_n}(2\pi)^3\delta_D\left(\boldsymbol{k}-\boldsymbol{k}_{12\ldots n}\right)G_n(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_n,\tau)\delta_m^{(1)}(\boldsymbol{k}_1,\tau)\ldots\delta_m^{(1)}(\boldsymbol{k}_n,\tau), \quad (2.33)$$

where the symmetric density and velocity divergence kernels F_n and G_n are recursively constructed out of Eq. (2.30). Note that, for $n = 1$,

$$F_1 = G_1 = 1. \quad (2.34)$$

For an EdS universe, where $D(\tau) = a$ and $f = \Omega_{\rm m} = 1$ (cf. Eqs. (2.26)–(2.23)) the kernels are time-independent [63]. The simple forms of F_n, G_n in EdS cosmology, however, give a very good approximation to those in other cosmologies [45, 64]. Hence it has become a common practice in SPT to compute these kernels in EdS cosmology and then apply them for other cosmologies, e.g. ΛCDM [59, 60, 45]. For $n = 2$, in an EdS universe [60],

$$F_2(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{5}{7} + \frac{2}{7}\frac{(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)^2}{k_1^2k_2^2} + \frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right), \quad (2.35)$$

$$G_2(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{3}{7} + \frac{4}{7}\frac{(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)^2}{k_1^2k_2^2} + \frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right). \quad (2.36)$$

Going back to real-space, the second-order density and velocity field can be expressed as [65]

$$\delta_m^{(2)}(\boldsymbol{x},\tau) = \frac{17}{21}\left[\delta_m^{(1)}(\boldsymbol{x},\tau)\right]^2 + \frac{2}{7}\left[K_{ij}^{(1)}(\boldsymbol{x},\tau)\right]^2 - s_{(1)}^i\partial_i\delta_m^{(1)}(\boldsymbol{x},\tau), \quad (2.37)$$

$$-\frac{1}{\mathcal{H}(\tau)f(\tau)}\theta^{(2)}(\boldsymbol{x},\tau) = \frac{13}{21}\left[\delta_m^{(1)}(\boldsymbol{x},\tau)\right]^2 + \frac{4}{7}\left[K_{ij}^{(1)}(\boldsymbol{x},\tau)\right]^2 - s_{(1)}^i\partial_i\delta_m^{(1)}(\boldsymbol{x},\tau), \quad (2.38)$$

where

$$\boldsymbol{s}_{(1)}(\boldsymbol{q},\tau) = \boldsymbol{x}^{(1)}(\tau) - \boldsymbol{q} = -\frac{\boldsymbol{\nabla}}{\nabla^2}\delta^{(1)}(\boldsymbol{q},\tau) \quad (2.39)$$

is the first-order Lagrangian displacement, and

$$K_{ij}^{(1)}(\boldsymbol{x},\tau) = \left[\frac{\partial_i\partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}\right]\delta_m^{(1)}(\boldsymbol{x},\tau) \quad (2.40)$$

is the first-order tidal field, both of which will be encountered in the next chapter.

To summarize, higher-order solutions of SPT are perturbative expansions of the linear solutions Eqs. (2.20)–(2.22) with symmetric kernels computed recursively from Eq. (2.30). The cosmological dependence of these solutions is encoded in the linear growth factor $D(\tau)$ in Eq. (2.19). As such, measurements of the linear growth factor, or equivalently, the logarithmic growth rate $f(\tau)$ provide a powerful connection between theoretical models of Dark Energy or modified gravity and observations of LSS and CMB [66]. The evolution of the two quantities and their second-order counterparts in three different cosmologies are illustrated in Figure 2.1.

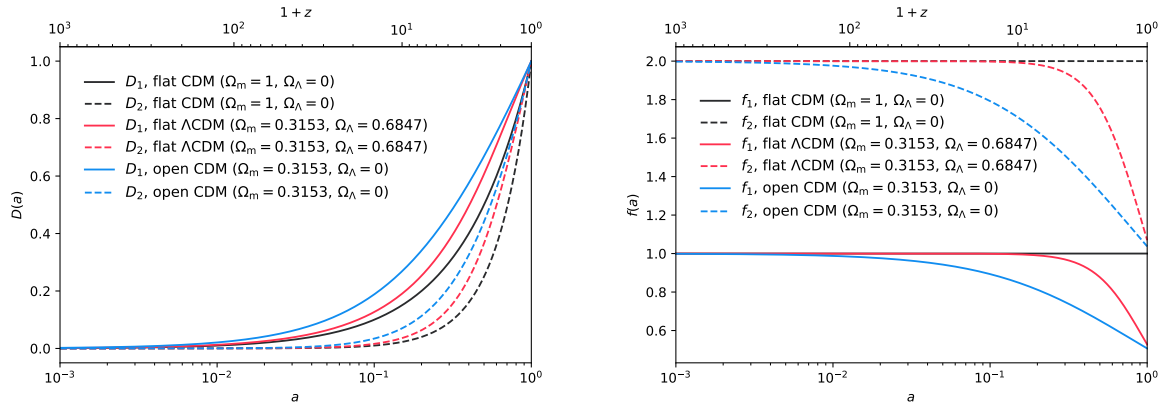

Figure 2.1: The linear (continuous) and second-order (dashed) growth factors D (left panel) and logarithmic growth rates $f = \frac{d\ln D}{d\ln a}$ (right panel) as functions of the scale factor a, or equivalently, redshift z, in three different cosmologies: EdS (gray), flat ΛCDM (red), and open CDM (blue).

2.2.2 Lagrangian Perturbation Theory

The Lagrangian description of the CDM fluid was first introduced in [67, 68], then later generalized in [69, 70] and extended to second- and third-order in [71, 72, 73]. In this picture, instead of fixating on a given comoving spatial position, one follows the trajectory of a given individual fluid element. The dynamical variable depicting our fluid element is then the *Lagrangian displacement* $\boldsymbol{s}$, which is a function of the comoving Lagrangian initial position $\boldsymbol{q}$ and the conformal time τ, such that

$$\boldsymbol{x}_{\rm fl}(\tau) = \boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q}, \tau), \tag{2.41}$$

where $\boldsymbol{x}(\tau)$ is the comoving Eulerian final position of the fluid element. Note that $\boldsymbol{s}(\boldsymbol{q}, \tau = 0) = 0$ and $\boldsymbol{x}(\tau = 0) = \boldsymbol{q}$. The total number of particles enclosed in and thus the total mass $dm_{\rm fl}$ of each fluid element is conserved such that

$$\delta_{\rm fl} = mn(\tau)d^3q = \overline{\rho}_m(\tau)d^3q. \tag{2.42}$$

Compare Eq. (2.42) to that in the Eulerian picture, we have

$$\overline{\rho}_m(\tau)d^3q = \rho(\boldsymbol{x}, \tau)d^3x = \overline{\rho}_m(\tau)[1 + \delta_m(\boldsymbol{x}, \tau)]d^3x. \tag{2.43}$$

Let us define the Jacobian (determinant) of the Lagrangian to Eulerian transformation as

$$\mathcal{J}(\boldsymbol{q}, \tau) = \left|\delta_{ij} + s_{i,j}(\boldsymbol{q}, \tau)\right| = 1 + s_{i,i} + \frac{1}{2}\left[(s_{i,i})^2 - s_{i,j}s_{j,i}\right] + \mathcal{O}(s^3_{i,j}), \tag{2.44}$$

where $s_{i,j} \equiv \frac{\partial s_i}{\partial q_j}$. Then Eq. (2.43) allows us to relate the Eulerian matter density fluctuations $\delta_m(\boldsymbol{x}, \tau)$ to the Jacobian $\mathcal{J}(\boldsymbol{q}, \tau)$

$$\mathcal{J}(\boldsymbol{q}, \tau) = \left|\frac{d^3x}{d^3q}\right| = \frac{1}{1 + \delta_m(\boldsymbol{x}, \tau)}. \tag{2.45}$$

Recall that the fluid trajectory in an expanding Universe follows the Euler equation Eq. (2.15), i.e.

$$\frac{d^2\boldsymbol{x}_{\rm fl}}{d\tau^2} + \mathcal{H}(\tau)\frac{d\boldsymbol{x}_{\rm fl}}{d\tau} = -\boldsymbol{\nabla}_{\boldsymbol{x}}\Phi, \tag{2.46}$$

where we have replaced $\boldsymbol{v}(\boldsymbol{x}(\boldsymbol{q}),\tau) = d\boldsymbol{s}(\boldsymbol{q},\tau)/d\tau$. Then, similar to Eq. (2.16), we can derive

$$\mathcal{J}(\boldsymbol{q},\tau)\nabla_{\boldsymbol{x}}\cdot\left[\frac{d^2\boldsymbol{x}_{\rm fl}}{d\tau^2} + \mathcal{H}(\tau)\frac{d\boldsymbol{x}_{\rm fl}}{d\tau}\right] - \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau)[J(\boldsymbol{q},\tau)-1] = 0, \tag{2.47}$$

using also Eq. (2.45). Next, let us replace the Eulerian variable $\boldsymbol{x}$ by the displacement $\boldsymbol{s}(\boldsymbol{q},\tau)$ following the chain rule

$$\frac{\partial}{\partial x_i} = \left[\delta_{ij} + s_{i,j}(\boldsymbol{q},\tau)\right]^{-1}\frac{\partial}{\partial q_j}, \tag{2.48}$$

so that Eq. (2.47) becomes

$$\mathcal{J}(\boldsymbol{q},\tau)\left[\delta_{ij} + s_{i,j}(\boldsymbol{q},\tau)\right]^{-1}\left[\frac{d^2 s_{i,j}(\boldsymbol{q},\tau)}{d\tau^2} + \mathcal{H}(\tau)\frac{ds_{i,j}(\boldsymbol{q},\tau)}{d\tau}\right] - \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau)[\mathcal{J}(\boldsymbol{q},\tau)-1] = 0. \tag{2.49}$$

Equation (2.49) is the Lagrangian equation of motion for the displacement field $\boldsymbol{s}(\boldsymbol{q},\tau)$. It describes the evolution of the CDM fluid. This equation can be solved perturbatively by expanding $\boldsymbol{s}(\boldsymbol{q},\tau)$, as similar to Eq. (2.11),

$$\boldsymbol{s}(\boldsymbol{q},\tau) = \sum_n^{\infty}\boldsymbol{s}^{(n)}(\boldsymbol{q},\tau). \tag{2.50}$$

In the following sections, we review its first- (LPT) and second-order (2LPT) solutions, which we employ as our gravitational forward models throughout Chapter 5, Chapter 6, Chapter 7.

Linear Lagrangian Perturbation Theory

As first-order approximations (see, e.g. [69, 70, 73]),

$$\mathcal{J}(\boldsymbol{q},\tau) \simeq 1 + s^{(1)}_{i,i}(\boldsymbol{q},\tau), \tag{2.51}$$

$$1 + \delta_m(\boldsymbol{x},\tau) = |\mathcal{J}(\boldsymbol{q},\tau)|^{-1} \simeq 1 - s^{(1)}_{i,i}(\boldsymbol{q},\tau), \tag{2.52}$$

$$\left[\delta_{ij} + s^{(1)}_{i,j}(\boldsymbol{q},\tau)\right]^{-1} \simeq \delta_{ij} - s^{(1)}_{i,j}(\boldsymbol{q},\tau). \tag{2.53}$$

We can rewrite Eq. (2.49) using these approximations as

$$\left[1 + s^{(1)}_{k,k}\right]\left[\delta_{ij} - s^{(1)}_{i,j}\right]\left[\frac{d^2 s^{(1)}_{i,j}}{d\tau^2} + \mathcal{H}(\tau)\frac{ds^{(1)}_{i,j}}{d\tau}\right] - \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau)s^{(1)}_{k,k} = 0, \tag{2.54}$$

keeping only first-order terms [45],

$$\frac{d^2 s_{i,i}^{(1)}(\boldsymbol{q},\tau)}{d\tau^2} + \mathcal{H}(\tau)\frac{d s_{i,i}^{(1)}(\boldsymbol{q},\tau)}{d\tau} - \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau) s_{i,i}^{(1)}(\boldsymbol{q},\tau) = 0 \tag{2.55}$$

where $s_{i,i}^{(1)} \equiv \nabla_{\boldsymbol{q}} \cdot \boldsymbol{s}^{(1)}$. Here we can, similar to Eq. (2.17), factor out the *same* linear growth factor $D_1(\tau)$ as

$$\boldsymbol{s}^{(1)}(\boldsymbol{q},\tau) = \frac{D_1(\tau)}{D_1(\tau_0)}\boldsymbol{s}^{(1)}(\boldsymbol{q},\tau_0), \tag{2.56}$$

such that the linear solution $\boldsymbol{s}^{(1)}(\boldsymbol{q},\tau)$ is simply given by [70, 73]

$$\nabla_{\boldsymbol{q}} \cdot \boldsymbol{s}^{(1)}(\boldsymbol{q},\tau) = -\delta_m^{(1)}(\boldsymbol{x},\tau) = -\frac{D_1(\tau)}{D_1(\tau=0)}\delta_m^{(1)}(\boldsymbol{x}=\boldsymbol{q}), \tag{2.57}$$

with the evolution of $D_1(\tau)$ following exactly Eq. (2.18).

The LPT solution can also be expressed in terms of the (Eulerian) position $\boldsymbol{x}_{\rm fl}(\boldsymbol{q},\tau)$ and peculiar velocity $\boldsymbol{v}_{\rm fl}(\boldsymbol{q},\tau)$ of the fluid element. The irrotational condition of the Lagrangian displacement[4],

$$\nabla_{\boldsymbol{q}} \times \boldsymbol{s}^{(1)}(\boldsymbol{q},\tau) = 0, \tag{2.58}$$

motivates the definition of the Lagrangian potential $\Phi^{(1)}(\boldsymbol{q},\tau)$ that satisfies

$$\boldsymbol{s}^{(1)}(\boldsymbol{q},\tau) = -\nabla_{\boldsymbol{q}}\Phi^{(1)}(\boldsymbol{q},\tau). \tag{2.59}$$

We can thus rewrite Eq. (2.57) as

$$\nabla_{\boldsymbol{q}} \cdot \boldsymbol{s}^{(1)}(\boldsymbol{q},\tau) = -\nabla_{\boldsymbol{q}}^2\Phi^{(1)}(\boldsymbol{q},\tau) = -\delta_m^{(1)}(\boldsymbol{x},\tau). \tag{2.60}$$

Hence [73, 45],

$$\boldsymbol{x}_{\rm fl}(\boldsymbol{q},\tau) = \boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q},\tau) = \boldsymbol{q} - \nabla_{\boldsymbol{q}}^{-1}\delta_m(\boldsymbol{x},\tau) = \boldsymbol{q} - \nabla_{\boldsymbol{q}}\Phi^{(1)}(\boldsymbol{q},\tau), \tag{2.61}$$

$$\boldsymbol{v}_{\rm fl}(\boldsymbol{q},\tau) = \mathcal{H}(\tau)f(\tau)\nabla_{\boldsymbol{q}}^{-1}\delta_m(\boldsymbol{x},\tau) = -\mathcal{H}(\tau)f(\tau)\nabla_{\boldsymbol{q}}\Phi^{(1)}(\boldsymbol{q},\tau). \tag{2.62}$$

An important realization from Eqs. (2.61)–(2.62) is that, in comoving coordinates, fluid elements travel on straight lines, along the direction set by their initial peculiar velocities [61].

The LPT solution is often referred to as *Zel'dovich approximation* (ZA) since it was proposed by Zel'dovich in [74], who used this solution to extrapolate fluid element trajectories into the quasi-linear regime where the condition $\delta_m(\boldsymbol{x},\tau) \ll 1$ no longer holds and examined the formation of the so-called *Zel'dovich pancake*. Later, it was used to analyze the growth of galactic spin from initial tidal field [75].

Although LPT (or ZA) breaks down at shell crossing, this simple solution can still provide us an intuitive picture of the formation of complex LSS such as sheets, filaments, DM

[4]This is a direct consequence of the single-stream approximation and hence only valid before shell crossing. It is worth noting that the SPT solution, evaluated at any order, fails at shell crossing as well.

halos and voids. Recall that the Jacobian matrix Eq. (2.44) is symmetric as a consequence of Eq. (2.59), this implies the "tensor of deformation" $\mathcal{D}_{ij}(\boldsymbol{q},\tau) \equiv \delta_{ij} + \boldsymbol{s}_{i,j}(\boldsymbol{q},\tau)$ can be brought to the diagonal form [74]

$$\mathcal{D}(\boldsymbol{q},\tau) = \begin{pmatrix} 1-\lambda_1(\boldsymbol{q})D(\tau) & 0 & 0 \\ 0 & 1-\lambda_2(\boldsymbol{q})D(\tau) & 0 \\ 0 & 0 & 1-\lambda_3(\boldsymbol{q})D(\tau) \end{pmatrix}. \tag{2.63}$$

Hence the evolution of the matter density fluctuations $\delta_m(\boldsymbol{x},\tau)$ in LPT can be expressed as

$$1+\delta_m(\boldsymbol{x},\tau) = \mathcal{J}(\boldsymbol{q},\tau)^{-1} = \Big\{ \left[1-\lambda_1 D(\tau)\right]\left[1-\lambda_2 D(\tau)\right]\left[1-\lambda_3 D(\tau)\right] \Big\}^{-1}, \tag{2.64}$$

where $\lambda_i(\boldsymbol{q})$ are local eigenvalues of the tidal tensor $s_{i,j}$ and positive eigenvalues correspond to growing modes, such that the evolution can be classified into these following four cases:

1. $\lambda_1 > 0$ and $\lambda_1 > \lambda_2$, λ_3 case correponds planar collapse or sheet formation;
2. λ_1, $\lambda_2 > 0$ and $\lambda_1 \simeq \lambda_2 > \lambda_3$ case corresponds to cylindrical collapse or filament formation;
3. $\lambda_1 \simeq \lambda_2 \simeq \lambda_3 > 0$ case corresponds to spherical collapse or DM halo/cluster[5] formation;
4. λ_1, λ_2, $\lambda_3 < 0$ case corresponds to underdense region expansion or void formation.

Various analytical and numerical comparisons have shown that, compared to linear SPT, LPT solution not only extends the validity regime of PT but also improves the accuracy of the evolved matter density field, specifically its first- and second-order moments [73, 72, 76, 77]. Especially in one-dimensional cases where LPT solution is exact (up until shell crossing), [78] proved that the solution is identical to that of SPT in the infinite loop limit.

Second-order Lagrangian perturbation theory

Despite the success of LPT over linear SPT, it still fails to produce evolved matter density field with sufficiently accurate skewness and higher order moments [71, 73]. In particular, [79] showed that initial conditions set by LPT could lead to incorrect second- and higher-order growing modes, which in turn excite non-linear decay modes in N-body simulations. These results motivate the study and use of 2LPT solution, which we review below.

To begin, we again expand the Jacobian $\mathcal{J}(\boldsymbol{q},\tau)$ (cf. Eq. (2.44)), this time replacing $\boldsymbol{s}(\boldsymbol{q},\tau) = \boldsymbol{s}^{(1)}(\boldsymbol{q},\tau) + \boldsymbol{s}^{(2)}(\boldsymbol{q},\tau)$ and keeping up-to-2nd-order terms [70, 71, 73]

$$\mathcal{J}(\boldsymbol{q},\tau) \simeq 1 + s^{(1)}_{i,i}(\boldsymbol{q},\tau) + s^{(2)}_{i,i}(\boldsymbol{q},\tau) + \frac{1}{2}\left[\left(s^{(1)}_{i,i}(\boldsymbol{q},\tau)\right)^2 - s^{(1)}_{i,j}(\boldsymbol{q},\tau)s^{(1)}_{j,i}(\boldsymbol{q},\tau)\right]. \tag{2.65}$$

[5] We have assumed that baryonic matter comoves with the CDM fluid, thus the formation of galaxy clusters is equivalent to that of their host CDM halos in this particular context.

Plugging Eq. (2.65) into Eq. (2.49), we retrieve

$$\left(\frac{d^2 s^{(2)}_{i,i}}{d\tau^2} + \mathcal{H}\frac{ds^{(2)}_{i,i}}{d\tau}\right) + s^{(1)}_{k,k}\left(\frac{d^2 s^{(1)}_{i,i}}{d\tau^2} + \mathcal{H}\frac{ds^{(1)}_{i,i}}{d\tau}\right) - s^{(1)}_{i,j}\left(\frac{d^2 s^{(1)}_{i,j}}{d\tau^2} + \mathcal{H}\frac{ds^{(1)}_{i,j}}{d\tau}\right) = \frac{3}{2}\mathcal{H}^2\Omega_{\rm m}\left[s^{(2)}_{k,k} + \frac{1}{2}\left(s^{(1)}_{k,k}\right)^2 - \frac{1}{2}s^{(1)}_{i,j}s^{(1)}_{j,i}\right]. \quad (2.66)$$

Due to symmetry, $\boldsymbol{s}^{(1)}_{i,j} = \boldsymbol{s}^{(1)}_{j,i} = -\Phi^{(1)}_{,ij}$, using that fact and Eq. (2.55), Eq. (2.66) can be simplified to

$$\left(\frac{d^2 s^{(2)}_{i,i}}{d\tau^2} + \mathcal{H}(\tau)\frac{ds^{(2)}_{i,i}}{d\tau} - \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau)s^{(2)}_{i,i}\right) + \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau)\left[\frac{1}{2}\left(s^{(1)}_{k,k}\right)^2 - \frac{1}{2}s^{(1)}_{i,j}s^{(1)}_{j,i}\right] = 0, \quad (2.67)$$

which can then be separated into time and spatial parts [70, 73]. The former evolves as

$$\boldsymbol{s}(\boldsymbol{q},\tau) = \frac{D_1(\tau)}{D_1(\tau_0)}\boldsymbol{s}^{(1)}(\boldsymbol{q},\tau_0) + \frac{D_2(\tau)}{D_2(\tau_0)}\boldsymbol{s}^{(2)}(\boldsymbol{q},\tau_0) \quad (2.68)$$

with

$$\frac{d^2 D_2(\tau)}{d\tau^2} + \mathcal{H}(\tau)\frac{dD_2(\tau)}{d\tau} - \frac{3}{2}\mathcal{H}^2(\tau)\Omega_{\rm m}(\tau)\left[D_2(\tau) + D_1^2(\tau)\right] = 0, \quad (2.69)$$

where, in a flat ΛCDM universe, $D_2(\tau) \simeq -\frac{3}{7}\Omega_{\rm m}^{-1/143}D_1^2(\tau)$ (see left panel of Figure 2.1) to better than 0.6% [73]. The later describes the tidal effect

$$s^{(2)}_{i,i}(\boldsymbol{q},\tau) = \frac{D_2(\tau)}{2D_1^2(\tau)}\sum_{i\neq j}\left[s^{(1)}_{i,i}(\boldsymbol{q},\tau)s^{(1)}_{j,j}(\boldsymbol{q},\tau) - s^{(1)}_{i,j}(\boldsymbol{q},\tau)s^{(1)}_{j,i}(\boldsymbol{q},\tau)\right] \simeq -\frac{3}{7}\Omega_{\rm m}^{-1/143}(\tau)\sum_{i>j}\left\{\Phi^{(1)}_{,ii}(\boldsymbol{q},\tau)\Phi^{(1)}_{,jj}(\boldsymbol{q},\tau) - \left[\Phi^{(1)}_{,ij}(\boldsymbol{q},\tau)\right]^2\right\}, \quad (2.70)$$

in which we have again used Eq. (2.59), and similar to that, let $\boldsymbol{s}^{(2)}(\boldsymbol{q},\tau) = \nabla_{\boldsymbol{q}}\Phi^{(2)}(\boldsymbol{q},\tau)$.

The 2LPT solution can be expressed in Eulerian position and velocity, similar to Eqs. (2.61)–(2.62),

$$\boldsymbol{x}(\boldsymbol{q},\tau) = \boldsymbol{q} - \frac{D_1(\tau)}{D_1(\tau_0)}\nabla_{\boldsymbol{q}}\Phi^{(1)}(\boldsymbol{q},\tau_0) + \frac{D_2(\tau)}{D_2(\tau_0)}\nabla_{\boldsymbol{q}}\Phi^{(2)}(\boldsymbol{q},\tau_0) \quad (2.71)$$

$$\boldsymbol{v}(\boldsymbol{q},\tau) = -\frac{D_1(\tau)}{D_1(\tau_0)}f_1(\tau)\mathcal{H}(\tau)\nabla_{\boldsymbol{q}}\Phi^{(1)}(\boldsymbol{q},\tau_0) + \frac{D_2(\tau)}{D_2(\tau_0)}f_2(\tau)\mathcal{H}(\tau)\nabla_{\boldsymbol{q}}\Phi^{(2)}(\boldsymbol{q},\tau_0), \quad (2.72)$$

where we have introduced the second-order logarithmic growth rate f_2. Even better, for the range of $\Omega_{\rm m}(a=1)$ values constrained by Planck CMB measurement [80] and results from LSS surveys [34, 33, 81], the quality of the two fits significantly improves [73].

Eqs. (2.61)–(2.62) and Eqs. (2.71)–(2.72) serve as the backbone of our gravitation forward model in the BORG inference framework, which we introduce in Chapter 4. It is worth

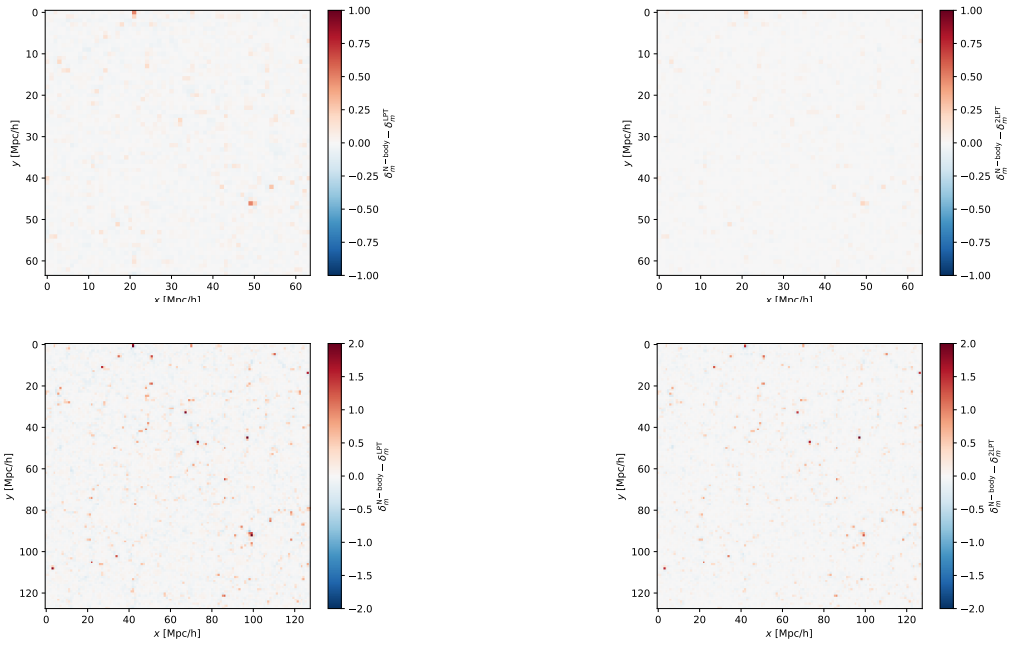

Figure 2.2: The difference between DM density field at redshift zero evolved with GADGET-2 N-body code and that evolved with LPT (left column) or 2LPT (right column), all starting from the same initial conditions. The box size is $L_{\rm box} = 2000\ h^{-1}$Mpc while the total number of particles is $N_{\rm part} = 1536^3$. Particles are assigned to a grid using the CIC kernel with two different smoothing scales, $L_{\rm grid} = 31.25\,h^{-1}$Mpc (top row) and $L_{\rm grid} = 15.625\,h^{-1}$Mpc (bottom row). We show the same slab through the center of the boxes, with a thickness equal to the smoothing scale used in each case.

noting that, in all of our analyses presented in Chapter 5-Chapter 7, the matter density fluctuations are always smoothed with a *cloud-in-cell* (CIC) kernel at fairly large scales of $L_{\rm grid} \sim 31.2 - 15.6\,h^{-1}$Mpc, where perturbative expansion of $\boldsymbol{s}(\boldsymbol{x}, \tau)$ should converge. Thus information encoded in the one- and two-point statistics should be well-preserved by both LPT and 2LPT [72, 76, 82]. In Figure 2.2, we show the difference between the density fields at redshift $z = 0$ evolved by LPT or 2LPT and that evolved by GADGET-2[6] [83], a tree-particle mesh (Tree-PM) N-body simulation code. Both LPT and 2LPT provide a fairly good approximation of the GADGET-2 simulation, especially at the larger smoothing scale. The difference in performance of LPT and 2LPT is only marginal on these smoothing scales. Additionally, we compare, in Figure 2.3, the power spectrum at $z = 0$ of the evolved matter density fields. The fractional difference between LPT (ZA) or 2LPT power spectrum and that of N-body are $\lesssim 5\%$ up to $k = 0.05\,hMpc^{-1}$, and $\lesssim 10\%$ up to $k = 0.1\,hMpc^{-1}$. Further, 2LPT should also reproduce very well the three-point statistics, as shown in [84, 82].

[6] `https://www.mpa.mpa-garching.mpg.de/gadget/`

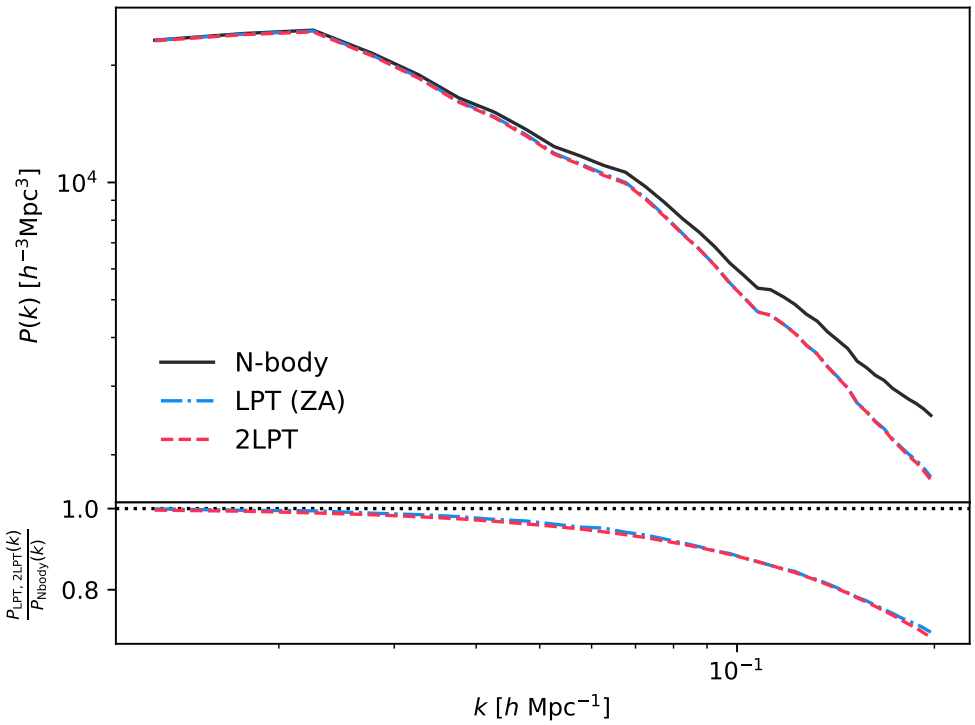

Figure 2.3: The matter power spectra $P_{mm}(k, z = 0)$ (top panel) and their ratio (bottom panel) measured for the same simulations in Figure 2.2. DM particles are assigned to a grid using the CIC kernel with $L_{\rm grid} = 15.625\,h^{-1}$Mpc. Shot noise was subtracted for all cases.

At higher resolutions, to properly generate realistic, physically-constrained realizations of the evolved matter density and velocity fields at low redshift, we would need to employ a particle-mesh (PM) or Tree-PM N-body solver such as the BORG-PMCIC or GADGET-2 used in Chapter 6.

One must be able to tie theoretical predictions on the statistical distribution of evolved matter density fluctuations $\delta_m(\boldsymbol{x}, \tau)$ to observations of galaxies in galaxy redshift surveys or halos in N-body simulations in order to retrieve cosmological information encoded in the evolution of LSS. That critical link is the subject of large-scale galaxy bias, which we turn to in the next chapter.

Chapter 3

Galaxy clustering: biased tracers of large-scale structure

Galaxy[1] clustering is a complicated subject, since it tries to bring together, on one hand, an approximate statistical description of quasi-linear matter fluctuations and, on another hand, distribution of discrete tracers that are results of a highly non-linear, complex formation process [85, 86, 87]. Nevertheless, the last decade has seen many theoretical and numerical advances in the understanding of bias, especially within the perturbative bias expansion framework, matching those discussed in the previous chapter for growth of matter density fluctuations. In this chapter, we review the key ideas and results of the general perturbative bias expansion. We will restrict our discussion to quasi-linear scales and second-order. Motivated readers are encouraged to see [65] for a recent, comprehensive review on this topic.

3.1 The perturbative theory of galaxy clustering: general bias expansion

The ultimate goal of the perturbative bias expansion is to be able to express the galaxy density fluctuations $\delta_g(\boldsymbol{x}, \tau)$ as a general function of large-scale properties of the galaxy's host environment. Such an expansion can be written down in the most general form allowed by principle of general covariance as [65]

$$\delta_g(\boldsymbol{x}, \tau) = \sum_O b_O(\tau) O(\boldsymbol{x}, \tau) + \epsilon(\boldsymbol{x}, \tau) + \sum_O \epsilon_O(\boldsymbol{x}, \tau) O(\boldsymbol{x}, \tau). \tag{3.1}$$

We will refer to the first term on the r.h.s. of Eq. (3.1) as *deterministic bias* in which the operators $O(\boldsymbol{x}, \tau)$ are statistical fields constructed out of the matter density field itself, and the expansion coefficients b_O are referred to as *bias parameters*. This deterministic term is

[1] Here and throuout, we use the words "galaxy" and "halo" (as in DM halo) interchangeably. Whenever not stated explicitly, our discussion applies for both and all biased tracers in general.

the focus of Section 3.1.1. In short, this term predicts the *mean*, or *expected* tracer number density $\overline{n}_g(\boldsymbol{x}, \tau)$ or density contrast $\overline{\delta}_g(\boldsymbol{x}, \tau)$ at the given location. In reality, whether or not a galaxy forms at this specific location depends on small-scale fluctuations in the initial conditions that are not captured in the PT expansion of the matter density field. This randomness introduces stochasticity into the deterministic mean relation above. We refer to the other two terms as *stochastic bias*. These are briefly discussed in Section 3.1.2. The detailed modeling of stochasticity is described in Section 4.1.4.

A non-trivial, remarkable result of restricting ourselves to quasi-linear scales is that, all the complications from complex, unmodeled physics of galaxy formation mentioned earlier can be absorbed into a *finite* number of bias parameters which can then be marginalized over [65]. Analogous to the EFT approach, wherein one would integrate out the small-scale degree of freedoms to obtain an effective theory which correctly describes the phenomena at the pre-determined length scale, the perturbative bias expansion provides the effective large-scale description of the full, complicated dynamics of galaxy formation. Thus, while not being able to extract information from small, highly non-linear scales of LSS evolution, we obtain a rigorous statistical description of galaxy clustering, which in turn allows for unbiased cosmological inference.

3.1.1 Deterministic bias

Local Lagrangian bias: Thresholding toy model

Let us examine how the first term on the r.h.s. of Eq. (3.1) originates from the process of galaxy formation by considering a toy model on a fixed time slice. In this model, galaxies reside inside their host DM halos, which in turn only form at peaks of the initial matter fluctuations linearly extrapolated to the given time τ (cf. Eqs. (2.19)–(2.20)) and filtered on some scale[2] R [51]; we will denote this field as $\delta_R^{(1)}$; a natural choice for R is the Lagrangian radius of the *proto-halo*, such that $R(M_h) = (3M/4\pi\overline{\rho}_m)^{1/3}$.

For a given halo mass M_h, let us define a universal critical threshold density for formation of halo equal and more massive than M_h as $\delta_{\rm cr}$, such that the Lagrangian number density of proto-halos can be written as

$$n_h^L(\boldsymbol{q}) \equiv \Theta_{\rm H}\left(\delta_R^{(1)}(\boldsymbol{q}) - \delta_{\rm cr}\right) \tag{3.2}$$

with $\Theta_{\rm H}$ being the Heaviside step function. We sketch an illustration of this toy model in Figure 3.1.

Recall from Section 2.2.1 and Section 2.2.2 that $\delta_m^{(1)}$ and hence $\delta_R^{(1)}$ are zero-mean Gaussian fields. This implies the statistics of $\delta_R^{(1)}$ are fully captured by the Lagrangian filtered matter two-point correlation function Eq. (2.5)

$$\xi_R^L(r) = \left\langle \delta_R^{(1)}(\boldsymbol{q})\, \delta_R^{(1)}(\boldsymbol{q}+\boldsymbol{r}) \right\rangle, \tag{3.3}$$

[2]See Section A.2 for some common shapes of filter.

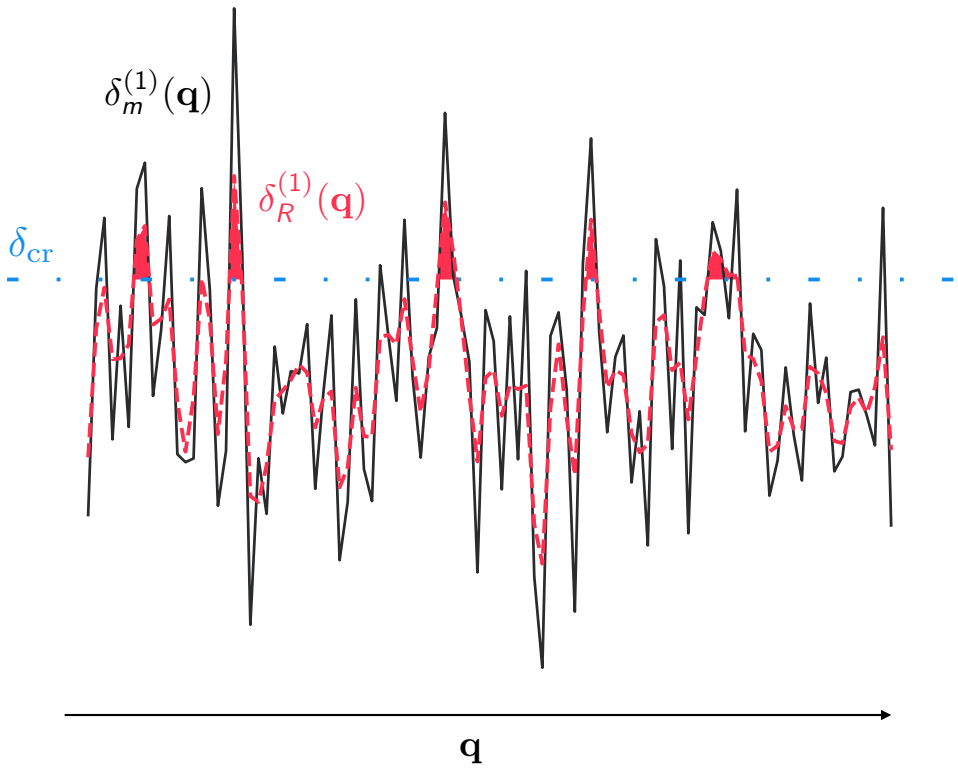

Figure 3.1: Sketch of the "thresholding" toy model considered in Section 3.1.1. The black solid line represents the underlying linear matter density field $\delta_m^{(1)}$; the dashed red line shows the linear density field as filtered by a spherically symmetric Gaussian filter of radius R (cf. Eq. (A.8)), $\delta_R^{(1)}$; the blue dotted-dotted-dashed line indicates the *barrier* set by the critical threshold density $\delta_{\rm cr}$.

where $\lim_{R\to 0}\xi_R^L(r) = \xi^L(r)$ and $\xi^L(0) = \sigma^2(R)$. The expected number density of proto-halos at a given point $\boldsymbol{q}$ is equal to the 1-point probability of the density fluctuations at that point being above the threshold, i.e. $\delta_R^{(1)}(\boldsymbol{q}) \geq \delta_{\rm cr}$,

$$\left\langle n_h^L(\boldsymbol{q})\right\rangle \equiv p_1(\boldsymbol{q}) = \int_{-\infty}^{\infty}\Theta_{\rm H}\left(\nu-\nu_c\right)\frac{1}{\sqrt{2\pi}}e^{-\nu^2/2}d\nu \tag{3.4}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{\nu_c}^{\infty}e^{-\nu^2/2}d\nu \tag{3.5}$$

$$= \frac{1}{2}\,{\rm erfc}\left(\nu_c/\sqrt{2}\right) \tag{3.6}$$

where we have let $\nu \equiv \delta_R^{(1)}(\boldsymbol{q})/\sigma(R)$ and $\nu_c \equiv \delta_{\rm cr}/\sigma(R)$. Similarly, we can define the 2-point probability of finding two proto-halos at $\boldsymbol{q}$ and $\boldsymbol{q}+\boldsymbol{r}$ using the bivariate zero-mean normal distribution (cf. Eq. (2.3))

$$\begin{aligned} p_2(\boldsymbol{q},\boldsymbol{q}+\boldsymbol{r}) &= \frac{1}{2\pi\sqrt{1-\left[\xi_R^L(r)/\sigma^2(R)\right]^2}} \\ &\quad\times\int_{\nu_c}^{\infty}\int_{\nu_c}^{\infty}\exp\left\{-\frac{\nu_1^2+\nu_2^2-2\left(\xi_R^L(r)/\sigma^2(R)\right)\nu_1\nu_2}{2\left[1-\left(\xi_R^L(r)/\sigma^2(R)\right)^2\right]}\right\}d\nu_1 d\nu_2 \\ &= \frac{1}{\sqrt{2}\pi}\int_{\nu_c}^{\infty}e^{-\nu_2^2/2}\int_{u_c}^{\infty}e^{-u_1^2}du_1 d\nu_2 \\ &= \frac{1}{2\sqrt{2\pi}}\int_{\nu_c}^{\infty}e^{-\nu^2/2}\,{\rm erfc}\left[\frac{\nu_c-\nu\hat{\xi}(r)}{\sqrt{2\left[1-\hat{\xi}^2(r)\right]}}\right]d\nu. \end{aligned} \tag{3.7}$$

where, in the second line, we have replaced $\hat{\xi}(r) \equiv \xi_R^L(r)/\sigma^2(R)$ and applied a change of variable $u_1 = \dfrac{\nu_1-\nu_2\hat{\xi}(r)}{\sqrt{2\left[1-\hat{\xi}^2(r)\right]}}$; in the last line, we have integrated over u_1 and, for simplification, written $\nu \equiv \nu_2$.

The corresponding proto-halo two-point correlation function is then given by [88, 89, 90]

$$\begin{aligned} \xi_h^L(r) &= \frac{p_2(\boldsymbol{q},\boldsymbol{q}+\boldsymbol{r})}{p_1(\boldsymbol{q})^2}-1 \\ &= \sqrt{\frac{2}{\pi}}\left[{\rm erfc}\left(\nu_c/\sqrt{2}\right)\right]^{-2}\int_{\nu_c}^{\infty}e^{-\nu^2/2}\,{\rm erfc}\left[\frac{\nu_c-\nu\hat{\xi}(r)}{\sqrt{2\left[1-\hat{\xi}^2(r)\right]}}\right]-1, \end{aligned} \tag{3.8}$$

which, for small values of $\xi_R^L(r)$, can be rewritten in the series form [89, 90, 65]

$$\begin{aligned}\xi_h^L(r) &= \sum_{N=1}^{\infty} \frac{1}{N!} \left\{ \sqrt{\frac{2}{\pi}} \left[\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^{-1} \frac{e^{\nu_c^2/2}}{\sigma^N(R)} H_{N-1}(\nu_c) \right\}^2 \left[\xi_R^L(r)\right]^N \\ &= \sum_{N=1}^{\infty} \frac{1}{N!} \left(b_N^L\right)^2 \left[\xi_R^L(r)\right]^N, \end{aligned} \tag{3.9}$$

where H_N denotes the Hermite polynomials. Let us consider the large-scale limit, i.e. $\lim_{r\to\infty} \xi_R^L(r) \to 0$, so that, up to second-order,

$$\xi_h^L(r) = \left(b_1^L\right)^2 \xi_R^L(r) + \frac{1}{2}\left(b_2^L\right)^2 \left[\xi_R^L(r)\right]^2 + \mathcal{O}\left(\left[\xi_R^L(r)\right]^3\right). \tag{3.10}$$

At very large scales, $r \gtrsim 30\ h^{-1}$Mpc, the proto-halo correlation function has the same behavior as the (filtered) matter correlation function, only that the amplitude of the former is enhanced or *biased* by a factor of $\left(b_1^L\right)^2$. As we reach smaller scales, $r \sim 20-30\ h^{-1}$Mpc, the contribution from the second-order term will become important. Eventually, at even smaller scales, $r \lesssim R_{\rm NL} \sim 10-20\ h^{-1}$Mpc (for redshift zero) [65], our perturbative description for the biased tracer and the evolved matter density fields no longer converges to the correct result, as higher-order terms are no longer smaller than lower-order terms. Note that $R_{\rm NL}$ becomes smaller at higher redshift [91]. We thus expect our perturbative approach to be able to access information at scales $r \sim 10-20\ h^{-1}$Mpc in deep galaxy surveys, while keeping theoretical error under rigorous control.

The b_N^L in Eq. (3.9) are referred to as Lagrangian *deterministic bias* parameters. Historically, Eq. (3.9) was also obtained through a series expansion of $\delta_h^L(\boldsymbol{q})$

$$\delta_h^L(\boldsymbol{q}) = b_1^L \delta_R^{(1)}(\boldsymbol{q}) + \frac{1}{2} b_2^L \left(\left[\delta_R^{(1)}(\boldsymbol{q})\right]^2 - \sigma^2(R)\right) + \mathcal{O}\left(\left[\delta_R^{(1)}(\boldsymbol{q})\right]^3\right) \tag{3.11}$$

Throughout this thesis, we adopt the convention of [65] and will refer to local bias relations of the form of Eq. (3.11) as *local-in-matter-density* (LIMD) bias. This notion stems from the fact that the proto-halo density is described by local functions of nothing else but the initial matter density field itself. Below, we will see that, as we take into account the time evolution and non-locality of galaxy formation, the LIMD bias relation in Eq. (3.11) is not complete, and we need to introduce two more terms, namely the tidal and leading higher-derivative terms, in order to complete our deterministic bias expansion at second-order.

Local Eulerian bias: Time evolution of bias relation for conserved tracers

In the previous section, we have arrived at the second-order deterministic local bias relation in Lagrangian space through the consideration of a toy model – which is a simplified version of the so-called "peak-background-split" model [51] – on a fixed time slice. Let us now follow the evolution of a galaxy sample – experiencing only gravitational effects – between two fixed time slices, the time slice of formation τ_* and the time slice of observation τ,

as illustrated in Figure 3.2. Our goal is to express the Eulerian local bias parameters at observation time, $b_n^E(\tau)$, in terms of the local bias parameters at formation time, $b_n^* \equiv b_n^L$. For the simplicity of our arguments and calculations below, we will assume that:

1. all galaxies in our sample simultaneously and instantaneously formed at time τ_*, and

2. gravitational evolution preserves the number of galaxies between two time slices.

Since a realistic galaxy sample in which galaxies might form and merge at different times can be considered as a superposition of many of our simplified samples, it is straightforward to generalize our results for the latter to describe the former. Indeed,

$$b_O(\tau) = \int b_O^E(\tau|\{b_{O'}^*\},\tau_*)p(\tau_*)d\tau_*, \tag{3.12}$$

where $p(\tau_*)$ denotes the normalized distribution function of galaxy formation times in the realistic sample and $b_O^E(\tau|\{b_{O'}^*\},\tau_*)$ denotes the Eulerian bias at time τ given a set of bias parameters $\{b_{O'}^*\}$ measured at each formation time τ_* [65].

Our starting point is then the continuity equation that governs the time evolution of such conserved tracers [92, 93, 94, 65]

$$\frac{D}{D\tau}\delta_g(\boldsymbol{x},\tau) = -\theta\left[1+\delta_g(\boldsymbol{x},\tau)\right], \tag{3.13}$$

where $\mathrm{D}/\mathrm{D}\tau \equiv \dfrac{\partial}{\partial\tau} + v^i\dfrac{\partial}{\partial x^i}$ denotes the Lagrangian time derivative[3], while v^i and $\theta = \nabla_{\boldsymbol{x}}\cdot\boldsymbol{v}$ are the peculiar velocity and the velocity divergence of the cosmic matter fluid, respectively. Note that this relation holds on large scales as we have assumed earlier that baryonic gas and galaxies comove with the CDM fluid (See Section 2.7 of [65] for a detailed discussion about velocity bias between galaxies and matter).

In order to solve for $\delta_g(\boldsymbol{x},\tau)$, we would need to integrate Eq. (3.13) along the fluid trajectory (black solid line in Figure 3.2) while making use of the PT solutions of δ_m and v^i (see Section 2.2.1 and Section 2.2.2), which we rewrite here with the Lagrangian derivative

$$\frac{\mathrm{D}}{\mathrm{D}\tau}\delta_m(\boldsymbol{x},\tau) = -\theta\left[1+\delta_m(\boldsymbol{x},\tau)\right], \tag{3.14}$$

$$\frac{\mathrm{D}}{\mathrm{D}\tau}\theta(\boldsymbol{x},\tau) = \mathcal{H}(\tau)\theta(\boldsymbol{x},\tau) - \left[\frac{\partial}{\partial x^i}v^j(\boldsymbol{x},\tau)\right]^2 - \frac{3}{2}\Omega_{\mathrm{m}}(\tau)\mathcal{H}^2\delta_m(\boldsymbol{x},\tau). \tag{3.15}$$

Dividing Eq. (3.13) and Eq. (3.14) by $(1+\delta_g)$ and $(1+\delta_m)$, respectively, yields

$$\frac{1}{1+\delta_g}\frac{\mathrm{D}}{\mathrm{D}\tau}\delta_g = \frac{1}{1+\delta_m}\frac{\mathrm{D}}{\mathrm{D}\tau}\delta_m = -\theta. \tag{3.16}$$

[3]Some texts also refer to this as convective or material derivative

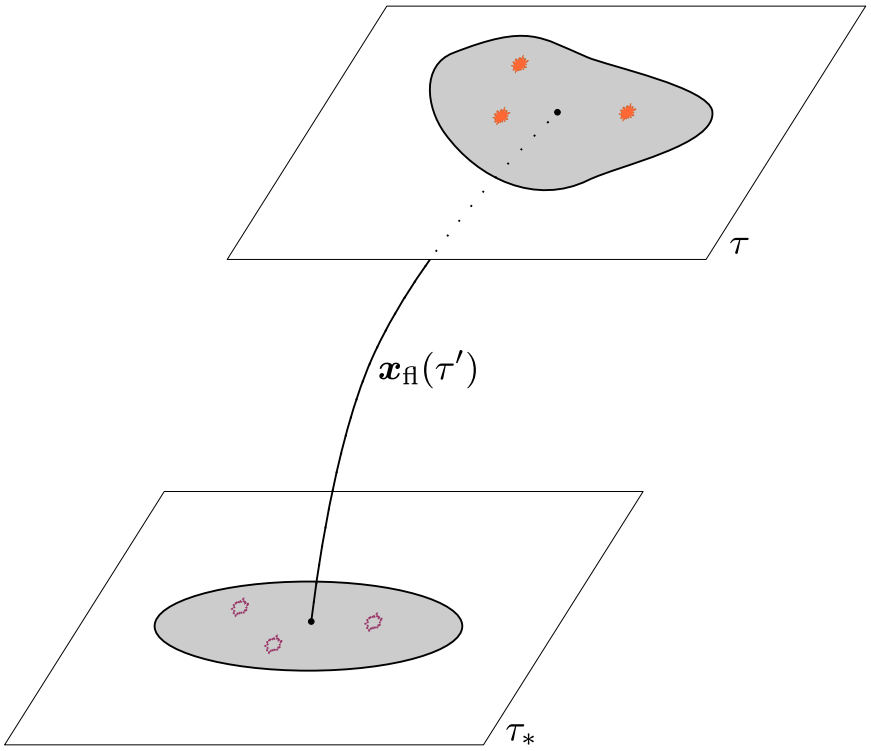

Figure 3.2: Sketch of the setup considered in Section 3.1.1, taken from [65]. Galaxies formed instantaneously at $\tau = \tau_*$, where they are described by the initial Lagrangian bias relation Eq. (3.11) at time slice τ_*. After formation, they comove with matter and evolve up to the observation time slice τ. The grey region denotes the Lagrangian volume encompassing these galaxies which gets deformed by non-linear gravitational evolution. Since we assume galaxies are assumed to comove with matter and their number to be conserved, their density is similarly affected.

Let us then write Eq. (3.16) in Lagrangian space (cf. Eq. (2.41)) where $\mathrm{D}/\mathrm{D}\tau \equiv \frac{\partial}{\partial\tau}$ and we can straightforwardly integrate both sides (along the fluid trajectory)

$$\ln\left[1+\delta_g(\boldsymbol{x}_{\rm fl}(\tau),\tau)\right] = \ln\left[1+\delta_m(\boldsymbol{x}_{\rm fl}(\tau),\tau)\right] + \ln\left[\frac{1+\delta_g(\boldsymbol{x}_{\rm fl}(\tau_*),\tau)}{1+\delta_m(\boldsymbol{x}_{\rm fl}(\tau_*),\tau)}\right], \tag{3.17}$$

which leads to

$$1+\delta_g\Big|_\tau = \frac{1+\delta_m|_\tau}{1+\delta_m|_{\tau_*}}(1+\delta_g|_{\tau_*}) \tag{3.18}$$

wherein $|_\tau$ implies that the corresponding quantity is evaluated on time slice τ and all quantities are evaluated along the same fluid trajectory. The simple result in Eq. (3.18) can be viewed as a direct consequence of general covariance principle, two comoving observers will observe no change in the ratio between the density of their fluid and that of the other's. By letting $\tau_* \to 0$ so that $\delta_* \to 0$ and $\delta_{g*} = \delta_g^L(\boldsymbol{q})$, we will retrieve the relation between Eulerian and Lagrangian bias in the special case of formation at $z_* = 0$ [65]

$$1+\delta_g\left(\boldsymbol{x}_{\rm fl}[\boldsymbol{q},\tau],\tau\right) = \left[1+\delta_m\left(\boldsymbol{x}_{\rm fl}[\boldsymbol{q},\tau],\tau\right)\right]\left[1+\delta_g^L(\boldsymbol{q})\right]. \tag{3.19}$$

To derive the Eulerian deterministic bias relation in general case, complete up to second-order, we need to solve Eq. (3.18) up to the corresponding order in PT. That is, we are going to expand δ_g, δ_m in Eq. (3.18) while keeping up to quadratic terms

$$\begin{aligned}
1+\delta_g^{(1)}+\delta_g^{(2)} &= \left[1+\delta_m^{(1)}+\delta_m^{(2)}\right]\left[1-(\delta_{m*}^{(1)}+\delta_{m*}^{(2)})+(\delta_{m*}^{(1)})^2\right]\left[1+(\delta_{g*}^{(1)}+\delta_{g*}^{(2)})\right] \\
&= 1+\left\{\delta_m^{(1)}-\delta_{m*}^{(1)}+\delta_{g*}^{(1)}\right\} \\
&\quad +\left\{\delta_m^{(2)}-\delta_{m*}^{(2)}+\delta_{g*}^{(2)}+(\delta_{m*}^{(1)})^2-\delta_m^{(1)}\delta_{m*}^{(1)}+\left[\delta_m^{(1)}-\delta_{m*}^{(1)}\right]\delta_{g*}^{(1)}\right\},
\end{aligned} \tag{3.20}$$

where we have used the notation $f_* \equiv f(\boldsymbol{x}_*,\tau_*)$ while $f \equiv f(\boldsymbol{x},\tau)$, and grouped the first- and second-order terms in the second and third line line, respectively. Note that the difference between $\boldsymbol{x}_*$ and $\boldsymbol{x}$ is itself first-order in LPT, as indeed suggested by Eq. (2.41) and Eq. (2.56)

$$\boldsymbol{x}_* \equiv \boldsymbol{x}_{\rm fl}(\tau_*) = \boldsymbol{q}+\boldsymbol{s}(\boldsymbol{q},\tau_*) = [\boldsymbol{x}-\boldsymbol{s}(\boldsymbol{q},\tau)]+\boldsymbol{s}(\boldsymbol{q},\tau_*) = \boldsymbol{x}+\left(\frac{D_*}{D}-1\right)\boldsymbol{s}^{(1)}(\boldsymbol{x},\tau), \tag{3.21}$$

where $D_*/D \equiv D(\tau_*)/D(\tau)$. This implies, going up to second-order, we only have to distinguish between $\boldsymbol{x}$ and $\boldsymbol{x}_*$ among the first-order terms of Eq. (3.20). In specific, we must expand $\delta_m^{(1)}(\boldsymbol{x}_*,\tau_*)$ around $\boldsymbol{x}$,

$$\begin{aligned}
\delta_m^{(1)}(\boldsymbol{x}_*,\tau_*) &= \delta_m^{(1)}(\boldsymbol{x},\tau_*)+(\boldsymbol{x}_*-\boldsymbol{x})\left.\frac{\partial\delta_m^{(1)}}{\partial x^i}\right|_{(\boldsymbol{x},\tau_*)} \\
&= \frac{D_*}{D}\delta_m^{(1)}(\boldsymbol{x},\tau)+\left(\frac{D_*}{D}-1\right)\left(s^{(1)}\right)^i(\boldsymbol{x},\tau)\partial_i\delta_m^{(1)}(\boldsymbol{x},\tau).
\end{aligned} \tag{3.22}$$

The only missing piece on the r.h.s. of Eq. (3.20) now is an expression for $\delta_{g*}^{(1+2)} = \delta_{g*}^{(1)} + \delta_{g*}^{(2)}$. Let us then write down the most general bias relation at second-order, as allowed by Eq. (3.1),

$$\begin{aligned}\delta_{g*}^{(1+2)} &= b_1^* \left[\delta_{m*}^{(1)} + \delta_{m*}^{(2)}\right] + \frac{1}{2} b_2^* \left[\delta_{m*}^{(1)}\right]^2 + b_{K^2}^* \left[K_{ij*}^{(1)}\right]^2 + \epsilon^*, \\ &= b_1^* \left[\left(\frac{D_*}{D}\right) \delta_m^{(1)} + \left(\frac{D_*}{D}\right)^2 \delta_m^{(2)}\right] + \frac{1}{2} b_2^* \left(\frac{D_*}{D}\right)^2 \left[\delta_m^{(1)}\right]^2 + b_{K^2}^* \left(\frac{D_*}{D}\right)^2 \left[K_{ij}^{(1)}\right]^2 + \epsilon^*,\end{aligned} \tag{3.23}$$

where we have introduced the tidal field in Eq. (2.40) and included the leading stochastic bias term ϵ^* – assumed to be uncorrelated with the matter density fluctuations δ_m (see Section 3.1.2 for more details). Let us now compare Eqs. (3.22)–(3.23) to Eq. (3.20) so that we can collect first-order terms [65]

$$\delta_{g*}^{(1)}(\boldsymbol{x}, \tau) = \left[1 + \frac{D_*}{D}\left(b_1^* - 1\right)\right] \delta_m^{(1)}(\boldsymbol{x}, \tau) \; + \; \epsilon^*, \tag{3.24}$$

and second-order terms [65]

$$\begin{aligned}\delta_{g*}^{(2)}(\boldsymbol{x}, \tau) = &\left[1 + \left(b_1^* - 1\right)\left(\frac{D_*}{D}\right)^2\right] \delta_m^{(2)} \\ &+ \left[\frac{D_*}{D}\left(b^* - 1\right) - \left(\frac{D_*}{D}\right)^2 \left(b_1^* - 1\right) \; + \; \frac{1}{2} b_2^* \left(\frac{D_*}{D}\right)^2\right] \left[\delta_m^{(1)}\right]^2 \\ &+ \; b_{K^2}^* \left(\frac{D_*}{D}\right)^2 \left[K_{ij}^{(1)}\right]^2 \; + \; \left(\frac{D_*}{D} - 1\right) \frac{D_*}{D} \left(b_1^* - 1\right) \left(s^{(1)}\right)^i \partial_i \delta_m^{(1)} \\ &- \left(\frac{D_*}{D} - 1\right) \epsilon^* \delta_m^{(1)} \; + \; \left(\frac{D_*}{D} - 1\right) \left(s^{(1)}\right)^i \partial_i \epsilon^*.\end{aligned} \tag{3.25}$$

By combining Eqs. (3.24)–(3.25), we can finally express the second-order galaxy density contrast as [65]

$$\begin{aligned}\delta_g^{(1)+(2)} = &\, b_1^E \left[\delta_m^{(1)} + \delta_m^{(2)}\right] + \frac{1}{2} b_2^E \left[\delta_m^{(1)}\right]^2 + b_{K^2}^E \left[K_{ij}^{(1)}\right]^2 \\ &+ \epsilon^* - \left(\frac{D_*}{D} - 1\right) \epsilon^* \delta_m^{(1)} + \left(\frac{D_*}{D} - 1\right) \left(s^{(1)}\right)^i \partial_i \epsilon^*\end{aligned} \tag{3.26}$$

with

$$b_1^E(\tau) = 1 + \frac{D_*}{D}\left(b_1^* - 1\right), \tag{3.27}$$

$$b_2^E(\tau) = \left(\frac{D_*}{D}\right)^2 b_2^* + \frac{8}{21}\left(1 - \frac{D_*}{D}\right)\left(b_1^E - 1\right), \tag{3.28}$$

$$b_{K^2}^E(\tau) = \left(\frac{D_*}{D}\right)^2 b_{K^2}^* - \frac{2}{7}\left(1 - \frac{D_*}{D}\right)\left(b_1^E - 1\right). \tag{3.29}$$

Eq. (3.26) is the general Eulerian local bias expansion (up to second-order) as derived from general considerations of a conserved galaxy sample exclusively undergoing gravitational evolution, including both deterministic (first line) and stochastic (second line) contributions [65]. We will discuss the latter further in Section 3.1.2. For the rest of this section, let us focus on the former, which reads

$$\left\langle \delta_g^{(1)+(2)} \right\rangle = b_1^E \left[\delta_m^{(1)} + \delta_m^{(2)}\right] + \frac{1}{2} b_2^E \left[\delta_m^{(1)}\right]^2 + b_{K^2}^E \left[K_{ij}^{(1)}\right]^2 . \tag{3.30}$$

Eulerian non-local bias: higher-derivative terms

To complete the deterministic bias relation, we need to go only one more step further – we must account for the fact that halos and galaxies are formed by collapsed matter, not at one specific spatial point, but within a *finite* region of space. In fact, the galaxy formation process might also be affected by baryonic gas pressure within the neighboring region (see Figure 3.3). This can be done by introducing the convolution [95, 65]

$$\begin{aligned} b_{\delta_m}(\tau)\delta_m(\boldsymbol{x},\tau) &\to \left[\int d^3y F_{\delta_m}(\boldsymbol{y},\tau)\right] \delta_m(\boldsymbol{x}+\boldsymbol{y},\tau) \\ &\to \left[\int_{\boldsymbol{y}} F_{\delta_m}(y,\tau)\right] \delta_m(\boldsymbol{x},\tau) + \left[\frac{1}{6}\int_{\boldsymbol{y}} |\boldsymbol{y}|^2 F_{\delta_m}(y,\tau)\right] \nabla_x^2 \delta_m(\boldsymbol{x},\tau) + \dots \\ &= b_{\delta_m}(\tau)\delta_m(\boldsymbol{x},\tau) + b_{\nabla^2\delta_m}(\tau)\nabla_x^2\delta_m(\boldsymbol{x},\tau) + \dots , \end{aligned} \tag{3.31}$$

where we have employed the earlier assumptions of statistical homogeneity and isotropy in the second line. As can be seen from Eq. (3.31) and Figure 3.3, this contribution is referred to as *higher-derivative* bias term since it involves more than two derivatives acting on Φ ($\nabla^2\delta_m \equiv \nabla^2\nabla^2\Phi$). As each spatial derivative is multiplied by R_* – the spatial scale set by galaxy formation process, $|b_{\nabla^2\delta_m}| \sim R_*^2$ and this contribution is sub-leading in the limit of very long-wavelength perturbations [65].

With the addition of the leading higher-derivative bias term $b_{\nabla^2\delta_m}\nabla^2\delta_m$, we have completed the deterministic bias expansion up to second-order,

$$\left\langle \delta_g^{(1)+(2)} \right\rangle = b_1^E \left[\delta_m^{(1)} + \delta_m^{(2)}\right] + \frac{1}{2} b_2^E \left[\delta_m^{(1)}\right]^2 + b_{K^2}^E \left[K_{ij}^{(1)}\right]^2 + b_{\nabla^2\delta_m}^E \nabla^2\delta_m . \tag{3.32}$$

Hereafter, we will refer to any specific choice for this deterministic relation as a *bias model*.

3.1.2 Stochastic bias

As mentioned before, small-scale modes – which are not correlated with large-scale modes for Gaussian initial conditions – still contribute to galaxy formation process, introducing stochasticity in the galaxy-matter bias relation. In the previous section, we have seen that this stochastic contribution and its coupling effect with gravitational evolution can

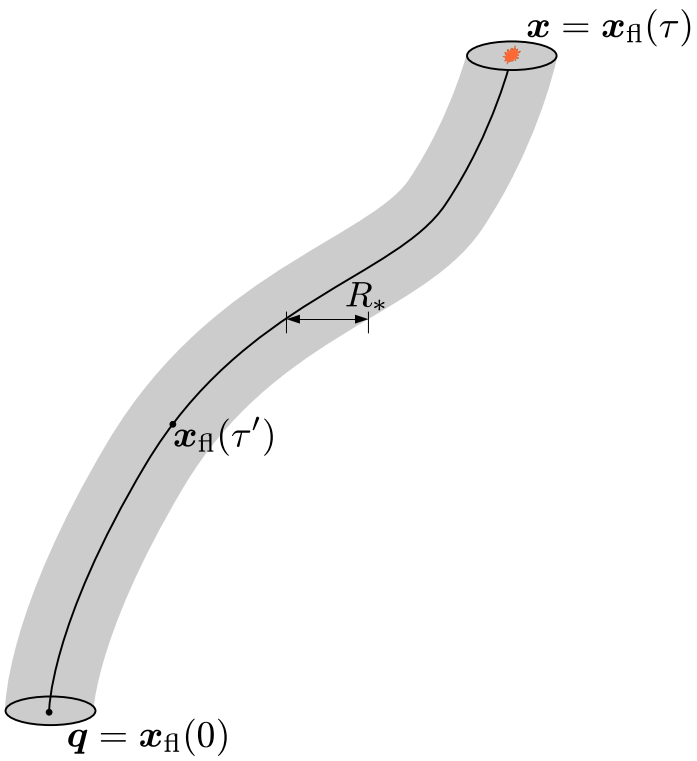

Figure 3.3: Sketch of galaxy formation process and the spatial region involved, taken from [65]. The black solid line represents the fluid trajectory starting from the initial Lagrangian position $\boldsymbol{q} = \boldsymbol{x}_{\rm fl}(\tau = 0)$ and ending at the final (observed) Eulerian position $\boldsymbol{x} = \boldsymbol{x}_{\rm fl}(\tau)$; the gray-shaded region illustrates the region within which matter distribution is relevant for the galaxy formation process.

be captured by terms involving ϵ_0 in Eq. (3.1) or ϵ^* in Eq. (3.26). In general, we can write [96, 65]

$$\epsilon_g(\boldsymbol{x}, \tau) = \epsilon(\boldsymbol{x}, \tau) + \sum_O \epsilon_O(\boldsymbol{x}, \tau) O(\boldsymbol{x}, \tau) \tag{3.33}$$

where ϵ and ϵ_O are uncorrelated with O, but correlated among themselves. Let us next highlight the behavior of the stochasticity power spectrum in two limits.

On one hand, in the large-scale limit, i.e. $k \ll 1/R_*$, the noise fields are Gaussian and analytic in k [97, 65]

$$\langle \epsilon_O(\boldsymbol{k}) \epsilon_{O'}(\boldsymbol{k}) \rangle' \equiv P^n_{\epsilon_O, \epsilon_{O'}}(k) = \sum_{n=0,2,4,\ldots} P^n_{\epsilon_O \epsilon_{O'}} k^n. \tag{3.34}$$

On the other hand, for discrete tracers, in the small-scale limit, i.e. $k \gg 1/R_*$, the noise field presumably follows *Poisson distribution* [96] such that

$$P^0_{\epsilon\epsilon} = \frac{1}{\overline{n}_g}, \tag{3.35}$$

where $\overline{n}_g$ is the tracer mean comoving number density.

Within the BORG framework, instead of directly sampling the stochastic bias parameters ϵ and ϵ_O, i.e. treating them equally with the deterministic bias parameters b_O, the scatter between the mean-field predicted galaxy density contrast $\delta_{g,\mathrm{det}}$ and the observed galaxy density contrast δ_g is modeled with the *conditional probability* or, for short, *likelihood*. This likelihood will be the focus of Section 4.1.4. At this point, it is worth to note that, on intermediate, quasi-linear scales, the stochasticity power spectrum deviates from the Poisson case which complicates our likelihood.

3.2 Observation of galaxy clustering: galaxy redshift surveys and galaxy clusters

While we exclusively work with DM halos in N-body simulations in Chapter 5 and Chapter 7, the work in Chapter 6 is built on results of BORG reconstruction of the SDSS-III/BOSS volume using BOSS LOW-Z and CMASS galaxy samples. In addition, we cross-correlate this data with the maxBCG galaxy cluster data. Thus in this section, we briefly review important observation effects in galaxy redshift surveys, as well as measurements of the mass-richness scaling relation for galaxy clusters. Those will be especially relevant for our discussion in Chapter 6.

3.2.1 Galaxy redshift survey

Galaxy redshift surveys attempt to map out the 3D distribution of galaxies by recording both their sky position and redshift.

Optimally, galaxy redshift should be measured from position of emission or absorption lines feature in galaxy spectra. This is known as *spectroscopic* redshift. A cheaper alternative is estimating galaxy redshift using multi-band photometry of which can be thought as low-resolution spectroscopy. *Photometric* redshift methods rely on measuring the signal in the photometric data arising from prominent "break" features in galaxy spectra, e.g. the 4000 Å break in early-type, red galaxies or the Lyman break at 912 Å in star-forming, blue galaxies.

Survey geometry and selection effects

In an N-body simulation, halos are uniformly sampled within the simulation box. In contrast, redshift surveys are generally flux-limited and masked. The two effects – often grouped together as "observational selection effects" – imply that galaxies which are, respectively, fainter than a specified limiting observable flux and within a particular "masked" solid angle on the sky will be excluded from the survey. This results into a *non-uniform* sampling of the galaxy distribution within the survey volume which must be taken into account during the galaxy clustering modeling process. In particular,

1. the *non-uniform* sampling of the 2D sky – be it due to survey geometry, foreground contaminations, instrumental aspects, operational conditions, etc. – is described by the *angular selection function* or *angular completeness mask* $\psi_a(\alpha, \delta)$ [98, 99];

2. the increasingly *sparse* sampling along the radial direction in flux-limited surveys is encapsulated in the *radial selection function* or *redshift completeness mask* $\psi_r(z) = \frac{\int_{L(z)}^{\infty} \Phi(L)}{\int_{L_0}^{\int} \Phi(L)}$[4] [100, 98, 99].

Both effects result in an underestimate of the true galaxy field. To correct for these variations in sky completeness and redshift completeness, the estimated density at any given point i must be, in general, weighted by the inverse of the total selection function [101, 99]

$$\psi^{-1}(\boldsymbol{x}) = \left(\psi_a(\alpha, \delta)\psi_r(z)\right)^{-1} . \tag{3.36}$$

Eq. (3.36) defines the probability of a galaxy at a given sky position and redshift meets the selection criteria to be registered in the survey sample.

As we will see in Chapter 4, these effects simply amount to a multiplication when galaxy clustering is directly modeled at the field-level [102, 43].

Redshift-space distortion

Yet another important observational effect is redshift-space distortion (RSD). This complication arises from the nature of our observation, mentioned at the beginning of this

[4] $\Phi(L)$ is the galaxy luminosity function defined such that $\Phi(L)dL$ is proportional to the number density of galaxies within luminosity L and $L + dL$.

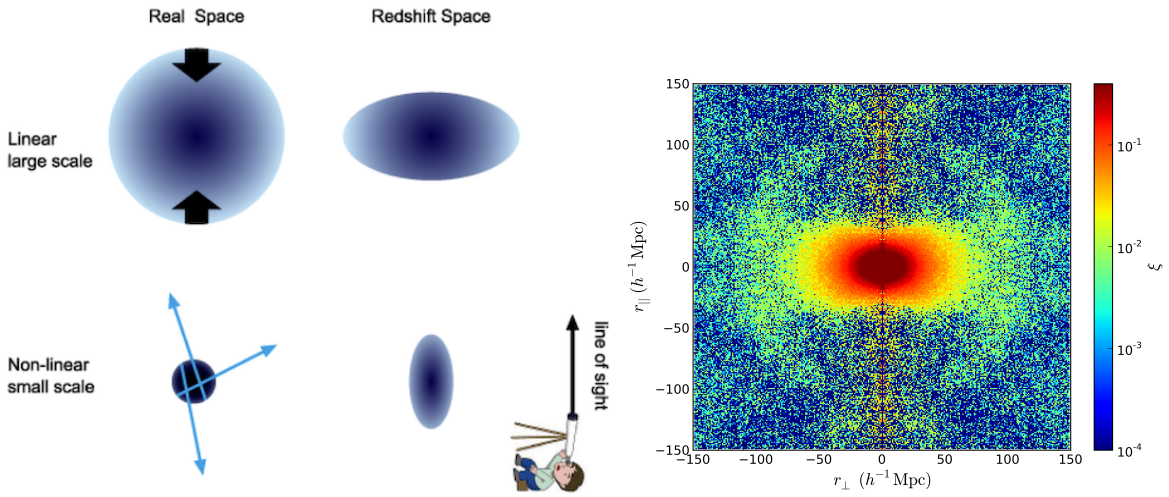

Figure 3.4: Left: A cartoon illustrates the redshift-space distortion effects that squash and stretch the galaxy clustering on, respectively, large and small scales of separation. Cartoon courtesy of my friends, Kimika and Shun Saito. Right: The 2D correlation function ξ – measured in bins of $1\,h^{-1}$Mpc $\times$ $1\,h^{-1}$Mpc – plotted as a function of separation along and perpendicular to the LOS, $r_{\parallel}$ and $r_{\perp}$. The color scale presents the amplitude of ξ. As visible from this plot, RSD introduces distinctive anisotropic features in the galaxy clustering amplitude. This plot is taken from [104].

section, that redshift serves as the only proxy for radial distance. Naturally, all cosmological observables that include radial distance are indeed measured in redshift-space, which, unfortunately is distorted by the line-of-sight (LOS) component of galaxies' peculiar motion in two prominent ways:

1. On large scales, the coherent motion of matter, and hence galaxies, towards an high density region induces a linear squashing effect which enhances the galaxy clustering amplitude. This is the so-called Kaiser effect [88].

2. At smaller scales, collapse and virialization processes give rise to random motions and the so-called *fingers-of-god* [103] which diminishes the clustering amplitude.

.

Notably, [88] showed that, at linear order, the Fourier amplitude of galaxy density contrast $\delta_g^{(s)}(\boldsymbol{k})$ in redshift-space is amplified over the Fourier amplitude of matter density contrast $\delta_m(\boldsymbol{k})$ in real-space by a factor

$$\delta_g^{(s)}(\boldsymbol{k}) = \left(b_1 + f_1\mu_{\boldsymbol{k}}^2\right)\delta_m(\boldsymbol{k}) \tag{3.37}$$

where $\mu_{\boldsymbol{k}} = \hat{\boldsymbol{z}}\cdot\hat{\boldsymbol{k}}$ is the cosine of the angle between the unit wavevector $\hat{\boldsymbol{k}}$ and the LOS $\hat{\boldsymbol{z}}$, while b_1 and f_1 are, respectively, the linear bias parameter and the logarithmic growth rate seen in Section 3.1 and Section 2.2. As one attempts to go to higher-order, RSD quickly

complicates one's modeling of the correlation function or power spectrum [105, 106, 107, 108]. As we will see in Chapter 4, RSD can be naturally accounted for within any forward modeling framework that incorporate gravitational dynamics.

3.2.2 Galaxy clusters from optical galaxy survey

Clusters of galaxies are arguably one of the most prominent features of LSS. These massive, gravitationally-bound objects are likely the observational counterpart of DM halos in N-body simulations; hence their observables play an important role in many cosmological probes[5] [110, 111, 112]. In particular, we will see in Chapter 6 that, the temperature anisotropies imprinted on the CMB by galaxy clusters could potentially provide a constraint on the abundance of baryons in the Universe. Below we briefly review how a cluster catalog and cluster mass proxy can be constructed from an *optical* galaxy survey, focusing on the concrete example of the maxBCG catalog [42] whose data we employ in Chapter 6.

Identification of galaxy clusters

The first cluster catalog constructed out of observations in optical wavelengths was that of [113]. The Abell catalog include clusters identified simply by visual inspection of photographic plates from the Palomar Observatory Sky Survey. Thenceforth, optically selected cluster catalogs have played a paramount role in cluster cosmology. Although this method – being sensitive to projection effects – faces the problem of finding a robust mass proxy that can minimize scatter in the scaling relation , it has multiple advantages as it is able to

1. extract clusters from the data of ground-based galaxy surveys,

2. reach further down the low mass end of galaxy clusters.

The above factors contribute, respectively, to the abundance, the vast amount of supplemental information and the large sample size of optically-selected cluster catalogs. With the exception of the adaptive matched filtering method adopted in [114], most optically-selected cluster catalogs have been constructed essentially based on the color information of bright, red-sequence galaxies [115, 42, 116, 117, 118]. By virtue of the 4000 Å break feature in their rest-frame spectra, these galaxies' colors are tightly correlated with their redshifts such that their color measurements presumably yield a reasonably accurate estimate of their redshifts. To be able to cover a broad range of redshifts, multi-color photometry is desired for tracking the intrinsic 4000 Å break feature of old stellar populations as it reddens. The five-band photometry of SDSS allows for such a selection. The maxBCG algorithm [115] utilizes a hierarchical maximum likelihood method – designed for g-r and r-i color data – on the assumption that red, brightest cluster galaxy (BCG) typically resides at the center of the cluster's member galaxy distribution, relatively at rest with the cluster and host halo center (hence the name maxBCG). Application of this cluster-finding method

[5] See [109] for a more detailed review.

on the SDSS optical imaging data yields a *volume-limited* catalog consists of 13823 galaxy clusters – each assigned with the photometric redshift that maximizes its BCG likelihood – covering 7500 deg^2 of the sky, spanning a redshift range of $z = 0.1 - 0.3$, being $> 90\%$ pure[6] for clusters whose $N_{200} > 10$ [42].

Proxies for galaxy cluster mass

Equally important to identifying the clusters is identifying a robust proxy to estimate cluster mass as X-ray, optical or weak lensing direct measurement of individual cluster's mass is often limited by available observing time and financial resources. The most common proxies are

1. optical richness,
2. X-ray luminosity,
3. integrated SZ flux.

For maxBCG clusters, a natural choice would be the optical richness, N_{200}, the number of galaxies within a radius R_{200} from the BCG where R_{200} is defined as the radius within which the local density contrast is 200 times higher than the critical density at that redshift. This quantity is automatically generated by the algorithm. The mean mass-richness relation of maxBCG clusters was first measured using weak lensing data [119, 120]. In [120], the cluster three-dimensional mean density and mass profile are first inferred from non-parametric inversions of the two-dimensional lensing shear profiles – measured by [119] – in twelve bins of optical richness extending all the way down to $N_{200} = 3$ and corrected for systematic errors. The mass-richness scaling relation is then obtained as a power-law fit of the form

$$M_{200}(N_{200}) = M_{200|20} \left(\frac{N_{200}}{20} \right)^{\alpha} \tag{3.38}$$

to the data which prefers a power-law index of $\alpha = 1.28 \pm 0.04$ [120]. Here, M_{200} denotes the projected total cluster mass within the radius R_{200}.

Later, by combining the above measurement and the X-ray measurement by [121][7], [122] was able to constrain the scatter in the mean $M - N_{200}$ relation as $\sigma_{M|N_{200}} = 0.45^{+0.20}_{-0.18}$ (95% CL) at $N_{200} \approx 40$ with a modest inferred non-Gaussian tail toward low masses [122]. For our analysis in Chapter 6, we adopt the $M_{180b} - N_{200}$ relation given by Eq. (A15) in [122] which presumably accounts for systematic bias induced by photometric redshift error in

[6]Tests of purity, i.e. false-positive rate, for maxBCG catalog are obtained by applying the cluster-finding algorithm on mock catalogs. Each identified cluster is later verified if it actually corresponds to a real DM halo whose physical extension is determined by its R_{200} – the radius within which the matter density contrast is 200 times above the critical density at that redshift. Each cluster that does not belong to any halo counts as a false-positive [42].

[7]In [121], the mean X-ray luminosities in nine optical richness bins were first measured by stacking X-ray emission from clusters in ROSAT All-Sky Survey data, the $L_X - N_{200}$ relation was then obtained assuming a power-law fit to the data after corrected for systematic bias.

lensing mass estimates of [119, 120]. We refer readers to the discussion leading to this equation for an in-depth context of systematics in measurements of the mean $M - N_{200}$ relation.

Chapter 4

Bayesian forward modeling and inference approach to galaxy clustering

This chapter provides an overview of the fundamental building blocks of the Bayesian forward approach in cosmological modeling of, and inference from, galaxy clustering (Section 4.1), with the explicit example of the BORG framework (Section 4.2). Figure 4.1 provides a simplified schematic flowchart of this approach. We have encountered the blocks of prior on initial conditions, (gravitational) forward model, (deterministic) bias expansion and observational effects in the previous two chapters. Below we will do a quick recap on some of these ingredients before focusing on the likelihood for stochastic bias, and the joint posterior on initial conditions plus cosmological parameters.

Note that some detailed steps shown in Figure 4.1 might differ for specific types of input data. For example, with the data based on N-body simulation, used in Chapter 5 and Chapter 7, there is no observational effects to be accounted for, hence $\vec{\delta}_{h,\mathrm{pred}} \equiv \vec{\delta}_{h,\mathrm{det}}$. Furthermore, depending on the application, the full Monte-Carlo Markov chain (MCMC) sampling step might be reduced to a profile likelihood estimate, as will be seen in Chapter 7.

Throughout the rest of this chapter, for simplicity, we consider a single tracer field $\vec{\delta}_h$ on a fixed time slice τ, which we leave implicit. These assumptions can be generalized straightforwardly as discussed in Section 3.1 and [65]. The main body of this chapter was published in [123].

4.1 Bayesian forward modeling of galaxy clustering

Our final goal is to derive a joint posterior for the initial density field $\vec{\delta}_{m,\mathrm{ini}}$, cosmological parameters $\{\theta\}$, plus "nuisance parameters", i.e. bias parameters and stochastic amplitudes, which describe the uncertainties in the formation process of our tracers. This posterior involves four key ingredients:

1. The prior on the initial conditions (see Section 2.1).

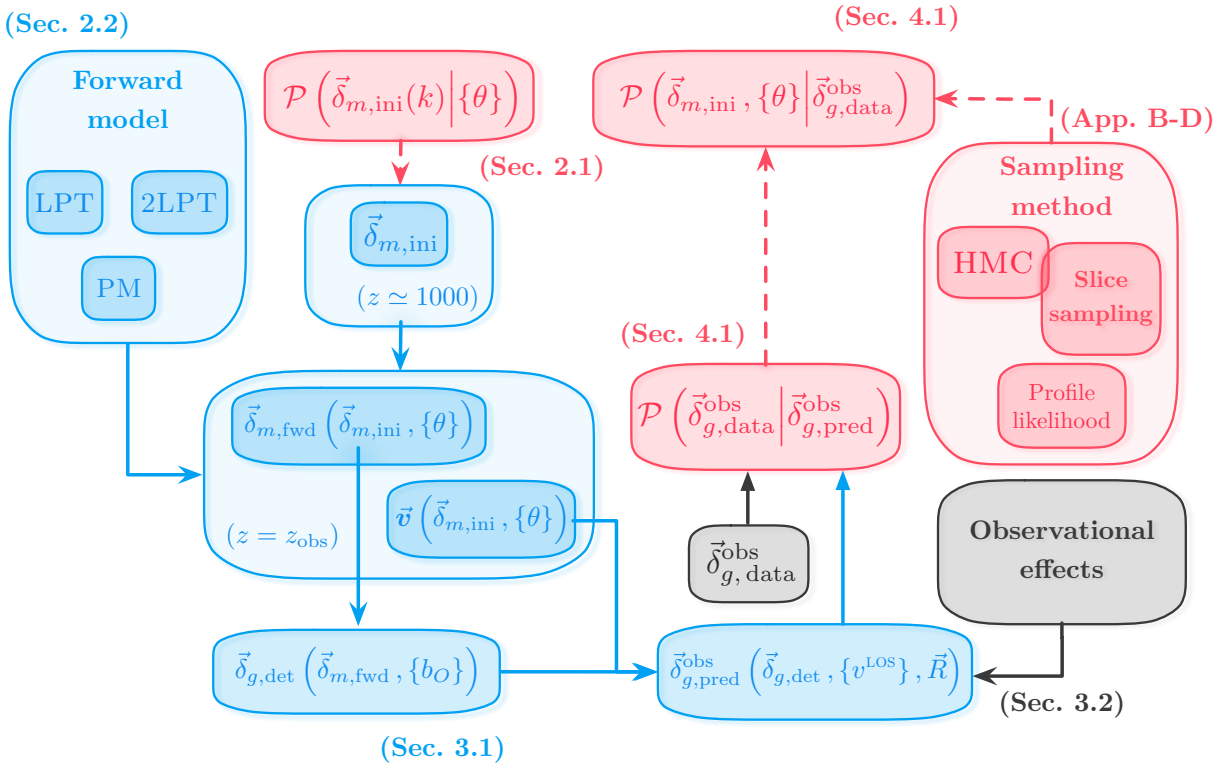

Figure 4.1: A schematic flowchart of the Bayesian forward modeling approach for galaxy clustering. $\{\theta\}$ denote the set of relevant cosmological parameters; $\vec{\delta}_{m,\mathrm{ini}}$ denotes the initial matter density fluctuations or initial Fourier modes of matter fluctuations at $z \simeq 1000$; $\vec{\delta}_{m,\mathrm{fwd}}$ denotes the gravitationally-evolved/forwarded matter density field at $z = z_{\mathrm{obs}}$, while $\vec{\boldsymbol{v}}$ denotes the associated peculiar velocity field; $\vec{\delta}_{g,\mathrm{det}}$ denotes the mean-field/deterministic galaxy field, while $\{b_O\}$ denote the set of bias parameters included in the modeling; $\vec{\delta}^{\mathrm{obs}}_{g,\mathrm{pred}}$ denotes the observationally predicted galaxy field after accounting for selection and systematic effects (Section 3.2.1), with $\vec{R}$ being the survey response function encapsulating completeness, angular and radial selection functions; $\vec{\delta}^{\mathrm{obs}}_{g,\mathrm{data}}$ denotes the actually observed galaxy field. The red boxes display the probability distributions including $P\left(\vec{\delta}_{m,\mathrm{ini}}(k)\middle|\{\theta\}\right)$ – the *Gaussian* prior on the initial power spectrum (Section 2.1), $P\left(\vec{\delta}^{\mathrm{obs}}_{g,\mathrm{data}}\middle|\vec{\delta}^{\mathrm{obs}}_{g,\mathrm{det}}\right)$ – the likelihood for modeling of stochasticity (see Section 4.1.4), $P\left(\vec{d}_{m,\mathrm{ini}}, \{\theta\}\middle|\vec{\delta}^{\mathrm{obs}}_{g,\mathrm{data}}\right)$ – the joint posterior of initial conditions and cosmological parameters (Section 4.1.5), and the method used to sample/evaluate the final posterior (Appendix B-Appendix E). The blue boxes represent the gravitational forward (Section 2.2) and deterministic bias models (Section 3.1.1). The gray boxes show the input data and observational effects. The solid arrows correspond to deterministic transitions, while the dashed arrows correspond to probabilistic transitions.

2. The gravitational forward model for matter fluctuations (see Section 2.2).

3. The deterministic bias model (see Section 3.1.1).

4. The conditional likelihood for the halo density at a given point (see Section 4.1.4).

Following our discussions in previous two chapters, we have seen that the first [51], second [45] and third constituents [65] are already on solid footings. Hence the last piece is the core open issue in this enterprise.

4.1.1 Recap: Gaussian prior on initial conditions

We assume adiabatic, growing-mode initial conditions (see Section 2.2), so that the initial conditions are given by a single field $\vec{\delta}_{m,\mathrm{ini}}$, the initial density field. As mentioned in Section 2.1, we additionally assume that our prior on $\vec{\delta}_{m,\mathrm{ini}}$ follows a *multivariate Gaussian* distribution, specified by the linear matter power spectrum $P_{\mathrm{L}}(k, \{\theta\})$, which in turn depends on a set of cosmological parameters $\{\theta\}$ as

$$\mathcal{P}_{\mathrm{prior}}\left(\vec{\delta}_{m,\mathrm{ini}}|\{\theta\}\right) = \mathcal{N}\left(\vec{\delta}_{m,\mathrm{ini}}\middle|\ \vec{\mu}=\vec{0},\ \mathrm{C}=\mathrm{FT}^{\dagger}[\mathrm{diag}\{P_{\mathrm{L}}(k_i,\{\theta\})\}]\mathrm{FT}\right)\,, \tag{4.1}$$

wherein $\mathcal{N}$ denotes a multivariate Gaussian distribution, while $\vec{\mu}$ is the vanishing expectation value, and C is the covariance matrix which is *diagonal* in Fourier space. This can be generalized to include primordial non-Gaussianity, and isocurvature perturbations between baryons and CDM, as pointed out in the conclusion of [123].

4.1.2 Recap: Gravitational forward model

The gravitational forward model yields the probability of finding an evolved density field $\vec{\delta}_{m,\mathrm{fwd}}$, given an initial density field $\vec{\delta}_{m,\mathrm{ini}}$ and cosmological parameters $\{\theta\}$. We can write this as a deterministic transition (see also [39, 102, 43])

$$\mathcal{P}\left(\vec{\delta}_{m,\mathrm{fwd}}\left|\vec{\delta}_{m,\mathrm{ini}},\{\theta\}\right.\right) = \prod_{i=1}^{N_g^3}\delta_D\left(\delta^i - \delta^i_{m,\mathrm{fwd}}[\vec{\delta}_{m,\mathrm{ini}},\{\theta\}]\right), \tag{4.2}$$

thanks to the deterministic nature of gravitational-only evolution. The actual physics of the forward model, e.g. first-/second-order Lagrangian perturbation theory (LPT/2LPT), etc. (see Section 2.2.2), is encoded in the non-linear non-local functional $\delta^i_{m,\mathrm{fwd}}[\vec{\delta}_{m,\mathrm{ini}},\{\theta\}]$. Any such forward model $\delta^i_{m,\mathrm{fwd}}[\vec{\delta}_{m,\mathrm{ini}},\{\theta\}]$ will certainly be imperfect (see Figure 2.2 and Figure 2.3 for the LPT/2LPT case) – both due to approximations made during the calculation and the fact that only modes down to some finite minimum scale are included, a fact which is neglected in Eq. (4.2). We will return to this below in the context of the likelihood.

Nevertheless, as long as certain conditions regarding mass and momentum conservation of matter are met, the small-scale uncertainties in the forward model for matter and gravity

can be effectively included in the conditional likelihood. In the effective field theory (EFT) context, the main requirement is that the forward model consistently includes all relevant terms up to a fixed order in PT, and that numerical truncation errors can be neglected. To be specific, in Chapter 7, we will assume that the error in the forward model of matter is at least third order in PT. While our results shown in Chapter 5-Chapter 7 are based on 2LPT or LPT, nothing apart from computational expense prevents our approach from being coupled to a full N-body simulation as forward model.

4.1.3 Recap: Deterministic bias model

Recall from Section 3.1, specifically Eq. (3.1), that our *deterministic* halo density field at any given point can be written as a linear superposition of operators, or fields, O,

$$\vec{\delta}_{h,\mathrm{det}}[\vec{\delta}_{m,\mathrm{fwd}}, \{b_O\}] = \sum_O b_O \vec{O}\left[\vec{\delta}_{m,\mathrm{fwd}}\right] , \tag{4.3}$$

in which $\vec{O}$ includes the complete linearly independent set of local gravitational observables at a certain order in PT: matter density and velocity divergence, tidal field, and so on (cf. Eq. (3.32)). At a given order in PT, there is only a finite, fixed number of linearly independent gravitational observables [124, 125]. Moreover, even though the gravitational observables include time derivatives, all of these operators can be expressed as non-local, non-linear transformations of the final density field $\vec{\delta}_{m,\mathrm{fwd}}$ [125] (see Section 2.5 of [65] for a review). b_O are the corresponding bias parameters. Note that the operators in Eq. (4.3), as denoted by $\vec{}$, are constructed from the evolved density field on the grid, i.e. *filtered on the grid scale*. This will become relevant in Chapter 7.

At this point, it is worth to emphasize that Eq. (4.3) only predicts the halo density field in a statistical sense. That is, if we imagine stacking many cells that have the same values of all operators O appearing in Eq. (4.3), then the mean density of halos in these cells should approach the prediction in Eq. (4.3). At any given position, the halo density can deviate from the prediction in Eq. (4.3), due to the random nature of the small-scale perturbations that we have integrated out in the bias expansion, yet are still relevant for halo formation. The conditional probability discussed below is supposed to account for this "scatter".

4.1.4 Stochasticity and conditional probability

The final ingredient needed in the Bayesian forward model is the probability for finding a certain number of halos (or galaxies) in a given cell, given the predicted deterministic field $\vec{\delta}_{h,\mathrm{det}}$ (as well as the matter density field). We will phrase this equivalently as the *conditional probability for finding a measured halo density field $\vec{\delta}_h$ given the predicted mean-field halo density $\vec{\delta}_{h,\mathrm{det}}$*. As discussed in the previous section, this probability should take into account the scatter induced by the small-scale modes that are not explicitly included in the forward model, which are nevertheless relevant for determining exactly where a halo

forms. Further, the conditional probability also needs to be able to capture deficiencies in the bias expansion Eq. (4.3), as well as in the forward model for matter and gravity.

One approach, followed by many if not most of the literature on this topic so far, is to assume that the size of the grid cells $R_{\rm cell}$ is much larger than the scale R_* that controls the higher-derivative contributions to the halo density in Eq. (4.3). That is, one assumes that on the scales resolved on the grid, halo formation can be effectively approximated as *spatially local*. In the EFT approach, the leading correction to this assumption is captured by the operator $\nabla^2\delta$ in Eq. (7.1), whose coefficient then should be (at least) of order R_*^2, i.e. $|c_{\nabla^2\delta}| \sim R_*^2$. Alternatively, one can explicitly include the difference between neighboring cells, as done in [126]. This approach thus assumes that the impact of all modes resolved on the grid, including the correlations of the halo density between different cells, is completely captured by a finite set of operators, in our case those appearing in Eq. (4.3).

Thus, building on the assumption on the locality of halo formation, the likelihood of a given halo density field *given* a matter density field and bias parameters is a product of conditional probabilities in each cell:

$$\mathcal{P}\left(\vec{\delta}_h \middle| \vec{\delta}_{m,\rm fwd}, \{b_O\}, \{\lambda_a\}\right) = \prod_{i=1}^{N_g^3} P^{(1)}\left(\delta_h^i - \delta_{h,\rm det}^i[\vec{\delta}_{m,\rm fwd}, \{b_O\}],\ \{\lambda_a\},\ \delta_{m,\rm fwd}^i\right), \tag{4.4}$$

where $P^{(1)}$ is the probability for finding, in a given cell, an overdensity δ_h^i given the prediction for the mean relation $\delta_{h,\rm det}^i$ from Eq. (4.3). Here we have allowed $P^{(1)}$ to depend on further parameters $\{\lambda_a\}$ (e.g., variance, skewness, etc.), as well as the matter density itself to take into account, for example, a larger variance in high-density regions.

Clearly, Eq. (4.4) still contains significant freedom to choose the form of the conditional probability $P^{(1)}$. Moreover, unlike the case for the bias expansion in Eq. (4.3), there is no guide from EFT considerations on what the form of $P^{(1)}$ should be, since $P^{(1)}$ arises from integrating out small-scale, fully non-linear modes whose PDF is not expected to be close to Gaussian. There is a limit for which the PDF is expected to asymptote to a known form: if the size of grid cells is much less than the mean separation between halos, then we expect the single-cell PDF to approach a Poisson distribution – by virtue of the law of rare events – given by

$$\mathcal{P}^{(1)}\left(n_h^i - n_{h,\rm det}^i[\vec{\delta}_{m,\rm fwd}, \{b_O\}]\right) = \frac{\left(n_{h,\rm det}^i\right)^{n_h^i}\ e^{-n_{h,\rm det}^i}}{(n_h^i)!}, \tag{4.5}$$

where we have replaced δ_h and $\delta_{h,\rm det}$ with n_h and $n_{h,\rm det}$ such that, for example, $n_h^i = \overline{n}_h\,(1+\delta_h^i)$, to respect the discrete nature of the Poisson distribution. Note that we have removed the variables $\{\lambda_a\},\ \delta_{m,\rm fwd}$ as a Poisson distribution is fully specified by its mean which, in this case, is simply $n_{h,\rm det}$.

Unfortunately however, this limit is not attainable within an EFT context, since the perturbative bias expansion Eq. (4.3) breaks down for such a small grid scale (in practice, the grid scale will then also be much smaller than R_*). Thus, in order to obtain a posterior that is under rigorous perturbative control, we must pursue a different route than Eq. (4.4).

4.1.5 The joint posterior of initial conditions and late-time large-scale structure

Let us now put together the ingredients presented above in order to obtain the final joint posterior for the initial density field, cosmological parameters, as well as nuisance parameters:

$$\mathcal{P}\left(\vec{\delta}_{m,\mathrm{ini}}, \{\theta\}, \{b_O\}, \{\lambda_a\} \middle| \vec{\delta}_h\right) = \mathcal{N}_P P_{\mathrm{prior}}(\vec{\delta}_{m,\mathrm{ini}}|\{\theta\}) P\left(\vec{\delta}_h \middle| [\vec{\delta}_{m,\mathrm{ini}}, \{\theta\}],\ \{b_O\},\ \{\lambda_a\}\right) , \tag{4.6}$$

where $\mathcal{N}_P$ is a normalization constant. Then, the desired cosmological constraints can be obtained by marginalizing over the initial phases $\vec{\delta}_{m,\mathrm{ini}}$ and nuisance parameters as

$$\mathcal{P}\left(\{\theta\} \middle| \vec{\delta}_h\right) = \int d\{b_O\} \int d\{\lambda_a\}\, \mathcal{P}_{\mathrm{prior}}(\{b_O\}, \{\lambda_a\}) \int \mathcal{D}\vec{\delta}_{m,\mathrm{ini}}\, \mathcal{P}\left(\vec{\delta}_{m,\mathrm{ini}}, \{\theta\}, \{b_O\}, \{\lambda_a\} \middle| \vec{\delta}_h\right) , \tag{4.7}$$

in which P_{prior} is a prior on the bias and PDF parameters, while $\int \mathcal{D}\vec{\delta}_{m,\mathrm{ini}}$ denotes the functional integral over the whole parameter space of initial conditions. Similarly, one can also obtain marginalized posteriors for the bias parameters b_O and likelihood parameters λ_a.

4.2 The BORG algorithm in a nutshell

The BORG algorithm [39, 127] provides a well-constructed, modular, numerical framework for the Bayesian forward modeling and inference approach to galaxy clustering. In essence, it allows for numerically efficient sampling of the joint posterior described in Section 4.1.5, given the huge parameter space of $\left\{\vec{\delta}_{m,\mathrm{ini}}, \{\theta\}, \{b_O\}, \{\lambda_a\}\right\}$[1], by employing a combination of Hamiltonian Monte-Carlo (HMC) [128] (for $\vec{\delta}_{m,\mathrm{ini}}$) and slice sampling [129] (for $\{\theta\}, \{b_O\}, \{\lambda_a\}$) methods for a series of sequential sampling blocks[2] [127, 102, 43]. In particular, the transition from the s-th sample to the $(s+1)$-th sample can be broken down

[1] Typically, to achieve a reasonably-high resolution of the reconstructed density and peculiar velocity fields, one needs a grid of size $N^3_{\mathrm{grid}} = (128)^3 - (256)^3$.

[2] This idea of factorization the high-dimensional joint posterior into multiple lower-dimensional conditional probabilities can be traced back to [130], and is similar to the idea of *Gibbs sampling* method [131, 132, 133],.

to the following steps:

$$(1)\ \vec{\delta}_{m,\mathrm{ini}}^{(s+1)} \curvearrowleft \mathcal{P}\left(\vec{\delta}_{m,\mathrm{ini}}\middle|\{\theta\}^s, \vec{\delta}_h, \overline{n}_h^s, \{b_O\}^s, \{\lambda_a\}^s\right) \tag{4.8}$$

$$(2)\ \overline{n}_h^{(s+1)} \curvearrowleft \mathcal{P}\left(\overline{n}_h\middle|\vec{\delta}_{m,\mathrm{ini}}^s, \{\theta\}^s, \vec{\delta}_h, \{b_O\}^s, \{\lambda_a\}^s\right) \tag{4.9}$$

$$(3)\ \{b_O\}^{(s+1)} \curvearrowleft \mathcal{P}\left(\{b_O\}\middle|\vec{\delta}_{m,\mathrm{ini}}^s, \{\theta\}^s, \vec{\delta}_h, \overline{n}_h^s, \{\lambda_a\}^s\right) \tag{4.10}$$

$$(4)\ \{\lambda_a\}^{(s+1)} \curvearrowleft \mathcal{P}\left(\{\lambda_a\}\middle|\vec{\delta}_{m,\mathrm{ini}}^s, \{\theta\}^s, \vec{\delta}_h, \overline{n}_h^s, \{b_O\}^s\right) \tag{4.11}$$

$$(5)\ \{\theta\}^{(s+1)} \curvearrowleft \mathcal{P}\left(\{\theta\}\middle|\vec{\delta}_{m,\mathrm{ini}}^s, \vec{\delta}_h, \overline{n}_h^s, \{b_O\}^s, \{\lambda_a\}^s\right). \tag{4.12}$$

Here, the $\curvearrowleft$ symbol denotes a sample on the l.h.s. being drawn from the conditional probability distribution on the r.h.s., while $\overline{n}_h$ denotes the comoving mean number density of halos. Step (1) utilizes the HMC sampling method (see Appendix B) while steps $(2)-(4)$ employ the slice sampling method (see Appendix C), to ensure a close-to-unity acceptance rate [102, 127]. Let us emphasize, at this point, that BORG seeks to construct a *fair ensemble* of points in the posterior parameter space, which presumably serves as a numerical approximation of the underlying posterior. It is not an iterative procedure, which would only provide a *single* realization of the reconstruction and hence would potentially bias the result as for the case of the Wiener filter [134]. More technical details about the BORG machinery will be discussed in the next chapter where we will perform an in-depth investigation on the performance of BORG.

Chapter 5

Inference of initial conditions

This chapter describes our numerical investigation of the quality of the three-dimensional initial conditions $\vec{\delta}_{m,\mathrm{ini}}$ inferred from Bayesian forward modeling of the biased tracer field $\vec{\delta}_h$, as implemented in the BORG framework. We restrict our tracers to DM halos throughout the rest of this chapter. Specifically, as discussed in the last chapter, two key ingredients for a quality inference are the deterministic bias expansion and the likelihood for stochastic bias. Below, we will directly compare the quality of reconstructions using different combination of these two ingredients. In addition, we also vary different factors in our setup, including tracer number density, grid resolution and gravitational forward model, in order to evaluate their effects on the reconstruction and determine the robustness of certain (deterministic) bias models and likelihoods.

The work presented here is in the advanced stage of preparation for publication. Major changes in the main results and conclusions are not expected in the final paper, although some minor technical details might differ. We will focus on presenting the main estimators, our findings and highligh the key challenges for any choice of bias model and likelihood. We then organize the rest of this chapter as follows. In Section 5.1, we characterize the setup for each BORG reconstruction, focusing on the elements under investigation in this chapter (see Section 4.1 for an overview of the BORG framework and main ingredients). We further detail the DM-only simulation and the DM halos we use here as the main input data in Section 5.2, before turning to the implementation in Section 5.3. In Section 5.4, we first define the estimators used to assess the quality of the reconstruction and then present our results in Section 5.5. The convergence and correlation tests we have performed for the MCMC chains included in this analysis are documented in Appendix D.

5.1 Setup

In the last chapter, we have reviewed all the four main ingredients of a BORG reconstruction. Below, we will focus on the relevant aspects for our comparison.

5.1.1 Choices of forward model

For this comparison, we choose the first- and second-order Lagrangian perturbation theory (LPT and 2LPT, see Section 2.2.2 for details) as our gravitational forward models (see Section 4.1.2). These choices should be particularly relevant for current and future analyses of large-volume surveys [43] as these approximations facilitate the numerical computation – as opposed to a full N-body approach – yet still capture the sub-leading effects of large-scale modes on the Baryon Acoustic oscillation (BAO) feature (see Appendix N).

5.1.2 Choices of bias models

In this work, we compare three non-linear deterministic bias models. While this is neither an exhaustive list of bias models included in BORG nor studied in the literature so far, it does include some of the most common choices in previous and ongoing analyses [135, 102, 43, 127, 123, 136]. Below, we review these models as implemented in the BORG framework where the halo and matter density fields are projected onto grid cells for numerical computation.

Power-law bias

In the power-law bias model, the expected halo density in a given grid cell i is related to the evolved matter density at that cell via only one free parameter, the power-law index β [135]

$$\delta^i_{h,\mathrm{det}} = \left(1 + \delta^i_{m,\mathrm{fwd}}\right)^\beta - 1. \tag{5.1}$$

This local bias relation belongs to the LIMD class of bias models – previously described in Section 3.1.1 – as the r.h.s. of Eq. (5.1) includes only one single operator constructed out of the local matter density field, namely $\delta^i_{m,\mathrm{fwd}}$. For small perturbations, $\delta^i_{m,\mathrm{fwd}} \ll 1$, Eq. (5.1) simplifies to

$$\delta^i_{h,\mathrm{det}} = \beta\delta^i_{m,\mathrm{fwd}} + \frac{\beta(\beta-1)}{2}\left(\delta^i_{m,\mathrm{fwd}}\right)^2 + \mathcal{O}\left(\delta^3_{m,\mathrm{fwd}}\right). \tag{5.2}$$

Thus, for cells where the evolved matter density fluctuations are still within the linear regime, the role of the power-law index reduces to that of the large-scale linear bias parameter, i.e. $\beta \simeq b_1$. As we will see below, when non-linear information is included, β deviates from b_1 significantly.

Broken power-law bias

The broken power-law bias model, as its name indicates, introduces an additional power-law cut-off into the power-law bias relation in Eq. (5.1), such that [102, 43, 127]

$$\delta^i_{h,\mathrm{det}} = \left(1 + \delta^i_{m,\mathrm{fwd}}\right)^\beta \exp\left[-\rho(1 + \delta^i_{m,\mathrm{fwd}})^{-\epsilon}\right] - 1. \tag{5.3}$$

This model essentially behaves like the power-law bias model but accounts for the exponential suppression of halo clustering in voids [137]. It is characterized by three bias parameters (β, ρ, ϵ), and can also be classified as LIMD bias.

Note that by definitions in Eq. (5.1) and Eq. (5.3), there is no guarantee that $\langle \delta_{h,\mathrm{det}} \rangle = 0$. Technically, $\langle \delta_{h,\mathrm{det}} \rangle \neq 0$ is absorbed by the mean halo density $\overline{n}_h = n_h^i \left(1 + \delta_h^i\right)^{-1}$ which is also being sampled in BORG.

Second-order bias

As discussed in Section 3.1, one should also account for the effect of tidal fields K_{ij} and the fact that halo and galaxy formation is a non-local process (see Section 3.1 and [65]). Following our work in [123, 136], we implement in BORG a non-local, second-order bias model given by

$$\delta_{h,\mathrm{det}}^i = b_1 \delta_{m,\mathrm{fwd}}^i + \frac{b_2}{2} \left[\left(\delta_{m\,\mathrm{fwd}}^i \right)^2 \right] + b_K \left[\left(K_{ij} \right)^2 \right] + c_{\nabla^2 \delta} \nabla^2 \delta_m, \tag{5.4}$$

where the extra square brackets $[O]$ denote *mean-subtracted* operators. Subtracting the mean-field values for these operators ensure that $\langle O \rangle = 0$,[1] including also $\langle \delta_{h,\mathrm{det}} \rangle$. The motivation behind this technical complication is to ensure that our inference of the bias parameters is not strongly dependent on the smoothing scale imposed by the BORG grid size. Note that the higher-derivative bias coefficient $c_{\nabla^2 \delta}$ is an effective coefficient which also absorbs higher-order contributions depending on the chosen smoothing scale.

5.1.3 Poisson likelihood

For simplicity and robustness, in this comparison, we only focus on one particular form of the conditional probability for the stochasticity in the actual halo distribution, namely the Poisson likelihood which is the basis for the main active likelihoods in BORG [135, 102, 43, 127]. It should be noted that the stochasticity in the one-point PDF (count-in-cell) of DM halos in N-body simulations follows Poisson distribution only for highly under-dense cells whose $1 + \delta_m \ll 1$, while it tends to *sub*-Poisson at $1 + \delta_m \simeq 1$ due to the halo exclusion effect, and finally becomes *super*-Poisson in highly over-dense cells where $1 + \delta_m \gg 1$ [139] (see also [137, 65] and references therein).

The single-cell Poisson likelihood implemented in BORG is identical to Eq. (4.5). For clarification, let us re-write it here:

$$\mathcal{P}^{(1)} \left(n_h^i \middle| n_{h,\mathrm{det}}^i \right) = \frac{\left(n_{h,\mathrm{det}}^i \right)^{n_h^i} \, e^{-n_{h,\mathrm{det}}^i}}{(n_h^i)!}. \tag{5.5}$$

The joint likelihood is then simply a product of Eq. (5.5) evaluated at all cells, as given by Eq. (4.4).

[1] This is one of the renormalization conditions prescribed in Appendix J and [123, 138]. We recommend readers to see also Section 2.10.4 of [65] for a detailed discussion on the motivation behind renormalization.

5.1.4 Choices of grid resolution

We run our reconstructions at three different resolutions correspond to the grid spacings of $L_{\rm grid} = 31.25, 20.83, 15.625\ h^{-1}$Mpc. For our box size of $L_{\rm box} = 2000\ h^{-1}$Mpc, these resolutions correspond to a grid of $N_{\rm grid} = 64^3, 96^3, 128^3$, respectively.

5.2 Input data

The main input data for our BORG reconstruction is one of the GADGET-2 [83] simulations presented in [140]. The simulation adopted a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $n_{\rm s} = 0.967$, $h = 0.7$, $\sigma_8 = 0.85$, for a box size of $L = 2000\ h^{-1}$Mpc, and a total number of $N_{\rm part} = 1536^3$ DM particles of mass $M_{\rm part} = 1.8 \times 10^{11}\, h^{-1} M_\odot$. The transfer function for the given cosmology was computed with the Boltzmann code CLASS[2] [141]. Initial conditions for the N-body run were then generated at redshift $z_{\rm ini} = 99$ using second-order Lagrangian perturbation theory (2LPT) [71] with the 2LPTic algorithm [79, 142]. DM halos are subsequently identified at redshift $z = 0$ as spherical over-densities (SO) [143, 144, 145] using the Amiga Halo Finder (AHF) algorithm[3] [146, 147], where we choose an over-density threshold of 200 times the background matter density. Note that we only include the main halos for our analysis. We further divide the halos into three mass bins of $\log_{10} M_h = [13.0, 13.5), [13.5 - 14.0), [14.0 - 14.5)\, h^{-1} M_\odot$[4]. Their corresponding number densities are $\overline{n}_h = 2.91 \times 10^{-5}, 1.10 \times 10^{-4}, 3.34 \times 10^{-4}\, (\,h\,{\rm Mpc}^{-1})^3$.

Additionally, we also generate mock halo fields from the same combinations of bias model and likelihood used in some of the reconstructions. We study reconstructions using this particular type of input data in Section 5.4.1 to better understand the optimal limit where the underlying mechanism that generates the data coincide with BORG data model, in particular, the combination of deterministic bias model and likelihood.

5.3 Implementation

We impose uniform priors for all bias parameters, with positivity constraints on $\overline{n}_h$ and β (for power-law and broken-powerlaw bias) or β_1 (for second-order bias). The bias parameters are sampled using the slice sampling method as mentioned in Section 4.2 and further described in Appendix C. We refer readers to Appendix D for details of how we ensure that all the MCMC chains have converged and how we only include independent samples from each chain to facilitate numerical computations.

To ensure the positivity of $n_{h,\rm det}$ (cf. Eq. (5.5)) in case of the second-order bias model, where $\delta_{h,\rm det}$ is not strictly non-negative (cf. Eq. (5.4)), we, at the boundary $\delta_{h,\rm det} = 0$,

[2] `https://http://class-code.net/`

[3] `http://popia.ft.uam.es/AHF/`

[4] For each $[M_{\rm lower}, M_{\rm upper})$ mass bin, we include halos whose $M_{\rm lower} \leq M_{\rm viral} < M_{\rm upper}$.

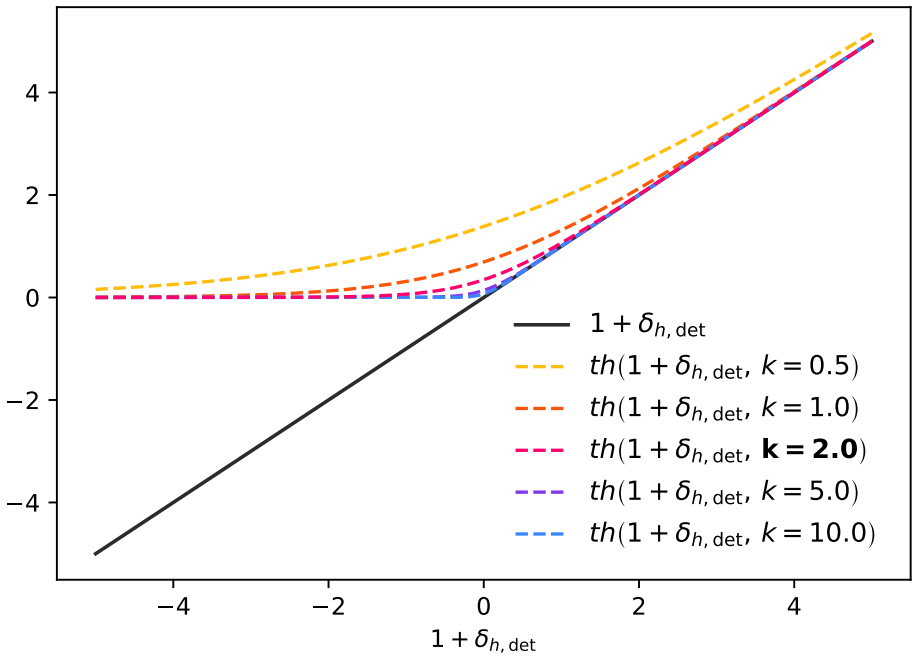

Figure 5.1: Comparison of the density thresholder (cf. Eq. (5.7)) for $1+\delta_{h,\mathrm{det}}$, with various choices of k (dashed color lines), and the exact value of $1+\delta_{h,\mathrm{det}}$ (continuous gray line).

impose a thresholder:

$$n_{h,\mathrm{det}} = \bar{n}_{h,\mathrm{grid}} \begin{cases} (1+\delta_{h,\mathrm{det}}) & \delta_{h,\mathrm{det}} > 0 \\ th\,(1+\delta_{h,\mathrm{det}}, k) & \text{otherwise,} \end{cases} \tag{5.6}$$

where

$$th\,(1+\delta_{h,\mathrm{det}}, k) = \ln\left(1+\exp\left[k\,(1+\delta_{h,\mathrm{det}})\right]\right)/k. \tag{5.7}$$

Here, $\bar{n}_{h,\mathrm{grid}} = n_h/N^3_{\mathrm{grid}}$ denotes the mean halo number in each grid-cell. We choose the value $k = 2.0$ to balance between numerical precision and performance, in particular the acceptance rate (see Appendix B), according to results from our numerical tests. Note that a higher value of k means a sharper transition around the boundary $\delta_{h,\mathrm{det}} = 0$ (see Figure 5.1).

5.4 Results

We define the Fourier-space *correlation coefficient* between two given fields δ_1, δ_2 as

$$r_{12}(k) = \frac{\langle \delta_1 \delta_2 \rangle_k}{\sqrt{\langle \delta_1 \delta_1 \rangle_k \ \langle \delta_2 \delta_2 \rangle_k}} \tag{5.8}$$

where $\langle \, \rangle_k$ denotes the average over wavenumber k.

Let us write down the correlation coefficient Eq. (5.8) between the reconstructed initial modes in BORG s'-th sample and those in the input as

$$r_{\rm ri}^{s'}(k) = \frac{\left\langle \delta_{\rm rec}^{s'} \delta_{\rm input} \right\rangle_k}{\sqrt{\langle \langle \delta_{\rm rec} \rangle_s \rangle_k \ \langle \delta_{\rm input} \rangle_k}} \tag{5.9}$$

where $\langle \, \rangle_s$ denotes the BORG ensemble average, and we have shortened the notation by $\delta_{m,\rm ini}^{\rm rec} \equiv \delta_{\rm rec}$ and $\delta_{m,\rm ini}^{\rm input} \equiv \delta_{\rm input}$, for readability. Note also that we use $\langle \langle \delta_{\rm rec} \rangle_s \rangle_k$ instead of $\left\langle \delta_{\rm rec}^{s'} \right\rangle_k$ in the denominator as an approximation, assuming that the nominator varies much faster than the denominator across BORG samples. This approximation effectively reduces the shot noise in the auto-correlation of the BORG reconstructed initial modes. We can then define our estimator of the correlation coefficient for each reconstruction (recall that each reconstruction consists of an ensemble of BORG samples) as

$$r_{\rm ri}(k) = \left\langle r_{\rm ri}^{s'}(k) \right\rangle_s \tag{5.10}$$

whose variance is given by the BORG sample variance:

$$\sigma_r^2(k) = \left\langle \left[r_{\rm ri}^{s'}(k) - \left\langle r_{\rm ri}^{s'}(k) \right\rangle_s \right]^2 \right\rangle_s . \tag{5.11}$$

Technically, the correlation coefficient $r_{\rm ri}$ can take any value between 1 and -1. These two limits correspond to, respectively, the reconstructed field and the input are perfectly correlated or anti-correlated, while a value of 0 would indicate that the two are completely uncorrelated. From a theoretical standpoint, one would expect $r_{\rm ri}$ get closer to 1 at *small* k, where the gravitational evolution of matter density fluctuations is linear (see Section 2.2), and to *decrease* at progressively *larger* k, where the evolution becomes more and more non-linear. We will return to this point below.

As a reference, let us derive the *large-scale* limit of $r_{\rm ri}$, i.e. where δ_m, δ_h, and ε_h (stochasticity) are all *Gaussian* fields, and additionally, we can assume a *linear* bias relation of the following form:

$$\delta_h(\boldsymbol{k}) = b_1 \delta_{m,\rm lin}^{\rm input}(\boldsymbol{k}) + \varepsilon_h(\boldsymbol{k}), \tag{5.12}$$

where $\delta_{m,\rm lin}^{\rm input}$ here denotes the matter density field *linearly extrapolated* to redshift zero (see Section 2.2). Since the halo stochasticity and the matter density field are assume to be *uncorrelated* in this limit, we can write

$$\left\langle \delta_{m,\rm lin}^{\rm rec}(\boldsymbol{k}) \delta_{m,\rm lin}^{\rm input}(\boldsymbol{k}) \right\rangle_k = P_{\rm L}(k), \tag{5.13}$$

where $P_{\rm L}$ denotes the linear matter power spectrum at z=0. In addition, the power spectrum of the reconstructed modes can be expressed as (cf. Eq. (5.12))

$$\begin{aligned}\left\langle \delta^{\rm rec}_{m,{\rm lin}}(\boldsymbol{k})\delta^{\rm rec}_{m,{\rm lin}}(\boldsymbol{k})\right\rangle_k &= \frac{1}{b_1^2}\left\langle \delta_h(\boldsymbol{k})\delta_h(\boldsymbol{k})\right\rangle_k \\ &= \frac{1}{b_1^2}\left(b_1^2 P_{\rm L}(k) + P_\varepsilon(k)\right) \quad = P_{\rm L}(k) + P_\varepsilon(k)/b_1^2. \end{aligned} \tag{5.14}$$

Putting Eqs. (5.13)–(5.14) into the definition in Eq. (5.9), assuming that each BORG sample s' gives us an *unbiased* estimate of the true underlying density field, we retrieve

$$r_{ri,{\rm lin}}(k) = \frac{1}{\sqrt{1 + P_\varepsilon(k)/\left(b_1^2 P_{\rm L}(k)\right)}}. \tag{5.15}$$

Eq. (5.15) clearly shows that there is a expected limit for the reconstruction, if only information from linear evolution of LSS is used, set by the halo stochasticity and bias.

For comparison purpose, since the halo stochasticity power spectrum is not precisely known, we approximate Eq. (5.15) by the correlation coefficient between the actual halo and the input initial Fourier modes, $r_{\rm hi}$, which from Eq. (5.8) reads

$$r_{\rm hi}(k) = \frac{\left\langle \delta_h \delta_{\rm input}\right\rangle_k}{\sqrt{\left\langle \delta_h\delta_h\right\rangle_k \left\langle \delta_{\rm input}\delta_{\rm input}\right\rangle_k}}. \tag{5.16}$$

Below, we show Eq. (5.16) in all our plots of correlation coefficient to serve as a reference guideline. Note that Eq. (5.16) can deviate from Eq. (5.15) on small scales, i.e. high k values, where the fields are non-Gaussian. However, at large scales, i.e. small k values, for whatever weights are used, Eq. (5.15) always reduces to Eq. (5.16).

The difference between $r_{\rm ri}$ and $r_{\rm hi}$ is then a quantitative measure of how good a reconstruction performs. We then define k_{80}, k_{50} and k_{20} as the wavenumber at which $r_{\rm ri}(k) = 0.8,\ 0.5,\ 0.2$ respectively.

In addition, we also measure the *amplitude bias* between each reconstruction sample s' and the input as

$$T^{s'}_{\rm rec}(k) = \sqrt{\frac{P^{s'}_{\rm rec}(k)}{P_{\rm true}(k)}}, \tag{5.17}$$

where $P(k) \equiv \left\langle \delta(\boldsymbol{k})\delta(\boldsymbol{k})\right\rangle_k$. The BORG ensemble mean and variance are then simply given by:

$$T_{\rm rec}(k) = \left\langle \sqrt{\frac{P^{s'}_{\rm rec}(k)}{P_{\rm true}(k)}} \right\rangle_s, \tag{5.18}$$

$$\sigma^2_T(k) = \left\langle T^{s'}_{\rm rec}(k) - \left\langle \sqrt{\frac{P^{s'}_{\rm rec}(k)}{P_{\rm true}(k)}} \right\rangle_s \right\rangle_s. \tag{5.19}$$

The value of $T_{\rm rec}(k)$ can, technically, vary between 0 and 1 with $T_{\rm rec}(k) = 1$ implies the inference is *unbiased*. In the reconstruction context, Eq. (5.17) can be thought of as a "transfer function".

Below we subsequently vary each element in Section 5.1 that could potentially affect the quality of the reconstruction, while keeping the rest of the setup fixed. For clarification, we summarize these factors below.

1. Tracer number density: $\overline{n}_h = 2.91 \times 10^{-5}$, 1.10×10^{-4}, $3.34 \times 10^{-4}\,(\,h\,{\rm Mpc}^{-1})^3$.
2. Grid resolution: $L_{\rm grid} = 31.25$, 20.83, 15.625 h^{-1}Mpc.
3. Forward model: LPT and 2LPT.
4. Bias model: power-law, broken power-law, and second-order bias.

5.4.1 Tracer number density

All results in this section are obtained with the following setup:

- 2LPT forward model.
- Power-law bias model.
- $L_{\rm grid} = 15.625\ h^{-1}$Mpc.

The only varying factor is the tracer number density as we consider three GADGET-2 DM halo catalogs divided into mass bins: $\log_{10} M_h = [13.0, 13.5)$, $[13.5 - 14.0)$, $[14.0 - 14.5)\,h^{-1}M_\odot$. Their corresponding number densities are $\overline{n}_h = 3.34{\times}10^{-4}$, $1.10{\times}10^{-4}$, $2.91{\times}10^{-5}\,(\,h\,{\rm Mpc}^{-1})^3$.

$\overline{n}_h$ $[(\,h\,{\rm Mpc}^{-1})^3]$	$\log_{10} M_h$ $[h^{-1}M_\odot]$	k_{80} $[\,h\,{\rm Mpc}^{-1}]$	k_{50} $[\,h\,{\rm Mpc}^{-1}]$	k_{20} $[\,h\,{\rm Mpc}^{-1}]$
3.34×10^{-4}	[13.0,13.5)	0.11	0.19	0.24
1.10×10^{-4}	[13.5,14.0)	0.09	0.17	0.22
2.91×10^{-5}	[14.0,14.5)	0.06	0.14	0.19

Table 5.1: We show the wavenumbers k_{80}, k_{50} and k_{20} corresponding to the wave number at which $r(k) = 0.8$, 0.5, 0.2 respectively, for three halo catalogs with different number density.

As can be clearly seen from the top row of Figure 5.2, the quality of the reconstructions is *scale-dependent*, as expected. We confirm that increasing tracer number density increases the quality of the reconstruction, i.e shifts $r(k)$ up, as can be consistently observed in both Figure 5.2 and Table 5.1. The same trend applies for the power bias as it decreases with higher tracer number density.

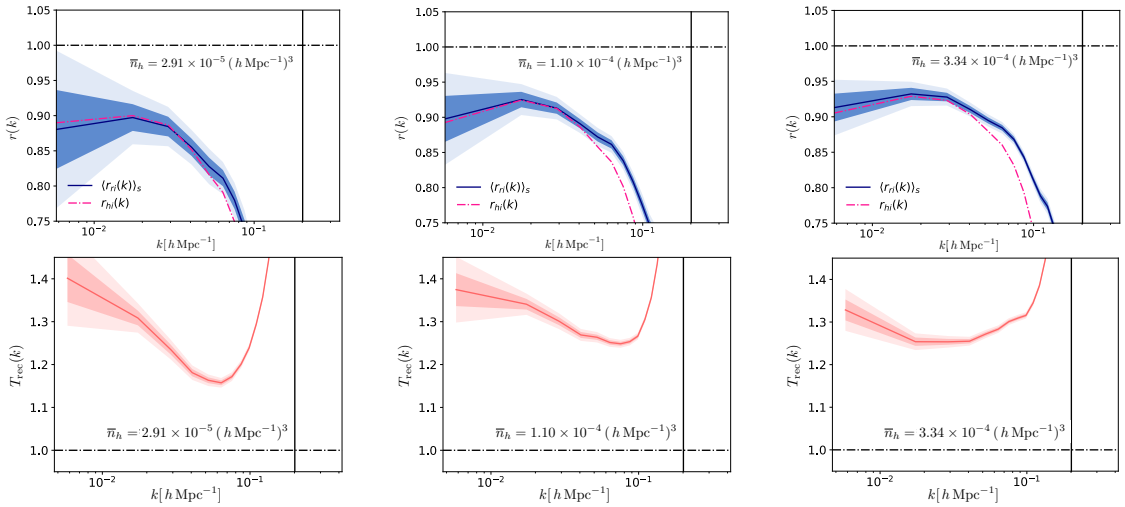

Figure 5.2: Comparison of correlation coefficient (top row) and amplitude bias (bottom), between reconstructed and input initial density modes, for three GADGET-2 halo catalogs of $\log_{10} M_h = [14.0 - 14.5)$, $[13.5 - 14.0)$, $[13.0, 13.5)\,h^{-1}M_\odot$ (left to right). The dark and light shaded regions correspond to the 1σ and 2σ uncertainties. The vertical solid black line denotes the $k_{\mathrm{nyq}} = \pi/L_{\mathrm{grid}}$. On top row, we additionally plot the correlation coefficient between the halo and input initial matter density modes in dot-dashed pink.

We attribute the fact that $r(k)$ is $\sim 10\%$ less than 1 at the smallest k to two factors that appear in the denominator of Eq. (5.15), which we will study further below:

1. incorrect bias model (see Figure 5.3),

2. tracer shot noise (see Figure 5.4).

To investigate the first point, we replace the GADGET-2 halo catalogs by mock catalogs generated directly from the power-law bias model combined with Poisson stochasticity on top of 2LPT-evolved matter density fields. We intentionally match the tracer densities of these mock catalogs to those of the GADGET-2 halo catalogs presented above. Note that we still sample the bias parameters $(\overline{n}_h, \beta)$ during the reconstructions. Figure 5.3 shows that, when the underlying mechanism that generates the input data is the same as the data model used for the reconstruction, the correlation coefficient r_{ri} is greatly improved and the amplitude bias T_{rec} is significantly reduced for all three tracer densities ($\lesssim 5\%$ for the case of highest tracer density) as there is no mismatch between the inferred and true bias parameters.

However, even in this particular case, there is still a limit set by the tracer stochasticity which is related to the second point above. To illustrate this effect of shot noise, we run the reconstruction on another mock catalog with an even higher tracer density of $\overline{n}_h = 3.41 \times 10^{-3}\,(h\,\mathrm{Mpc}^{-1})^3$ and show the result in Figure 5.4. Clearly, both $r(k)$ and $T_{\mathrm{rec}}(k)$ now approach 1 at small k as the effect of shot noise is significantly suppressed, as predicted by Eq. (5.15).

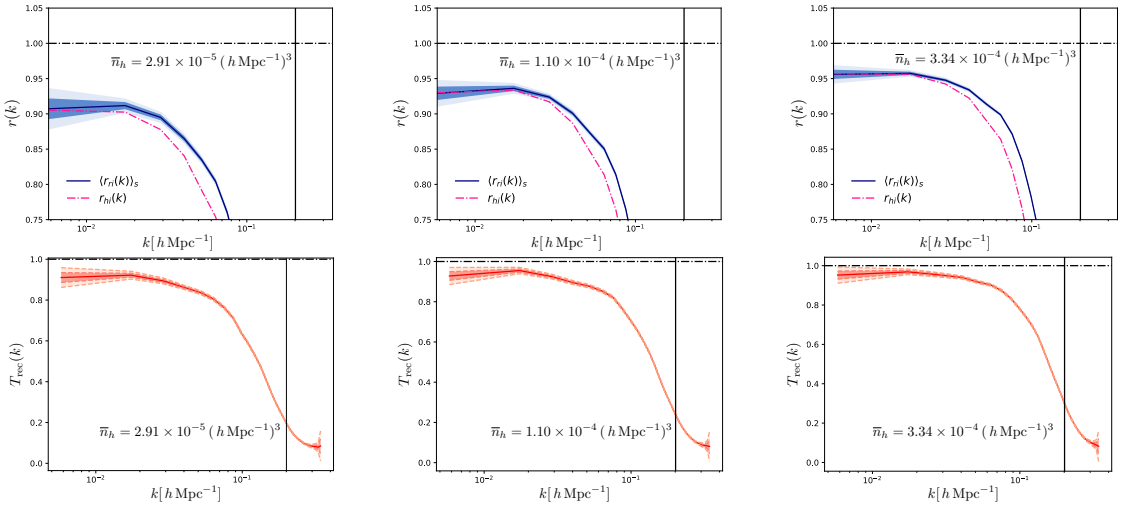

Figure 5.3: Correlation coefficient (top row) and amplitude bias (bottom) between reconstructed and input initial density modes. The mock catalogs are generated on top of a 2LPT-evolved matter density field by applying the same power-law deterministic bias model and Poisson stochasticity that are used later in the BORG reconstructions. We show, from left to right, results for different tracer densities of $\overline{n}_h = 2.91 \times 10^{-5}$, 1.10×10^{-4}, $3.34 \times 10^{-4}\,(h\,\mathrm{Mpc}^{-1})^3$.

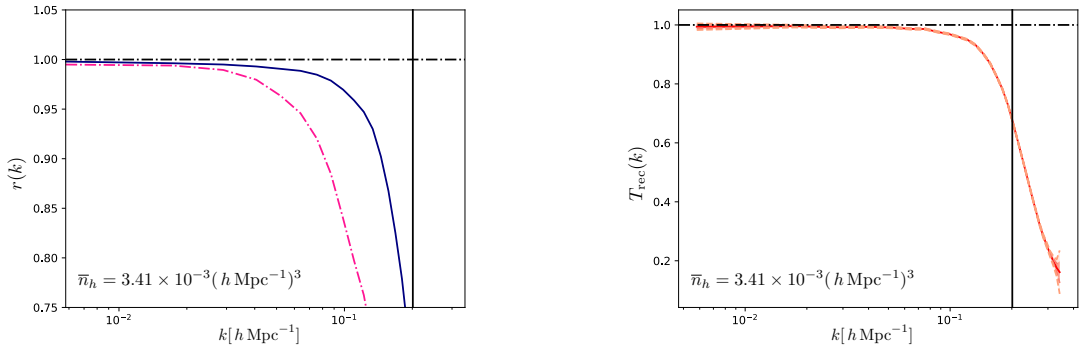

Figure 5.4: Correlation coefficient (left) and amplitude bias (right) for BORG reconstruction where the input mock catalog has an enhanced tracer number density of $\overline{n}_h = 3.41 \times 10^{-3}\,(h\,\mathrm{Mpc}^{-1})^3$. Note the uncertainty regions are now shrank significantly as the high density tracers provide much tighter constraints on both r_{ri} and T_{rec}.

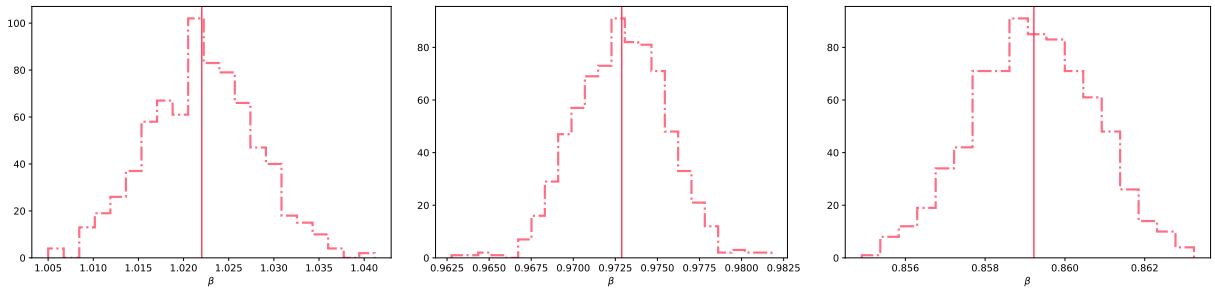

Figure 5.5: Histograms of the power-law index β – the only bias parameter in the power-law bias model – as inferred by BORG in the three reconstructions using DM halos of (from left to right) $\log_{10} M_h = [14.0 - 14.5)$, $[13.5 - 14.0)$, $[13.0, 13.5)\, h^{-1} M_\odot$ as tracers. The vertical dash line represents the BORG ensemble mean.

It should also be mentioned that, within each mass bin, we do not weight the halos by their massess but instead treating them all as equal mass tracers. For analysis in which the tracer mass is available, it is advisable to weight the halos by mass to reduce their shot noise [148, 149].

Additionally, we notice a rather *scale-independent* bias in the amplitude of reconstructed and true initial density modes. This trend can be universally observed in the rest of the results in this chapter. We attribute this to the fact that the real-space Poisson likelihood tends to underestimate the first-order bias parameters, i.e. β and b_1, for the range of halos considered in this work. As shown in Figure 5.5 specifically for the power-law bias model currently considered, $\langle\beta\rangle < 1$ for the first two halo mass bins while $\simeq 1$ for the most massive bin. On the other hand, the large-scale values of b_1 for these halo mass bins can be computed from the best-fit in [150, 151], for our cosmology, as $b_1 = 1.1742, 1.5597, 2.2690$, respectively. These deviations all point to significant contributions from second- and higher-order terms of the evolved matter density field in Eqs. (5.1)–(5.2).

It is worth pointing out here that there is a small yet important difference between the large-scale values of bias parameters b_O^L, i.e. bias parameters measured through n-point function, and the values inferred by count-in-cell data from BORG. The latter is often referred to as *moment bias* $b_O^m(L_{\rm grid})$, and it depends on the smoothing scale $L_{\rm grid}$ in BORG. At leading-order, however, the relation $b_1^L = b_1^m$ should hold so long as non-linear correction of order $\sigma^2(L_{\rm grid})$ is negligible (see Section 4.2 of [65] for a detailed discussion). We note that for the input data employed in this study, $\sigma^2(L_{\rm grid}) \simeq 0.09 - 0.33$ respectively for $L_{\rm grid} = 31.25 - 16.625\ h^{-1}$Mpc.

5.4.2 Grid resolution

All results in this section are obtained with the following setup:

- 2LPT forward model.

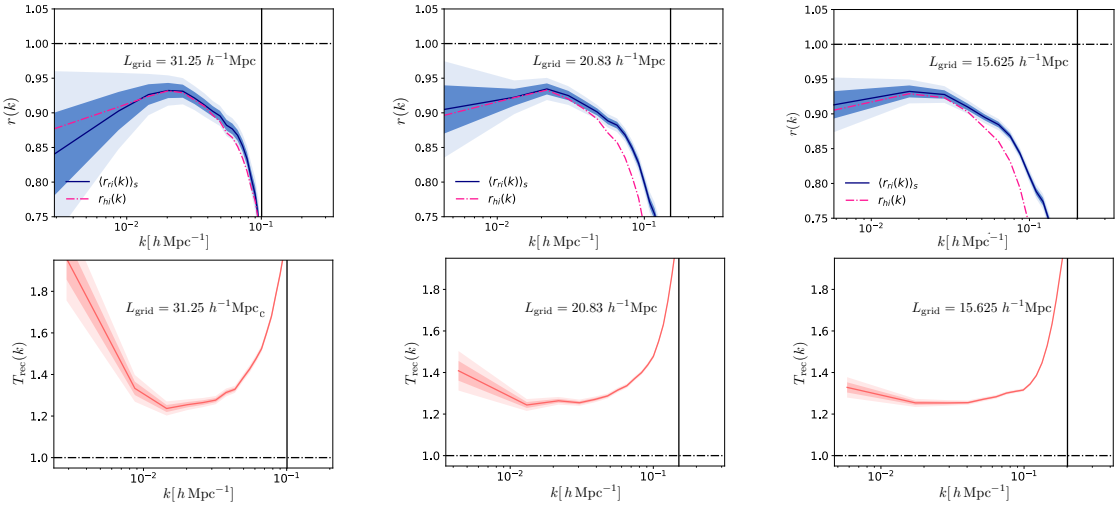

Figure 5.6: Comparison of correlation coefficient (top row) and amplitude bias (bottom row) for three grid resolutions of $L_{\rm grid} = 31.25,\ 20.83,\ 15.625\ h^{-1}{\rm Mpc}$ (left to right) corresponding to $N_{\rm grid} = 64^3,\ 96^3,\ 128^3$, respectively. In all three cases, we use the same GADGET-2 halo catalog of $\log_{10} M_h = [13.0, 13.5)\, h^{-1} M_\odot$ as input data, for maximum tracer density. The uncertainty regions and reference lines are defined similar to Figure 5.2.

- Power-law bias model.
- DM halos, $\log_{10} M_h = [13.0, 13.5) h^{-1} M_\odot$.

The only varying factor here is the grid resolution. We run three reconstructions at grid sizes of $L_{\rm grid} = 31.25,\ 20.83,\ 15.625\ h^{-1}{\rm Mpc}$. Increasing the resolution, i.e. decreasing the grid size $L_{\rm grid}$, reduces the number of small scale modes the stochasticity has to absorb[5] – at the expense of computational resources. However, there exists a limiting physical scale under which LPT and 2LPT will break down.

In Figure 5.6, we observe the same scale-dependent behavior of $r(k)$ in Figure 5.6. Interestingly, higher resolutions improve $r(k)$ and $T_{\rm rec}(k)$ even at small k (see Table 5.2), albeit rather marginal going from $L_{\rm grid} = 20.83\ h^{-1}{\rm Mpc}$ to $L_{\rm grid} = 15.625\ h^{-1}{\rm Mpc}$. This trend suggests that the real-space Poisson likelihood is not a good model for halo stochasticity at large grid sizes.

5.4.3 Forward model

All results in this section are obtained from the following setup:

- Power-law bias model.

[5]It should be noted that, for higher resolutions, although the *unfiltered* evolved matter density fields in the reconstruction will also be closer to the true, unfiltered matter density field at redshift zero, what really matters for the reconstruction is only the density *filtered* by the CIC kernel specified by $L_{\rm grid}$.

$L_{\rm grid}$ $[\,h^{-1}{\rm Mpc}]$	$N_{\rm grid}$	k_{80} $[\,h\,{\rm Mpc}^{-1}]$	k_{50} $[\,h\,{\rm Mpc}^{-1}]$	k_{20} $[\,h\,{\rm Mpc}^{-1}]$
31.25	64^3	0.08	0.13	0.16
20.83	96^3	0.10	0.16	0.20
15.625	128^3	0.11	0.19	0.24

Table 5.2: We show the wavenumbers k_{80}, k_{50} and k_{20} corresponding to the wave number at which $r(k) = 0.8,\ 0.5,\ 0.2$ respectively, for the halo mass bin $\log_{10} M_h = [13.0, 13.5)\, h^{-1} M_\odot$, at three resolutions, from left to right, $L_{\rm grid} = 31.25,\ 20.83,\ 15.625\ h^{-1}{\rm Mpc}$.

- Poisson likelihood.
- DM halos, $\log_{10} M_h = [13.0, 13.5) h^{-1} M_\odot$.
- $L_{\rm grid} = 15.625\ h^{-1}{\rm Mpc}$.

We run two reconstructions using LPT and 2LPT as forward model. In terms of numerical performance, we assert that LPT significantly reduces the computation resources require while maintaining the same convergence rate.

We find the quality of reconstructions using LPT mirror that of reconstructions running with 2LPT extremely well, as shown in Figure 5.7 (see also Table 5.3), even at the highest resolution considered here, $L_{\rm grid} = 15.625\ h^{-1}{\rm Mpc}$. Both exhibit the same behavior in term of power bias, as illustrated in Figure 5.7. The ensemble means of the reconstructed initial density fields using LPT and 2LPT are $\gtrsim 99\%$ spatially-correlated. We additionally verify that the same trend applies for reconstruction using the broken power-law bias instead of the simple power-law bias model; the results from LPT and 2LPT agree down to the k_{80}, k_{50}, k_{20} values, for both bias models. We summarize the results in Table 5.3.

This finding, couples with results from varying grid resolutions in the previous section, strongly suggest that a field-level Bayesian forward modeling such as BORG is actually applicable for future surveys where the grid resolution will be anyway limited due to the larger volumes. In the next chapter, we will see such an application on BOSS/SDSS3 data.

Forward model	k_{80} $[\,h\,{\rm Mpc}^{-1}]$	k_{50} $[\,h\,{\rm Mpc}^{-1}]$	k_{20} $[\,h\,{\rm Mpc}^{-1}]$
LPT	0.11	0.19	0.24
2LPT	0.11	0.19	0.24

Table 5.3: We show the wavenumbers k_{80}, k_{50} and k_{20} corresponding to the wave number at which $r(k) = 0.8,\ 0.5,\ 0.2$ respectively, for reconstructions using LPT and 2LPT forward model.

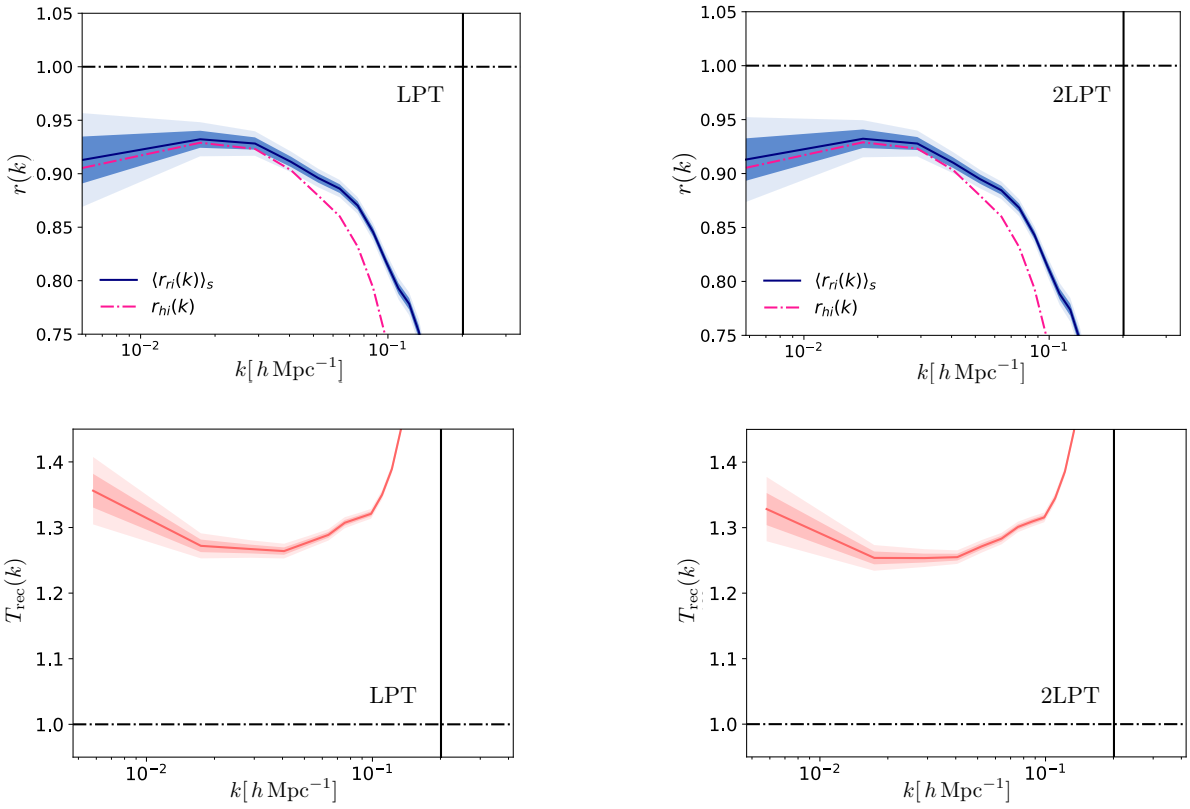

Figure 5.7: Comparison of correlation coefficient (top row) and amplitude bias (bottom row) from reconstructions using LPT (left) and 2LPT (right) as forward model. Both reconstructions use the GADGET-2 halo catalog of $\log_{10} M_h = [13.0, 13.5)\, h^{-1} M_\odot$ as input data, and the power-law bias model coupled with the Poisson likelihood. They also share the same grid resolution of $L_{\rm grid} = 15.625\ h^{-1}$Mpc, the highest resolution considered in our analysis.

5.4.4 Bias model

All results in this section are obtained with the following setup:

- LPT forward model.
- DM halos, $\log_{10} M_h = [13.0, 13.5) h^{-1} M_\odot$.
- $L_{\rm grid} = 15.625\ h^{-1}$Mpc.

We consider here three deterministic bias models which include the power-law, broken power-law and second-order bias.

Surprisingly, all the bias models considered here essentially perform equally (see Figure 5.8 and Table 5.4), suggesting (again) that the key factor controlling the reconstructions is the likelihood.

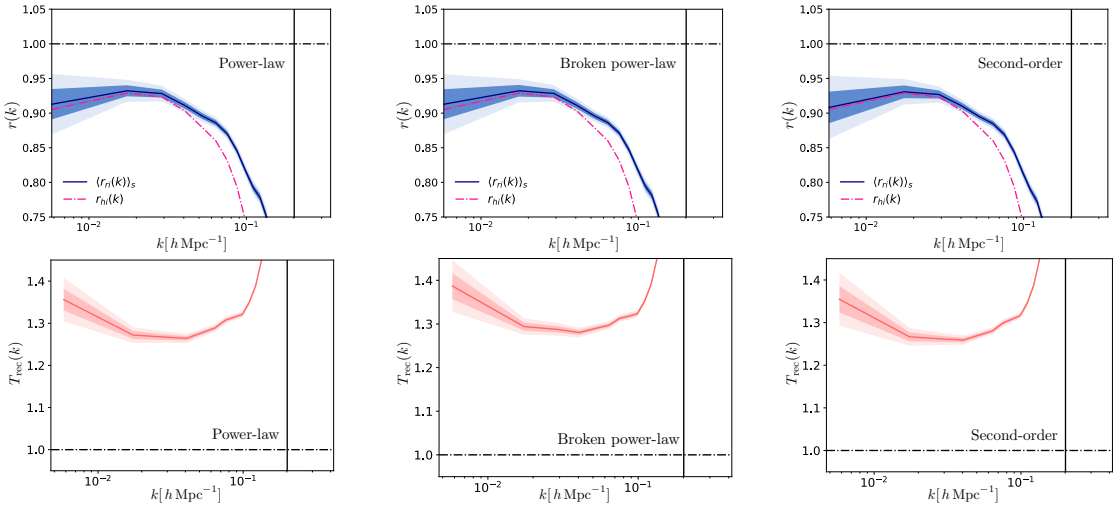

Figure 5.8: Comparison of correlation coefficient (top row) and amplitude bias (bottom row) from reconstructions using three (deterministic) bias models: power-law, broken power-law, and second-order bias (left to right). All three use the same input GADGET-2 halo of $\log_{10} M_h = [13.0, 13.5)\, h^{-1} M_\odot$, the same LPT forward model, and the same grid resolution of $L_{\rm grid} = 15.625\ h^{-1}$Mpc.

Bias model	k_{80} $[\,h\,{\rm Mpc}^{-1}]$	k_{50} $[\,h\,{\rm Mpc}^{-1}]$	k_{20} $[\,h\,{\rm Mpc}^{-1}]$
Power-law	0.11	0.19	0.24
Broken power-law	0.11	0.19	0.24
Second-order	0.11	0.18	0.30

Table 5.4: We show the wavenumbers k_{80}, k_{50} and k_{20} corresponding to the wave number at which $r(k) = 0.8, 0.5, 0.2$ respectively, for GADGET-2 halos in $\log_{10} M_h = [13.0, 13.5)\, h^{-1} M_\odot$, at $L_{\rm grid} = 15.625\ h^{-1}$Mpc, with three different (deterministic) bias models.

5.5 Discussion and conclusion

We have presented an investigation into the impact of various ingredients in the quality of the initial conditions reconstructed by Bayesian forward modeling of clustering of DM halos in GADGET-2 simulation. For this purpose, we have implemented in the BORG framework a non-local, second-order bias model based on the EFT approach to galaxy clustering (see Section 3.1.1 and [65]). We further quantify the performance of each reconstruction via a Fourier-space correlation coefficient $r(k)$ between inferred and input initial density modes. We additionally measure the bias in the amplitude of reconstructed initial fluctuations.

Universally, we have found that the real-space, Poisson likelihood underestimates the linear bias factor which in turn induces a scale-independent bias on the amplitude of the reconstructed modes. Further investigation is needed to examine whether this effect is specific to the Poisson likelihood or general for all the real-space likelihoods. The only implication from our results is that non-linear corrections are important for the grid sizes considered here. In Chapter 7, we will derive and study a *Fourier-space* likelihood which applies a *sharp-k* filter on the halo and evolved matter fields in order to keep these non-linear contributions under rigorous control.

We further confirm that increasing (decreasing) tracer number density shifts the correlation up/down and reduces (enhances) amplitude bias of the inferred initial conditions. Similar conclusions can be drawn for grid resolutions, albeit to a lesser extent. Care should be taken in choosing the grid size to balance between performance and computational cost.

The choices of gravitational forward model considered here, being LPT and 2LPT, do not significantly affect the performance of BORG reconstructions, even for fairly small grid size $L_{\rm grid} = 15.625\ h^{-1}$Mpc. It would be interesting to push this limit by increasing the number of grid size $N_{\rm grid}$, and/or using a smaller simulation box. This is of particular importance since information from small, non-linear scales, if modeled correctly, might improve the reconstruction even on large scales, as seen here (see also [152, 153]. Similarly, the performances of three deterministic bias models examined in this chapter are almost identical up to $L_{\rm grid} = 15.625\ h^{-1}$Mpc.

To our knowledge, this is the first stringent test for Bayesian forward modeling of halo clustering implemented in BORG. Our input data is taken directly from GADGET-2 simulations instead of being generated from the same data model used in the analysis (as done in [39, 127, 154]). A more realistic test case would also include survey geometries and selection effect on top of a full N-body simulation. In addition, further investigation of the likelihood is also an interesting topic, given the sub-optimal performance of the Poisson likelihood employed here. We defer such studies for future work.

Chapter 6

Inference of galaxy cluster large-scale bulk flow and measurement of kinematic Sunyaev-Zel'dovich effect

In this chapter, we introduce a concrete astrophysical application of the Bayesian forward inference approach to galaxy clustering data: measurement of the kinematic Sunyaev-Zel'dovich (kSZ) effect in the CMB data using the Bayesian forward inferred large-scale bulk flow of galaxy clusters. The project described in this chapter is currently in the final stage, being prepared for publication. Results presented here and conclusions drawn from them are expected to be consistent with the final paper.

This chapter is structured as follows. In Section 6.1, we review the theory and motivation behind the study of kSZ effect, including its theoretical model, its potential as a cosmological probe, previous measurements in the literature, and what Bayesian forward inference could add to this picture. In Section 6.2, we review the datasets used in this work and the physical modeling of these components. We further introduce the general data model and the kSZ likelihood in Section 6.3. We report our measurements of the kSZ effect and the large-scale bulk flow of maxBCG clusters, with their associated uncertainties, in Section 6.4. Two null tests are performed and described in Section 6.5. Finally, we summarize our findings and discuss potential future improvements of kSZ measurement in Section 6.6. Other relevant details are documented in Appendix G–Appendix I.

6.1 Theory and motivation

As Cosmic Microwave Background (CMB) photons pass through a cloud of free electrons, they might gain or lose energy due to Compton scattering processes. These phenomena leave imprints of cosmic large-scale structure (LSS), specifically ionized baryonic gas inside clusters of galaxies, on the primary CMB temperature anisotropies, and is referred to collectively as Sunyaev-Zel'dovich (SZ) effects [155]. In specific, the *kinematic* SZ (kSZ) effect designates the temperature anisotropies induced by the coherent bulk motion of the

ionized, free electron gas (with respect to the CMB or the comoving observer rest frame[1]). Due to peculiar motion (of the galaxy cluster), in the cluster electron gas rest frame, the primary CMB radiation appears anisotropic, hence the Compton scattering[2] would marginally re-isotropize this radiation. This, in the CMB or comoving observer rest frame, results into the radiation field appears slightly anistropic along the LOS of the scattering electron gas. The anisotropies induced by such coherent motion, in the non-relativistic limit, simply appear as a shift in the CMB blackbody temperature – as opposed to the spectral distortion imprinted by the random thermal motion of free electron, referred to as *thermal* SZ (tSZ) effect. Let us consider a single point source, located at position $\boldsymbol{x}$ along direction $\hat{\boldsymbol{n}}$ on the sky. At first-order, the kSZ signal imprinted by the cluster onto the CMB is given by [155, 156, 157, 158]:

$$\frac{\Delta T_{\rm kSZ}(\hat{\boldsymbol{n}})}{T_0} = -\sigma_T \int {\rm d}l \, \left(\frac{\boldsymbol{v}_e(\boldsymbol{x})\cdot\hat{\boldsymbol{n}}}{c}\right) n_e(\boldsymbol{x}) \tag{6.1}$$

where $T_0 = 2.725\times10^6\,\mu K$ is the CMB blackbody temperature, $n_e(\boldsymbol{x})$ is the free electron number density at position $\boldsymbol{x}$, and $\boldsymbol{v}_e(\boldsymbol{x})$ is the peculiar velocity of free electrons. σ_T and c denote the Thomson scattering cross-section and the speed of light in vacuum, respectively. The integral $\int dl$ is performed along the line-of-sight (LOS) $\hat{\boldsymbol{n}}$. It is generally assumed that the bulk motion of free electrons and their host clusters follows the large-scale motion of Dark Matter (DM) [65], i.e. $\boldsymbol{v}_e = \boldsymbol{v}_{DM} = \boldsymbol{v}$, and since the correlation length of the latter is much larger than the physical size of a typical galaxy cluster [159], Eq. (6.1) could be further reduced to

$$\frac{\Delta T_{\rm kSZ}(\hat{\boldsymbol{n}})}{T_0} = -\tau(\boldsymbol{x},\hat{\boldsymbol{n}})\left(v^{\rm LOS}(\boldsymbol{x},\hat{\boldsymbol{n}})/c\right), \tag{6.2}$$

where

$$v^{\rm LOS}(\boldsymbol{x},\hat{\boldsymbol{n}}) = \boldsymbol{v}(\boldsymbol{x})\cdot\hat{\boldsymbol{n}} \tag{6.3}$$

denotes the velocity along the LOS $\hat{\boldsymbol{n}}$, and we have defined

$$\tau(\boldsymbol{x},\hat{\boldsymbol{n}}) = \sigma_T \int {\rm d}l \; n_e(\boldsymbol{x}) \tag{6.4}$$

to be the LOS projected optical depth. Eq. (6.4) is written for a point source. If a cluster is resolved, we need to integrate Eq. (6.4) over the size θ_i of the cluster. For example, assuming a spherically symmetric Gaussian profile for the electron gas, Eq. (6.4) becomes

$$\tau(\boldsymbol{x}_i,\theta) = \frac{\tau_{0,i}}{\sqrt{2\pi\theta_i^2}}\int_0^\theta d^2\theta e^{-\theta^2/2\theta_i^2}, \tag{6.5}$$

where $\tau_{0,i}$ is the *integrated* optical depth specific for cluster i. We will return to this in more details in Section 6.2.2 and the discussion following Eq. (6.16).

[1]The two frames are equivalent in such a homogeneous and isotropic Universe described by the FLRW metric.

[2]For typical clusters, the non-relativistic limit $v_{\rm pec}/c \ll 1$ applies, and the kinematic SZ effect is better described by Thomson scattering, which leads to a great simplification in relevant computations.

A measurement of the kSZ signal thus directly constrains the product $(\boldsymbol{v}(\boldsymbol{x}) \cdot \hat{\boldsymbol{n}})\tau$ on the r.h.s. of Eq. (6.2). Assuming an external constraint on τ (see, e.g. [160]), the kSZ signal allows for a measurement of the peculiar velocity field $\boldsymbol{v}(\boldsymbol{x})$ – which, in comparison to that of its divergence $\vec{\nabla} \cdot \boldsymbol{v}(\boldsymbol{x})$, is less affected by nonlinear evolution – and hence for a test of modified gravity and Dark Energy models [161, 162, 163] as well as a constraint on the sum of neutrino masses [164].

On another hand, if one could measure or reconstruct the peculiar velocity field $\boldsymbol{v}(\boldsymbol{x})$, e.g. from GRS data, measurement of the kSZ signal then offers a probe for τ, i.e. the abundance and profile of free baryons inside galaxy clusters [165, 166]. Two unique features of the kSZ signal that one could notice from Eq. (6.2) is that it does not depend on temperature of the gas and only scales linearly with the gas density. These indeed make kSZ measurement the perfect candidate for detecting and measuring the otherwise elusive Warm-Hot Intergalactic Medium (WHIM) [167, 168, 165, 166] – a diffuse form of free baryonic gas with temperatures of $10^5 - 10^7 K$ which is too cold to show up in X-ray, and even tSZ measurements – that is suggested to host a large fraction of baryons in our Universe that are still missing compared to the number predicted by our standard cosmological model [167, 168].

Both approaches described above will benefit tremendously from upcoming high-resolution CMB experiments and large-volume galaxy redshift surveys. Indeed, the advent of CMB-S4 [29] will allow for the application of kSZ tomography, i.e. measurements of kSZ signal at different redshifts. The result of which could then be cross-correlated with datasets from DESI [18] or LSST [20] to, for example, either precisely map out the radial component of peculiar velocity field, which in turn would significantly improve current constraints on scale-dependent galaxy bias and *local* primordial non-Gaussianities [169], or tightly constrain the cosmic gas fraction and models of ionized gas evolution [170].

Consequently, despite the fact that the kSZ signal is deeply buried beneath the primary CMB anisotropies, there have been several attempts to measure this effect using various datasets and estimators. These efforts have resulted in $\simeq 2-4\sigma$ evidence of the kSZ effect using the kSZ *pairwise* estimator [171, 159, 172, 173] and the cross-correlation between CMB maps and reconstructed velocity field [159, 174, 166].

Previous analyses of the kSZ (see, e.g. [159, 174, 166]) usually derived $\boldsymbol{v}(\boldsymbol{x})$ – assuming a certain cosmology with Hubble parameter H and cosmic linear growth rate $f = d\ln\delta/d\ln a$ – from the inversion of the *linearized* continuity equation in either real-space $\boldsymbol{x}$ [159]

$$\boldsymbol{\nabla} \cdot \boldsymbol{v}(\boldsymbol{x}) = -aHf\delta(\boldsymbol{x}), \tag{6.6}$$

or redshift-space $\boldsymbol{s}$ [174], ,

$$\boldsymbol{\nabla} \cdot \boldsymbol{v}(\boldsymbol{s}) + f\boldsymbol{\nabla} \cdot [(\boldsymbol{v}(\boldsymbol{s}) \cdot \hat{\boldsymbol{n}})\,\hat{\boldsymbol{n}}] = -aHf\delta(\boldsymbol{s}), \tag{6.7}$$

where the DM density field $\delta(\boldsymbol{x})$ is simply obtained from a smoothed galaxy density field $\delta_g(\boldsymbol{x})$ by assuming a *local*, *linear* bias relation of the form $\delta(\boldsymbol{x}) = \delta_g(\boldsymbol{x})/b_1$. This simple method, however, ignores various uncertainties and systematics in the velocity reconstruction [159, 165, 174].

Below, we apply our Bayesian inference approach to the measurement of kSZ signal – extracted by traditional/adaptive aperture photometry filters [165, 159, 175] – from the Planck 2018 SMICA CMB map [44] at locations of a sub-sample of maxBCG clusters [42]. In particular, the kSZ signal of each cluster is modeled using the LOS component of its peculiar velocity assigned from velocity fields in samples of the BORG-SDSS3 reconstruction [43]. These samples are essentially the prior in our Bayesian analysis. It is worth emphasizing that the peculiar velocity field reconstruction using the BORG framework naturally takes into account the *non-linear*, *non-Gaussian* formation and dynamics of LSS associated with statistics of the density field beyond the 2-point correlation function, redshift-space distortions and light-cone effects [102, 43]. As such, our analysis not only coherently utilizes the information from higher-order statistics of LSS[3] but also accounts for systematics and uncertainties introduced by those. To our knowledge, this is the first time the uncertainty in velocity reconstruction is properly included in the final uncertainty of the kSZ measurement.

For consistency, throughout we assume the same flat ΛCDM cosmology with the same cosmological parameters as assumed by the BORG-SDSS3 reconstruction in [43], $\Omega_{\rm r} = 0$, $\Omega_K = 0$, $\Omega_{\rm m} = 0.2889$, $\Omega_{\rm b} = 0.048597$, $\Omega_\Lambda = 0.7111$, $w = -1$, $n_s = 0.9667$, $\sigma_8 = 0.8159$, $H_0 = 67.74\,{\rm kms}^{-1}{\rm Mpc}^{-1}$.

6.2 Data sets

6.2.1 CMB data

Planck SMICA CMB map

Our CMB data is the SMICA map from the Planck 2018 release[4] [44] (SMICA2018 hereafter). SMICA(Spectral Matching Independent Component Analysis) [176] linearly combines Planck frequency channels with multipole-dependent weights, including multipoles up to $\ell = 4000$ [44].

It is worth noting that, due to the finite resolution and detector noise associated with any CMB instrument, the observed temperature anisotropies $\Delta T^{\rm obs}$ is a convolution of the true anisotropies[5] ΔT with the instrumental beam function B, plus the instrumental noise $\Delta T_{\rm instr}$, i.e.

$$\Delta T^{\rm obs}(\boldsymbol{\theta_i}, \boldsymbol{\theta}) = \int d\boldsymbol{\theta'}\, \Delta T(\boldsymbol{\theta_i}, \boldsymbol{\theta'}) B(\boldsymbol{\theta_i}, \boldsymbol{\theta} - \boldsymbol{\theta'}) + \Delta T_{\rm instr}(\boldsymbol{\theta}), \tag{6.8}$$

where we have replaced the three-dimensional LOS unit vector $\hat{\boldsymbol{n}}$ by the two-dimensional vector $\boldsymbol{\theta}$ and used $\boldsymbol{\theta_i}$ to denote position of cluster i. The SMICA2018 map has a 5-arcmin

[3] See, e.g. [65, 123], for a general discussion about this topic.

[4] https://pla.esac.esa.int/

[5] This includes both primary, i.e. primordial CMB, and secondary anisotropies, e.g. kSZ, tSZ, integrated Sachs-Wolfe, etc.

resolution which corresponds to a Gaussian beam of

$$B(\vec{\theta_i}, \vec{\theta}) \equiv B(\theta_i, \theta) = \frac{1}{\sqrt{2\pi\theta^2_{\rm beam}}} \exp\left[-\frac{(\theta-\theta_i)^2}{2\theta^2_{\rm beam}}\right] \tag{6.9}$$

where $\theta_{\rm beam} = 5/\sqrt{8\ln(2)} \approx 2.1233'$.

Signal extraction

To extract the kSZ signal from a CMB map, an aperture photometry (AP) filter of radius θ_f is applied at the location of all clusters. Below, specifically in Section 6.3.1,Section 6.4.1, and Section 6.5.1, we adopt an *adaptive* aperture photometry (AAP) filter whose radius $\theta_{f,i}$ scales with the effective apparent size of cluster i. This ensures that the AAP filter always probes the same fraction of baryonic gas for each cluster assuming a universal gas profile.

The extracted flux ΔT^{θ_f} can then be expressed as a convolution of the observed flux $\Delta T^{\rm obs}$ with a radial weight function W^{θ_f} associated with that AP filter, i.e.

$$\Delta T^{\theta_f}(\boldsymbol{\theta_i}) = \int d\boldsymbol{\theta} W^{\theta_f}(\boldsymbol{\theta}-\boldsymbol{\theta_i}) \Delta T^{\rm obs}(\boldsymbol{\theta_i}, \boldsymbol{\theta}) \tag{6.10}$$

The specific form of the weight function W^{θ_f} is given by

$$W^{\theta_f}(\boldsymbol{\theta}-\boldsymbol{\theta_i}) \equiv W^{\theta_f}(\theta-\theta_i) = \begin{cases} 1 & 0 \le (\theta-\theta_i) < \theta_f \\ -1 & \theta_f \le (\theta-\theta_i) < \sqrt{2}\theta_f \\ 0 & \text{otherwise.} \end{cases} \tag{6.11}$$

As can be seen from Eq. (6.11), these *compensated* filters are designed to reduce contributions from primary CMB anisotropies, (and also other low-redshift sources of contamination[6],) which vary on scales larger than θ_f in the extracted flux. Thus, as θ_f increases, the filtered flux will include more contamination from these sources. In our analysis, we measure the kSZ signal as a function of the filter size θ_f. Naturally, we expect the signal to be underestimated at small filter sizes where parts of the signal fall outside the inner disks and thus being subtracted out. Once the whole cluster is encompassed by the filter, we expect the signal to asymptote a limit value[7], while the uncertainties increase due to large-scale noise, as explained above.

In practice, the application of an AP or AAP filter amounts to taking the difference between the pixel-averaged temperature anisotropies within the inner disk and that within the outer ring. For this estimate of the kSZ signal, the primary noise source on large scale

[6]Contributions from structures below redshift range of maxBCG clusters and the BORG-SDSS3 volume might manifest themselves as large-scale anisotropies in the observed CMB.

[7]The exact value of this asymptotic limit depends on various factors, it however should be proportional to the free baryonic fraction within the clusters [174].

is still the primary CMB, while on small scale – where the number of pixels encompassed by the inner disk or outer ring is small – the instrumental noise dominates.

Our method of extracting the kSZ flux is similar to that in [165, 159, 174, 177, 175]. It is worth mentioning here that the typical apparent size of maxBCG clusters selected for our analysis are very close to the Planck beam (see Section 6.2.1 and Section 6.2.2); a filtering technique that is insensitive to details of the cluster gas profile, like the one adopted here, is thus preferable. We defer a more optimal filtering method (which would require more specific assumptions on the form of the gas profile), such as the matched filter[8], to applications on future CMB data with higher-resolutions.

6.2.2 Galaxy cluster data

MaxBCG cluster catalog

Our galaxy clusters are taken from the public version of maxBCG catalog – a *volume-limited*, red-sequence galaxy cluster sample – that includes clusters of scaled richness $N_{200} = 10 - 188$, spanning a redshift range of $z = 0.1 - 0.3$ [42], overlapping with the redshift range of LOWZ galaxy sample. We use the mean richness-mass relation given by Eq. (A15) in [122] (see also the discussion leading up to it) to convert maxBCG cluster scaled richness N_{200} into M_{200}, the cluster projected mass within the R_{200} radius, accounting for the difference in mass definitions and cosmology[9]. If we assume that the projected gas distribution in cluster i, with physical size $R_{200,i}$ and angular diameter distance $D_{A,i}$, can be approximated by a Gaussian profile (cf. Eq. (6.5)), then the width of its profile is given by

$$\theta_{\mathrm{eff},i} = \sqrt{\theta_{200,i}^2 + \theta_{\mathrm{beam}}^2} \tag{6.12}$$

where $\theta_{200,i} = R_{200,i}/D_{A,i}$.

Note that the scatter in the $N_{200} - M_{200}$ relation and any uncertainty in the estimated M_{200} would affect only the signal amplitude but not the signal-to-noise (S/N), which is what really important for the purpose of kSZ detection. The catalog mean cluster mass and redshift are $M_{200} = 1.288 \times 10^{14} M_\odot$ and $z = 0.23$ respectively. Below, we describe various selection cuts that we applied on the original maxBCG catalog, and the resulting sub-sample of maxBCG clusters used in our analysis.

Firstly, to avoid any possible tSZ contamination, we exclude clusters whose $M_{200} > 0.85 \times 10^{14} M_\odot$ from our analysis (see Appendix F). Next, we select only clusters within regions where the BORG-SDSS3 reconstruction are well-constrained by data, i.e. sky regions where LOWZ and CMASS galaxies are actually observed. Finally, we remove clusters outside of the Planck 2018 common confidence mask recommended for temperature analysis, which covers $f_{\mathrm{sky}} = 77.9\%$ [44].

This leaves us with a final sub-sample consisting of 3512 clusters from the original

[8] See, e.g. [178, 172, 166]

[9] See the appendix of [120] and references therein.

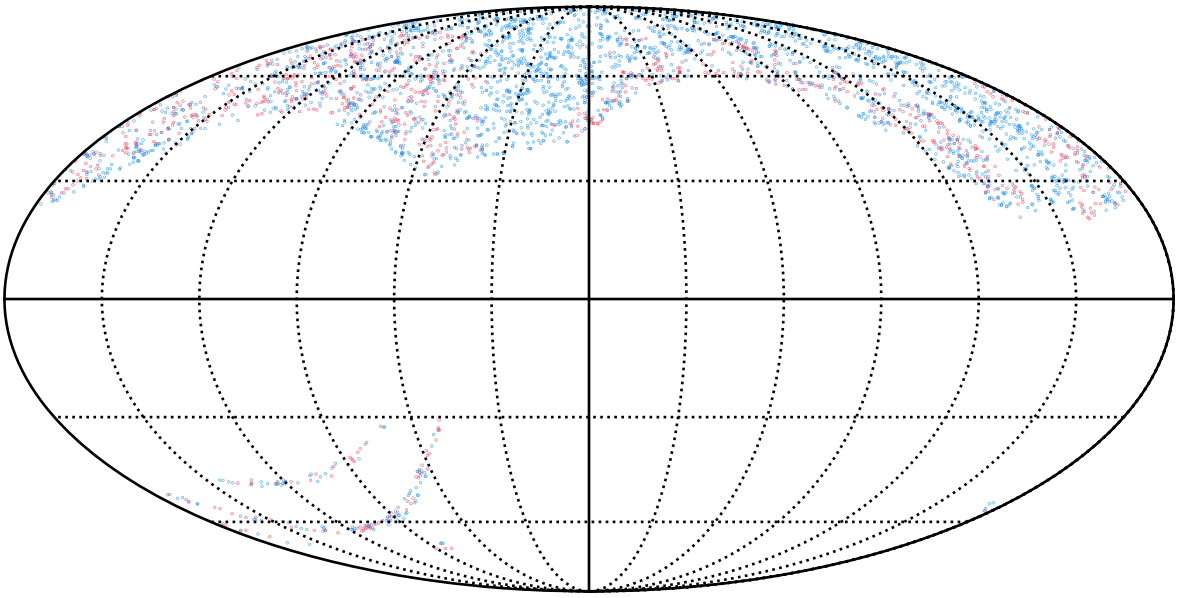

Figure 6.1: HEALPix projected map of 3512 clusters selected for our analysis, with 908 clusters with BCG spectroscopic redshift in red circles and 2604 clusters with photometric redshift in blue circles. The size of the circles are scaled with $\log_{10}(M_{200})$.

maxBCG catalog. We show in Figure 6.1 a HEALPix map[10] [179, 180] of these clusters in Galactic coordinate. We further divide our cluster sample into two datasets:

a) the spectroscopic set including 908 clusters, each has redshift of its brightest cluster galaxy (BCG) member measured by spectroscopy, denoted by $z_{\rm spec}$, and

b) the photometric set including 2604 clusters with only photometric redshift measurements, denoted by $z_{\rm photo}$.

The mean and median mass of our cluster sample (including both datasets) are $M_{200}^{\rm mean} = 7.18\times10^{13}M_\odot$ and $M_{200}^{\rm mean} = 7.31\times10^{13}M_\odot$, while the mean and median redshift are $z^{\rm mean} = z^{\rm median} = 0.257$. This results in a mean and a median apparent size of $\theta_{\rm eff}^{\rm mean} = \theta_{\rm eff}^{\rm median} = 3.9'$. The full histograms of redshift and apparent angular size for both sets of clusters are shown in left and right panels of Figure 6.2, respectively.

[10] `https://healpix.sourceforge.io`

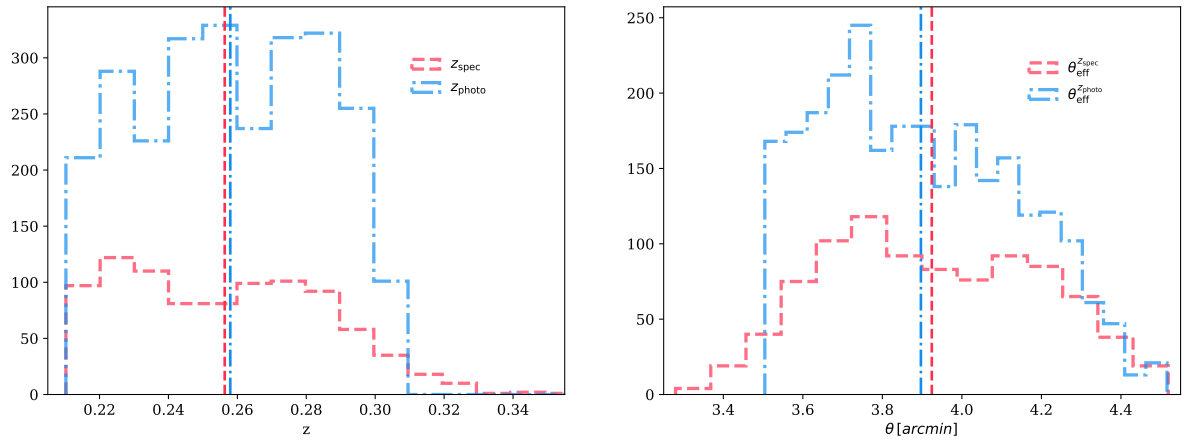

Figure 6.2: Redshift (left) and apparent size (right) distributions of 3512 maxBCG clusters selected for our analysis. The vertical lines present the mean of the $z_{\rm spec}$ and $z_{\rm photo}$ distributions.

Modeling maxBCG cluster optical depth

As the fraction of electrons in neutral gas is presumably small, we will assume that all of baryons in all the clusters being considered here are fully ionized, i.e. setting $f_{\rm free} = 1$ in $N_e = f_{\rm free} f_{\rm gas} M_{200,i}/\mu_e m_p$. We further adopt a universal gas-mass fraction $f_{\rm gas} = f_b \equiv \Omega_b/\Omega_m = 0.16$ following the cosmological baryon abundance and a mean particle weight per electron $\mu_e = 1.17$. Our expression for the integrated cluster optical depth defined in Eq. (6.5) then becomes

$$\tau_{0,i} = \frac{\sigma_T}{D^2_{A,i}} \frac{f_b M_{200,i}}{\mu_e m_p} \tag{6.13}$$

6.2.3 BORG-SDSS3 reconstructed velocity field

BORG-SDSS3 reconstruction

We employ the *non-linear* velocity fields – traced by the SDSS3-BOSS galaxy sample [16] – reconstructed using the BORG algorithm [43, 39]. This reconstruction presents a reconstructed cosmic velocity field within the SDSS volume that fully accounts for many systematics effects, e.g. galaxy bias, light-cone effect, survey geometry, plus other selection and multiplicative systematic effects (see [43] and references therein for further details). The BORG algorithm systematically explores the high-dimensional parameter space $N_{\rm dim} = 256^3$ consisting of the three-dimensional initial conditions at $z \sim 1000$, augmented by the galaxy bias parameters. This is only possible thanks to the introduction and development of Hamiltonian Monte Carlo (HMC) sampling technique (see Appendix B) for LSS inference [40, 39, 102, 43]. The final density fields at $z = 0$ is linked to the initial conditions by a LPT forward model (see Section 2.2.2). BORG thus searches for three-dimensional matter distributions that are physically compatible with the constraints from the observed galaxy

distribution. The result is a fully probabilistic inference of the cosmic matter and velocity fields, taking into account known and unknown systematic effects within some pre-defined limits. The setup of the inference is given as follows. The initial conditions are generated on a comoving grid consisting of 256^3 cells and covering a comoving volume of $4000^3\,h^{-3}$ Mpc^3, which results into a grid resolution of $L_{\rm grid} = 15.624\,h^{-1}$Mpc. The SDSS3-BOSS data are projected in a sub-volume with the observer located at $\boldsymbol{x} = \{200, 0, -1700\}\,h^{-1}$Mpc with respect to the center. There was a total number of 10360 MCMC samples collected [43].

Initial power-spectrum analysis in [43] found that the MCMC chain converged after ~ 1000 samples. Here we are interested in the convergence of the late-time velocity field hence we further remove all samples whose identifier is less than $s = 2000$. To facilitate the storage and process of the MCMC samples, the chain is thin by a factor of 10, more details can be found in [43] (see also Appendix D for similar steps applied for the reconstructions in Chapter 5). Taking the BORG-constrained initial conditions of the SDSS3 volume [43] as input, we run DM-only simulations for the same cosmological parameters used in the reconstruction using a PM algorithm with CIC projection of particles (BORG-PMCIC hereafter), to obtain the large-scale velocity field at the maxBCG catalog mean redshift $z = 0.23$. This includes 837 BORG-PMCIC simulations, one for each of our BORG-constrained initial conditions, with a resolution of $N_{\rm part} = 1024^3$. These are used for the estimation of cluster LOS velocities as well as their uncertainties.

We additionally generate a GADGET-2 [83] simulation at resolution of $N_{\rm part} = 2048^3$ from initial conditions specified by sample $s = 9000$ of BORG-SDSS3 reconstruction (see Appendix I for details). We use this full N-body, high resolution simulation to speficically:

1. estimate the small-scale motion of clusters unresolved by the BORG-SDSS3 reconstruction and BORG-PMCIC re-simulation (see Section 6.2.3),

2. verify that our kSZ estimators are unbiased (see Section 6.3.1),

3. measure the cluster signal profile (see Section 6.3.2).

Modeling the large-scale bulk flows of galaxy clusters

We model the large-scale bulk flow of galaxy clusters in Eq. (6.2) as a sum of two components

$$v^{\rm LOS}(\boldsymbol{x}, \hat{\boldsymbol{n}}) = v_L^{\rm LOS}(\boldsymbol{x}, \hat{\boldsymbol{n}}) + \epsilon_S^{\rm LOS}(\boldsymbol{x}, \hat{\boldsymbol{n}}) \tag{6.14}$$

where $v_L^{\rm LOS}$ is the *large-scale* LOS bulk-flow estimated from the BORG-SDSS3 reconstruction posterior while $\epsilon_S^{\rm LOS}$ is the unresolved *small-scale* LOS velocity. We further assume that, for all clusters:

$$\epsilon_S^{\rm LOS} \sim \mathcal{N}\left(0, \sigma^2_{\epsilon_S^{\rm LOS}}\right) \tag{6.15}$$

where $\sigma^2_{\epsilon_S^{\rm LOS}} \simeq 4.7 \times 10^4\,{\rm km^2 s^{-2}}$ is estimated from the variance of $\left(v^{\rm LOS}_{{\rm NBODY},i} - v^{\rm LOS}_{{\rm BORG\text{-}PMCIC},i}\right)$ distribution (see Figure 6.3) in which $v^{\rm LOS}_{{\rm NBODY},i}$ and $v^{\rm LOS}_{{\rm BORG\text{-}PMCIC},i}$ refer to the LOS velocity of halo i as respectively measured from the previously mentioned GADGET-2 simulation and from BORG-PMCIC simulation of the same BORG-SDSS3 sample, $s = 9000$.

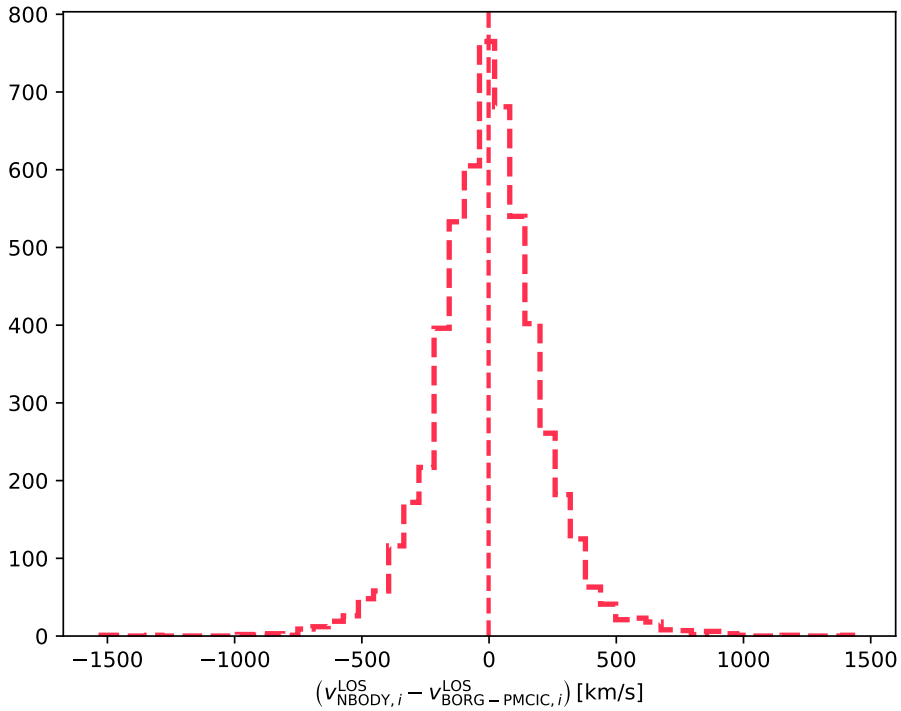

Figure 6.3: Histogram of $\left(v^{\rm LOS}_{{\rm NBODY},i} - v^{\rm LOS}_{{\rm BORG\text{-}PMCIC},i}\right)$ measured from DM halos within LOWZ-CMASS sky regions in our GADGET-2 and BORG-PMCIC simulations of BORG-SDSS3 sample $s = 9000$. We apply the same redshift and mass cuts that are applied for maxBCG clusters. The vertical line represents the sample mean. The sample variance is measured at $\sigma^2 = 4.67 \times 10^4\,{\rm km}^2{\rm s}^{-2}$.

6.3 Analysis methods

Given a BORG ensemble of inferred large-scale LOS velocity for each galaxy cluster i, $v_{L,i}^{\rm LOS}$, our goal is to construct a likelihood for the extracted kSZ signals at locations of all galaxy clusters in our sample. Below, we will derive this likelihood in two cases of input data.

1. The single-cluster signal is extracted at individual physical scales. This yields multiple measurements of the signal, each using information from a specific scale.

2. The single-cluster signal is simultaneously extracted at multiple scales. This yields one single measurement of the signal combining information from all scales.

While the former can be applied for an analysis focusing on using the kSZ effect to study galaxy cluster gas profile, we expect the latter to be less sensitive to CMB noise at large filter sizes, as information from all scales is combined. Our derivations assume that the kSZ measurements at individual cluster locations are independent, i.e. there is not significant overlapping between the AP/AAP filters. We verified this is indeed the case for our cluster sample selected from the maxBCG catalog (see Section 6.2.2).

6.3.1 kSZ likelihood: individual-scale signal

Let us express our data model as

$$\Delta T_{{\rm kSZ},i}^{\theta_f}/T_0 = -\alpha^{\theta_f}\,\tau_i\left(v_{L,i}^{\rm LOS}/c\right) - \tau_i\left(\epsilon_{S,i}^{\rm LOS}/c\right) + \epsilon_{0,i}^{\theta_f}\,, \tag{6.16}$$

where we have introduced α^{θ_f} as the amplitude of the kSZ signal and the large-scale bulk flow of cluster i. Here and below, we simply use $\tau_i = \tau_{0,i}$, letting α^{θ_f} absorbs all specific details about the cluster gas profile.

We expect to measure a value of α^{θ_f} consistent with zero in the case of no detection, whereas a value of order of unity at a filter size that is large enough to encompass the whole cluster, i.e. $\theta_f \geq \sqrt{\theta_{\rm vir}^2 + \theta_{\rm beam}^2}$, corresponds to simple model expectation[11].

We note also that the small-scale noise in LOS velocity $\epsilon_{S,i}^{{\rm LOS},s}$ (see Section 6.2.3) scales with cluster optical depth τ_i. The other noise term $\epsilon_{0,i}^{\theta_f}$ denotes the residual of primary CMB anisotropies plus inhomogeneous instrumental noise, which we assume to be a Gaussian random noise with zero mean and variance $\left(\sigma_{0,i}^{\theta_f}\right)^2$. Eq. (6.16) holds so long as the tSZ plus other foreground contaminations cancel out due to them being uncorrelated with the LOS large-scale velocity $v_{L,i}^{\rm LOS}$. Given the cluster mass cut introduced in Section 6.2.2, we confirmed that this condition holds (see Appendix F).

Both the signal amplitude and noise are functions of the AP filter size θ_f (or AAP filter scale φ_f) – as indicated by the superscript θ_f; however, for the sake of readability, we will omit the superscript θ_f in all following equations.

[11] Within the simple model considered here, $\alpha = 1$ also indicates that the baryonic fraction inside the clusters follows the cosmic fraction.

The corresponding likelihood distribution for a single BORG-SDSS3 s-th sample is given by

$$\mathcal{P}\left(\{\Delta T_{\mathrm{kSZ},i}/T_0\}|\alpha, \{\tau_i v_{L,i}^{\mathrm{LOS},s}/c\}\right) = \prod_i \frac{1}{\sqrt{2\,\pi\,\sigma_i^2}} \exp\left\{-\frac{1}{2}\frac{\left(\Delta T_{\mathrm{kSZ},i}/T_0 + \alpha\,\tau_i v_{L,i}^{\mathrm{LOS},s}/c\right)^2}{\sigma_i^2}\right\} \quad (6.17)$$

where

$$\sigma_i^2 \equiv \sigma_{0,i}^2 + (\tau_i/c)^2 \sigma^2_{\epsilon_S^{\mathrm{LOS}}}. \quad (6.18)$$

We now seek to construct a posterior distribution for α, marginalized over N BORG-SDSS3 samples

$$\begin{aligned}\mathcal{P}\left(\alpha|\{\Delta T_{\mathrm{kSZ},i}/T_0\}\right) &= \int \mathrm{d}x \mathcal{P}\left(\alpha, x|\{\Delta T_{\mathrm{kSZ},i}/T_0\},\right) \\ &= \int \mathrm{d}x \mathcal{P}\left(\alpha, x\right) \frac{\mathcal{P}\left(\{\Delta T_{\mathrm{kSZ},i}/T_0\}|\alpha,\, x\right)}{\mathcal{P}\left(\{\Delta T_{\mathrm{kSZ},i}/T_0\}\right)} \\ &\propto \mathcal{P}\left(\alpha\right) \int \mathrm{d}x \mathcal{P}(x) \mathcal{P}\left(\{\Delta T_{\mathrm{kSZ},i}/T_0\}|\alpha,\, x\right)\,, \end{aligned} \quad (6.19)$$

where we have used $x = \tau_i\left(v_{\mathrm{LOS},i}/c\right)$ and introduced the prior on α explicitly as $\mathcal{P}(\alpha)$. The BORG algorithm provides a sampled approximation

$$\mathcal{P}\left(x\right) \approx \frac{1}{N}\sum_{s=1}^{N} \delta^D(x - x^s)\,, \quad (6.20)$$

where $\delta^D(x)$ denotes the Dirac delta distribution and x^s denote the respective samples provided by BORG. Then

$$\begin{aligned}\mathcal{P}\left(\alpha|\{\Delta T_{\mathrm{kSZ},i}/T_0\}\right) &\propto \mathcal{P}\left(\alpha\right) \frac{1}{N}\sum_{s=1}^{N} \mathcal{P}\left(\{\Delta T_{\mathrm{kSZ},i}/T_0\}|\alpha,\, x^s\right)\,, \\ &\propto \mathcal{P}\left(\alpha\right) \frac{1}{N}\sum_{s=1}^{N} \prod_i \frac{1}{\sqrt{2\,\pi\,\sigma_i^2}} \exp\left\{-\frac{1}{2}\frac{\left(\Delta T_{\mathrm{kSZ},i}/T_0 + \alpha\, x^s\right)^2}{\sigma_i^2}\right\}\,, \end{aligned} \quad (6.21)$$

which is a Gaussian mixture distribution. Each mixture component consists of an individual BORG-SDSS3 sample s, associated with a mixture weight λ_s given by (see Appendix G for a detailed derivation):

$$\lambda_s = \frac{\mathrm{e}^{\omega_s + \frac{1}{2}\ln(2\pi\,(\sigma_s)^2)}}{\sum_s^N \mathrm{e}^{\omega_s + \frac{1}{2}\ln(2\pi\,(\sigma_s)^2)}}, \quad (6.22)$$

in which

$$\sigma_s^2 = \left[\sum_i \left(\frac{-\tau_i v_{L,i}^{\mathrm{LOS},s}/c}{\sigma_i}\right)^2\right]^{-1}, \tag{6.23}$$

and

$$\omega_s \equiv \mu_s^2/\left(2\sigma_s^2\right), \text{ with } \mu_s = \frac{\sum_i \left[\frac{\left(\Delta T_{\mathrm{kSZ},i}/T_0\right)\left(-\tau_i v_{L,i}^{\mathrm{LOS},s}/c\right)\right)}{\sigma_i^2}\right]}{\sum_i \left(\frac{-\tau_i v_{L,i}^{\mathrm{LOS},s}/c}{\sigma_i}\right)^2}. \tag{6.24}$$

Note, that λ_s defines a probability over BORG-SDSS3 samples which gives preference to better-fitting realizations, and $\sum_s^N \lambda_s = 1$.

For simplicity, in what follows, we assume a uniform prior on α such that $\mathcal{P}(\alpha) = 1$. Then, the ensemble mean estimate of α is then given by

$$\langle\alpha\rangle_s = \sum_s^N \lambda_s \mu_s = \sum_s^N \lambda_s \frac{\sum_i \frac{\left(\Delta T_{\mathrm{kSZ},i}/T_0\right)\left(-\tau_i v_{L,i}^{\mathrm{LOS},s}/c\right)}{\sigma_i^2}}{\sum_i \left(\frac{-\tau_i v_{L,i}^{\mathrm{LOS},s}/c}{\sigma_i}\right)^2}, \tag{6.25}$$

while its variance is given by (see Appendix G)

$$\sigma_\alpha^2 = \sum_s^N \lambda_s \sigma_s^2 + \sum_s^N \lambda_s \left(\mu_s - \langle\alpha\rangle_s\right)^2. \tag{6.26}$$

By up-weighting samples that fit the data better, our estimator in Eq. (6.25) minimizes the de-correlation of reconstructed velocity fields in our ensemble – caused by uncertainties in galaxy observation. It is worth noting that Eq. (6.25) is identical to Eq. (9) in [174] if one takes all $\lambda_s = 1$ and neglects the uncertainty in the velocity reconstruction. As can be seen in Eq. (6.26), however, our uncertainty properly includes uncertainties in the reconstructed velocity field.

We further test our estimator in Eq. (6.25) on mock input data where a kSZ signal template – generated by DM halos in the GADGET-2 simulation of sample $s = 9000$ (see Appendix I), assuming a Gaussian gas profile – is injected into a `SMICA`2018-like map (including both CMB and instrumental noise). By artificially varying the noise level, we verify that our estimator is indeed *unbiased*. In the limit of vanishing noise, $\lambda_{s=9000} \to 1$, correctly singles out the sample used to generate the mock signal.

However, the exact value of α is sensitive to, in addition to f_{free} or f_{gas}, systematics like bias in the amplitude of reconstructed velocity field and the weak lensing mass calibration in the $M_{200}-N_{200}$ relation, hence an interpretation of α as f_b would require careful modeling of these systematics, which is not the main focus of this chapter.

6.3.2 kSZ likelihood: multi-scale signal

For simplicity, in this case, let us restrict ourselves to the AP filter whose θ_f does not depend on the cluster effective apparent size. To combine measurements at different filter sizes θ_f then, it is necessary to modify Eqs. (6.16)–(6.21) to include the cluster gas profile $f(\theta_f)$ as

$$\tau_i = \tau_{i,0}\, f^{\theta_f}, \tag{6.27}$$

where we have assumed that this profile is universal. In this work, we estimate this profile by applying our pipeline on a pure-kSZ signal template generated from all DM particles found in the GADGET-2 high-resolution simulation (see details in Appendix I). We show below the measurement of f^{θ_f} at locations of DM halos identified in said simulation using the Rockstar halo finder[12] [181, 182], an adaptive hierarchical friends-of-friends (FoF) algorithm in six-dimensional phase-space. Note that we use the same velocity field to assign LOS velocity to the DM particles and halos.

Eq. (6.16) now becomes

$$\overrightarrow{\Delta T}_{\mathrm{kSZ},i}/T_0 = -\alpha\,\tau_{i,0}\vec{f}\left(v_{L,i}^{\mathrm{LOS}}/c\right) - \tau_{i,0}\vec{f}\left(\epsilon_{S,i}^{\mathrm{LOS},s}/c\right) + \vec{\epsilon}_{0,i}\,, \tag{6.28}$$

where we have again omitted the superscript θ_f and instead expressed quantities that depend on θ_f as data vectors to stress the fact that all measurements at different filter scales are now combined in Eq. (6.28). Note also that α is now simply a scalar instead of a function of θ_f. Eq. (6.17) can then be rewritten as

$$\begin{aligned}\mathcal{P}\left(\{\overrightarrow{\Delta T}_{\mathrm{kSZ},i}/T_0\}|\alpha,\,\tau_{i,0}\vec{f}\left(v_{L,i}^{\mathrm{LOS},s}/c\right)\}\right) &\propto \prod_i |\mathcal{C}_i|^{-1/2} \\ \times \exp\left\{-\frac{1}{2}\left[\overrightarrow{\Delta T}_{\mathrm{kSZ},i}/T_0 + \alpha\,\tau_{i,0}\vec{f}\left(v_{L,i}^{\mathrm{LOS},s}/c\right)\right]^{\intercal}(\mathcal{C}_i)^{-1}\left[\overrightarrow{\Delta T}_{\mathrm{kSZ},i}/T_0 + \alpha\,\tau_{i,0}\vec{f}\left(v_{L,i}^{\mathrm{LOS},s}/c\right)\right]\right\}\end{aligned} \tag{6.29}$$

where the covariance matrix $\mathcal{C}_i$ for cluster i is given by

$$C_i^{\theta_f\theta'_f} = \left\langle \frac{\Delta T_{\mathrm{CMB}}^{\theta_f}(\boldsymbol{\theta_i})}{T_0}\frac{\Delta T_{\mathrm{CMB}}^{\theta'_f}(\boldsymbol{\theta_i})}{T_0}\right\rangle + \left\langle \frac{\Delta T_{\mathrm{instr}}^{\theta_f}(\boldsymbol{\theta_i})}{T_0}\frac{\Delta T_{\mathrm{instr}}^{\theta'_f}(\boldsymbol{\theta_i})}{T_0}\right\rangle + \tau_{i,0}^2 f^{\theta_f} f^{\theta'_f}\sigma^2_{\epsilon_S^{\mathrm{LOS}}}, \tag{6.30}$$

in which we have separated the primary CMB anisotropy and the instrumental noise into the first and second term, respectively. Although we do include the last term on the r.h.s. of Eq. (6.30) in our covariance estimate, we note that it is negligible compared to the first two terms – as it scales with τ^2 – and the exclusion of this term affects neither the signal amplitude nor S/N. The second term is highly inhomogeneous due to the scanning strategy of the Planck satellite and requires instrumental-specific mocks to estimate. To this end, it is worth mentioning that Planck does provide a very limited set of 300 noise and

[12] `https://bitbucket.org/gfcstanford/rockstar/`

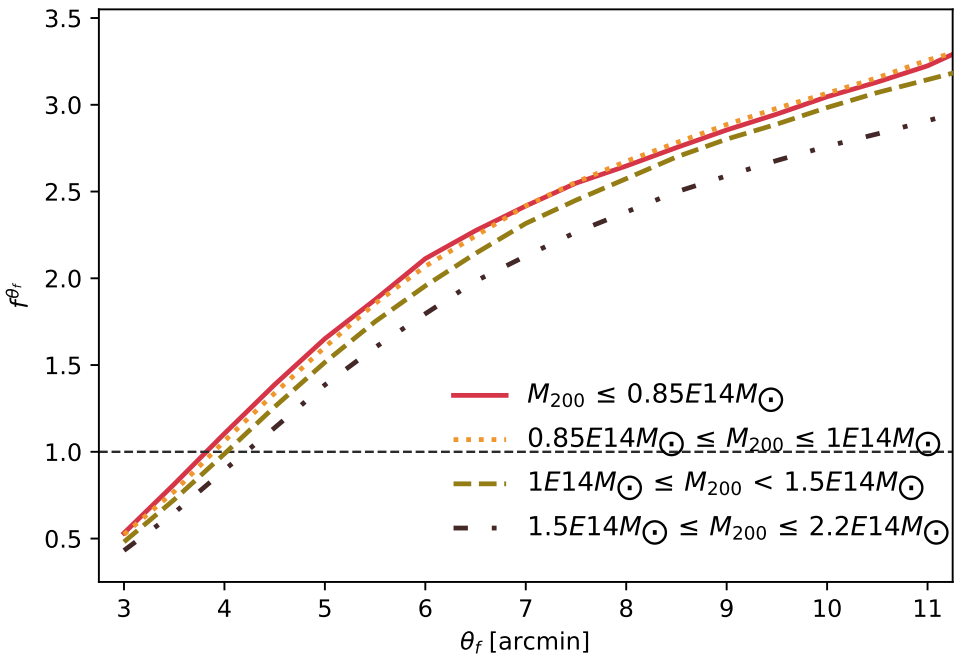

Figure 6.4: We measure the profile f^{θ_f} at locations of DM halos from our mock kSZ template generated by all DM particles in the GADGET-2 simulation detailed in Appendix I. As a consistency check, we divide our DM halo sample of $M_{200} \leq 2.2E14M_\odot$ into 4 mass bins as shown in the plot. Below $M_{200} \leq 1E14M_\odot$, the profile is not very sensitive to M_{200}. Only the result for $M_{200} \leq 0.85E14M_\odot$ is used in our analysis.

residual systematics simulations for SMICA2018 [44]. For this work, we instead choose to generate 2500 instrumental noise maps in which the noise value of each pixel is drawn from a zero-mean Gaussian whose variance is given by the corresponding temperature intensity variance in the Planck 2018 HFI Sky Map (Full Mission) at frequency 143GHz [183, 184] (see details in Appendix H). While this estimate is likely conservative since SMICA is a weighted linear combination of multiple frequency channels, we expect it to be robust and stable, especially at small filter sizes where this term dominates. The first term could be estimated analytically using the Planck 2018 best-fit ΛCDM power spectrum [80]. In the *flat-sky* limit,

$$\begin{aligned}\Delta T^{\theta_f}_{\rm CMB}(\boldsymbol{\theta_i}) &= \int d\boldsymbol{\theta'} W^{\theta_f}(\boldsymbol{\theta'}-\boldsymbol{\theta_i})\Delta T^{\rm obs}_{\rm CMB}(\boldsymbol{\theta'}) \\ &= \int \frac{d\boldsymbol{\ell}}{(2\pi)^2}\, \exp\left(i\boldsymbol{\ell}\cdot\boldsymbol{\theta_i}\right)(\pi\theta_f^2)\, W(\ell\theta_f)\, \Delta T^{\rm obs}_{\rm CMB}(\boldsymbol{\ell}),\end{aligned} \tag{6.31}$$

where $W(l\theta_f)$ is the Fourier transform of the AP filter

$$W\left(\ell\theta_f\right) = 2\left[W_{\rm TH}\left(\ell\theta_f\right) - W_{\rm TH}\left(\sqrt{2}\ell\theta_f\right)\right], \qquad W_{\rm TH}\left(\ell\theta_f\right) = 2\frac{J_1\left(\ell\theta_f\right)}{\ell\theta_f}, \tag{6.32}$$

while $\boldsymbol{\ell}$ is the two-dimensional wavevector perpendicular to the LOS, $\ell = |\boldsymbol{\ell}|$, and

$$\Delta T^{\rm obs}_{\rm CMB}(\boldsymbol{\ell}) = \Delta T_{\rm CMB}(\boldsymbol{\ell}) B(\boldsymbol{\ell}). \tag{6.33}$$

The CMB covariance matrix in Eq. (6.30) is then explicitly given by (see Appendix H)

$$\mathcal{C}^{\theta_f\theta'_f}_{{\rm CMB},i} = \frac{\pi\theta_f^2(\theta'_f)^2}{2T_0^2}\int_0^\infty d\ell\, \ell\, W\left(\ell\theta_f\right)\, W\left(\ell\theta'_f\right)\, C_\ell^{\rm CMB}, \tag{6.34}$$

where $C_\ell^{\rm CMB}$ denotes the Planck 2018 ΛCDM best fit angular power spectrum. We show in Figure 6.5 the CMB correlation matrix evaluated from Eq. (6.34), the instrumental noise correlation matrix estimated, for one single cluster, from 2500 instrumental noise mocks, and the corresponding total correlation matrix of the AP measurements (including both CMB and instrumental noise contributions) at different filter sizes estimated by Eq. (6.30). Note that the last two vary between cluster locations due to the inhomogeneity of the Planck instrumental noise.

We use the covariance matrix estimated by Eq. (6.30) for both the individual-scale and multi-scale measurements of the signal. Specifically, for the individual-scale case which employs the AAP filter, we interpolate $\sigma^2_{{\rm CMB},i}$ (cf. Eq. (6.18)) from a 1024x1024 $\mathcal{C}^{\theta_f\theta'_f}_{\rm CMB}$ matrix.

The posterior of α given the measurements at all filter sizes can be constructed similarly

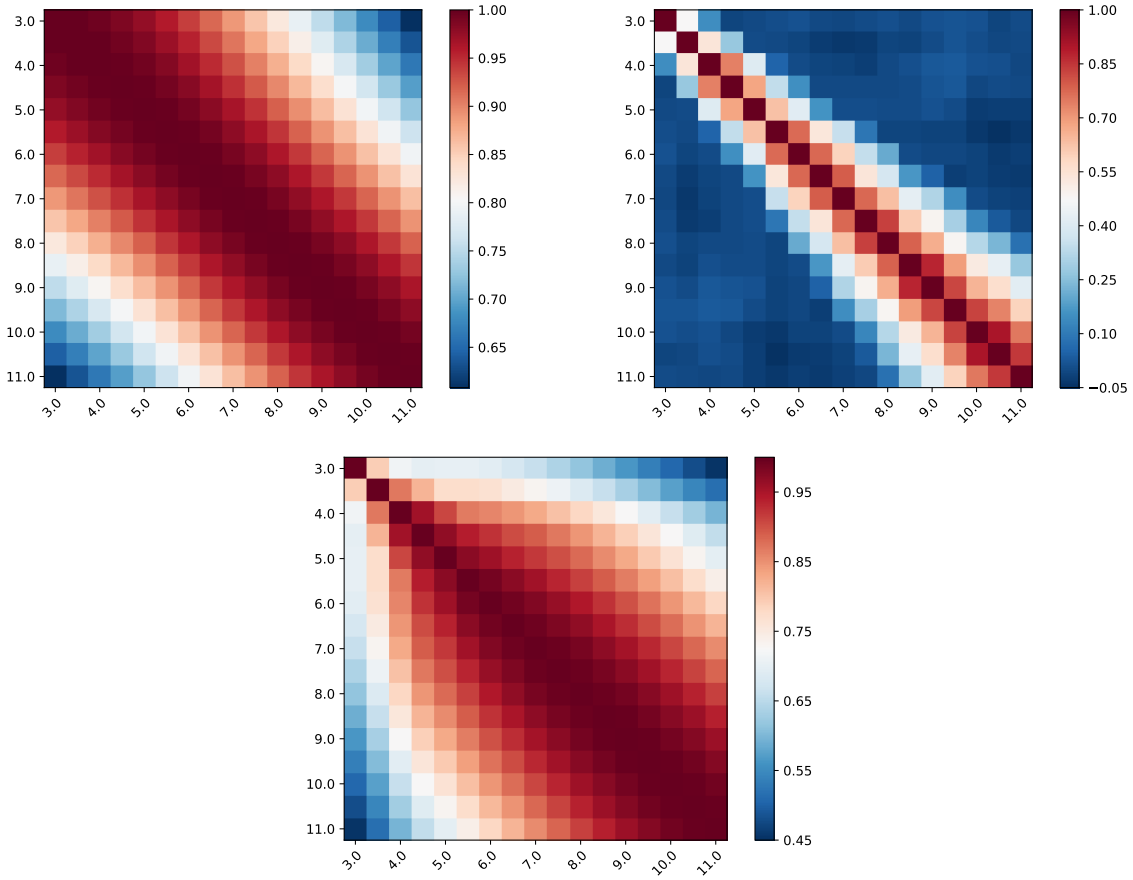

Figure 6.5: *Top:* Correlation matrices of the primary CMB (top-left panel) and instrumental noise (top-right panel), as described by the first and second term on the r.h.s. of Eq. (6.30). *Bottom:* The total noise correlation matrix as given by the Eq. (6.30).

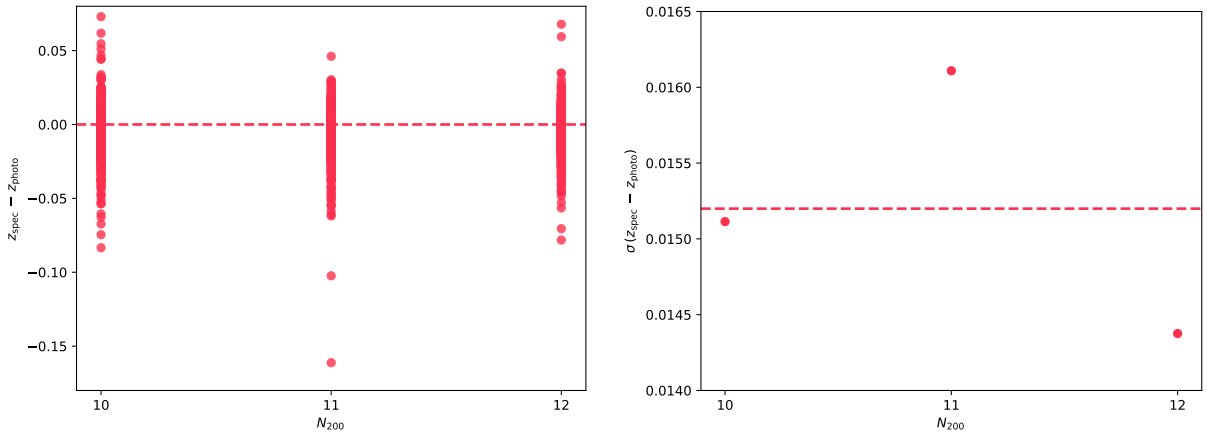

Figure 6.6: *Left panel:* The difference between BCG spectroscopic redshift and cluster photometric redshift for 2129 maxBCG clusters below $M_{200} = 0.85E14M_{\odot}$ (red dots). The dashed line presents the mean of the difference in each richness bin of $N_{200} = [10, 11, 12]$. *Right panel:* Variance of the difference between BCG spectroscopic redshift and cluster photometric for maxBCG clusters below $M_{200} = 0.85E14M_{\odot}$. The horizontal dashed line presents the scatter averaged over all three richness bins.

to Eq. (6.19)

$$\mathcal{P}\left(\alpha|\{\overrightarrow{\Delta T}_{\mathrm{kSZ},i}/T_0\}\right) \propto \mathcal{P}\left(\alpha\right)\frac{1}{N}\sum_{s=1}^{N}\prod_{i}|\vec{C}_i|^{-1/2} \exp\left\{-\frac{1}{2}\left[\overrightarrow{\Delta T}_{\mathrm{kSZ},i}/T_0 + \alpha\,\vec{x}^s\right]^{\intercal}\left(\vec{C}_i\right)^{-1}\left[\overrightarrow{\Delta T}_{\mathrm{kSZ},i}/T_0 + \alpha\,\vec{x}^s\right]\right\} \tag{6.35}$$

where we have similarly used $\vec{x}^s \equiv \left\{\tau_{i,0}\vec{f}\left(v_{L,i}^{\mathrm{LOS},s}/c\right)\right\}$.

The expressions of $\langle\alpha\rangle_s$ and σ_α^2 are similar to those in Eq. (6.25) and Eq. (6.26) with modifications to μ_s and ω_s as described in Appendix G.

6.3.3 Modeling photo-z uncertainty

To account for uncertainties in $v_{L,i}^{\mathrm{LOS},s}$ induced by photometric redshift error (see left panel of Figure 6.6), we introduce an additional sampling step. Specifically, we generate a sample of N_r realizations of maxBCG cluster positions in redshift-space, in which we keep fixed the redshifts of z_{spec} clusters while sampling those of z_{photo} clusters as

$$z_{\mathrm{photo},i}^{r} = z_{\mathrm{photo},i}^{0} + \delta z_i^r\,, \tag{6.36}$$

wherein $z_{\mathrm{photo},i}^{0}$ is the *fiducial* photometric redshift of cluster i and

$$\delta z_i^r \sim \mathcal{N}\left(0, \sigma_z\right) \tag{6.37}$$

with σ_z is the mean of the scatter $\sigma_z(N_{200}) = \sigma[z_{\rm spec}(N_{200}) - z_{\rm photo}(N_{200})]$ averaged over three $N_{200} = [10, 11, 12]$ bins (see right panel of Figure 6.6). When drawing from the Gaussian distribution in Eq. (6.37), we limit the range of $z^r_{{\rm photo},i}$ in our realizations within $[0.05, 0.5]$.

We then introduce an additional sum over all N_r realizations in Eq. (6.21)

$$\mathcal{P}\left(\alpha|\{\Delta T_{{\rm kSZ},i}/T_0\}\right) = \mathcal{P}\left(\alpha\right)\frac{1}{N_r}\sum_{r=1}^{N_r}\frac{1}{N_s}\sum_{s=1}^{N_s}\prod_i \frac{1}{\sqrt{2\,\pi\,\sigma_i^2}}\exp\left\{-\frac{1}{2}\frac{\left(\Delta T_{{\rm kSZ},i}/T_0 + \alpha\,\tau_i v_{L,i}^{{\rm LOS},sr}/c\right)^2}{\sigma_i^2}\right\}. \tag{6.38}$$

Note that, in the case of the AAP filter, for all redshift samples, we keep the filter size in Eq. (6.38) fixed as $\theta_f = \varphi_f\,\theta^0_{200,i}$ where $\theta^0_{200,i}$ is computed from the fiducial values $z^0_{{\rm photo},i}$. Given the range in which the apparent size of selected maxBCG clusters varies (see right panel of Figure 6.2), this approximation does not, by any mean, affect our estimator significantly.

Since the two indices s and r in Eq. (6.38) are mathematically equivalent – in the sense that they only appear in $v_{L,i}^{{\rm LOS},sr}$ – we can rewrite them as

$$n \equiv \{s,r\} \tag{6.39}$$

and

$$v_{L,i}^{{\rm LOS},\ n} \equiv v_{L,i}^{{\rm LOS},s}\left(z^r_{{\rm photo},i}\right) \tag{6.40}$$

so that

$$\mathcal{P}\left(\alpha|\{\Delta T_{{\rm kSZ},i}/T_0\}\right) = \mathcal{P}\left(\alpha\right)\frac{1}{N}\sum_{n=1}^{N}\prod_i \frac{1}{\sqrt{2\,\pi\,\sigma_i^2}}\exp\left\{-\frac{1}{2}\frac{\left(\Delta T_{{\rm kSZ},i}/T_0 + \alpha\,\tau_i v_{L,i}^{{\rm LOS},n}/c\right)^2}{\sigma_i^2}\right\}, \tag{6.41}$$

where $N = N_r\,N_s$.

The derivation of the estimator $\langle\alpha\rangle_n$ and its uncertainty σ^2_α is then again similar to that in Section 6.3 and Appendix G, with s replaced by n.

6.4 Results

6.4.1 Individual-scale signal measurements

We show in Figure 6.7 our measurements of individual-scale large-scale bulk-flow amplitude $\langle\alpha^{\varphi_f}\rangle_n$ (cf. Eq. (6.25)), as a function of the AAP filter scale φ_f, using each (top panel) and both datasets (bottom panel) described in Section 6.2.2. We emphasize again that our uncertainty reported here includes also uncertainties in the reconstructed large-scale velocity field.

As can be noted from the top panel of Figure 6.7, we obtain consistent results between the two datasets $z_{\rm spec}$ and $z_{\rm photo}$, which can be combined as shown in the bottom panel

of Figure 6.7. Note that the slightly larger uncertainty region of $\langle\alpha\rangle_n$ measurement using $z_{\rm spec}$ set – compared to that of $\langle\alpha\rangle_s$ measurement using $z_{\rm photo}$ set – is mostly caused by their limited quantity, as suggested by the ratio between the two uncertainties being approximately constant across all filter scales. Both panels of Figure 6.7 show that most of the information come from small scales. As the filter scale increases, the AAP estimate picks up more and more contribution from primary CMB anisotropies, as well as other large-scale sources of contamination, and quickly loose its constraining power.

In addition, we provide the S/N ratio at each filter scale for each dataset in Table 6.1. Both $z_{\rm spec}$ and the combined set show peaks of S/N at $\varphi = 0.9$, as one should expect. Further, for $z_{\rm photo}$ set, the photo-z uncertainty shows a bigger impact at small filter sizes, which is also an expected behavior, since all large scales, (primary) CMB is still the dominant noise source.

For clarity, we additionally show the mixture weights (cf. Eq. (6.22)) in Figure 6.8, since it can provide some insights on how the information in $z_{\rm spec}$ and $z_{\rm photo}$ sets are combined. One can see that the distributions of λ_s in the three panels on the right of Figure 6.8, which show λ_s for the combined set, are consistent with those of λ_s in the leftmost and middle columns, which respectively show λ_s for $z_{\rm spec}$ and $z_{\rm photo}$ set. This implies that Eq. (6.41) consistently combines information from the $z_{\rm spec}$ and $z_{\rm photo}$ to simultaneously and correctly pick out the better-fitting BORG samples and redshift realizations.

φ_f	$z_{\rm spec}$		$z_{\rm photo}$		$z_{\rm spec} + z_{\rm photo}$	
	$\langle\alpha^{\varphi_f}\rangle_n$	S/N	$\langle\alpha^{\varphi_f}\rangle_n$	S/N	$\langle\alpha^{\varphi_f}\rangle_n$	S/N
0.7	0.58	1.26	-0.66	-1.82	-0.02	-0.05
0.8	0.98	1.54	0.16	0.22	0.72	1.67
0.9	1.62	**1.99**	0.63	0.82	1.09	**2.26**
1.0	0.63	0.62	0.94	0.87	0.90	1.16
1.1	1.05	0.81	0.27	0.19	0.79	0.82
1.2	0.80	0.48	1.00	0.54	1.07	0.81
1.3	0.31	0.15	1.64	0.84	0.99	0.68
1.4	-0.35	-0.13	1.56	0.64	0.67	0.37
1.5	-0.67	-0.21	3.44	**1.09**	1.31	0.53
1.6	-0.18	-0.05	2.67	0.77	1.44	0.54

Table 6.1: We show the ensemble mean of our MAP estimates of the signal amplitude α for different AAP filter scales. We compute the statistical significance as S/N=$\langle\alpha^{\varphi_f}\rangle_n/\sigma_\alpha$. The maximum S/N ratios for each case are highlighted.

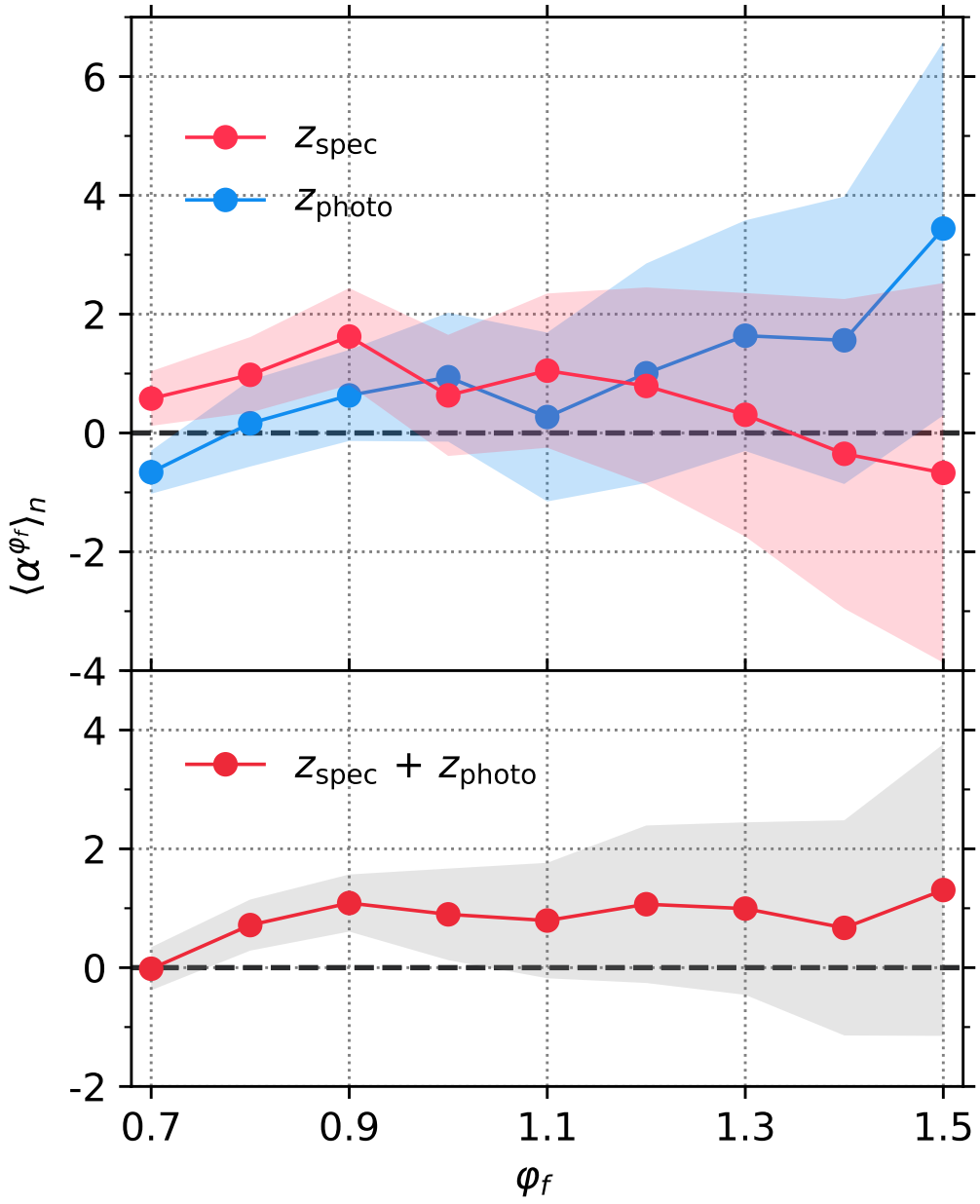

Figure 6.7: *Top panel:* The large-scale bulk-flow amplitude $\langle \alpha \rangle_n$ measured at different AAP filter scales $\varphi_f = [0.7, 1.6]$ using $z_{\rm spec}$ (red) and $z_{\rm photo}$ (blue) dataset. The shaded regions denote the corresponding $1\,\sigma$ uncertainties, including both uncertainties in CMB anisotropies and the reconstructed velocity field. *Bottom panel:* Same as top panel but combining both datasets.

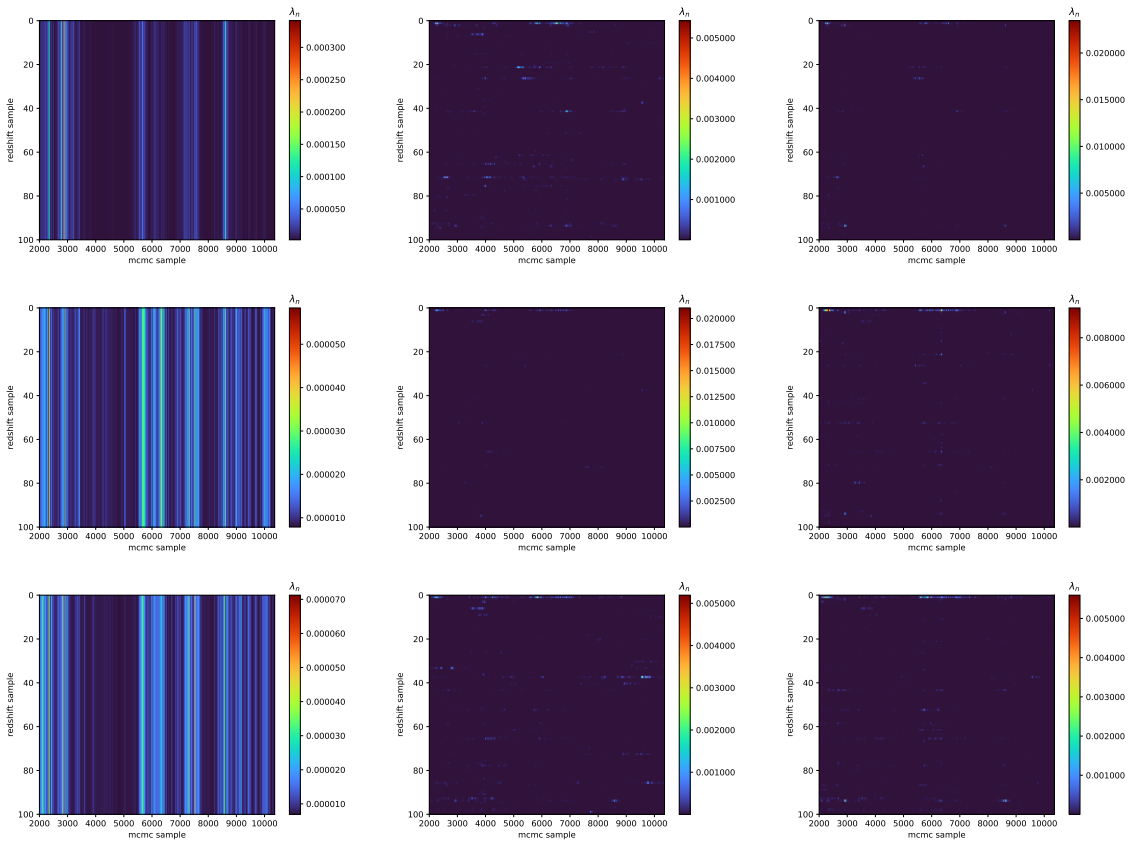

Figure 6.8: We show the distributions of λ_n over 837 BORG samples with identifier 2000-10360 (x-axes) and 100 redshift realizations (y-axes) for the cases of $z_{\rm spec}$ (left), $z_{\rm photo}$ (middle) and both datasets combined (right) at three individual filter scales of $\varphi = \{0.9, 1.0, 1.1\}$ (top to bottom). Note that we do no sample the redshifts of clusters in $z_{\rm spec}$ set, which explains the λ_n being evenly distributed among redshift realizations (y-axes) in the leftmost column.

6.4.2 Multi-scale signal measurement

We show in Figure 6.9 our measurements of multi-scale large-scale bulk-flow amplitude $\left\langle \alpha^{\theta_f} \right\rangle_s$ (cf. Eq. (6.25)), as a cumulative function of the AP filter radius θ_f, using each (top panel) and both datasets (bottom panel) described in Section 6.2.2.

As can be seen from both cases of individual- and multi-scale signal, the uncertainty in photometric redshift, when accounted for by double sampling, hinders the constraining power of the $z_{\rm photo}$ sample, i.e. adding the clusters from the $z_{\rm photo}$ set does not significantly improve our S/N. Although this problem might be alleviated by increasing the number of redshift realizations, further investigation is required to identify the optimal way to combine information from both spectroscopic and photometric data, given current constraints on computational resources. We defer such an investigation to future work.

The S/N ratio of the cumulative multi-scale signal for each dataset is summarized in Table 6.2. What is consistent between Table 6.1 and Table 6.2 is that, most of the information is limited at filter sizes below and about the apparent size of maxBCG clusters in our samples. Above that scale, CMB noise severely limits our S/N. This suggests that data from current CMB experiments with higher resolutions, $\sim$ 1 arcmin [27, 26], can improve our S/N significantly. On another hand, even at the modest resolution of $\sim 2-3$ arcmin, we expect CMB-S4, with much lower instrumental noise, to also have a significant impact on this measurement. In both cases, the details of the gas profile would become important and one would almost certainly need to go beyond the simple Gaussian profile assumed here.

θ_f [arcmin]	$z_{\rm spec}$		$z_{\rm photo}$		$z_{\rm spec} + z_{\rm photo}$	
	$\langle \alpha^{\varphi_f} \rangle_n$	S/N	$\langle \alpha^{\varphi_f} \rangle_n$	S/N	$\langle \alpha^{\varphi_f} \rangle_n$	S/N
3.0	1.62	1.50	1.78	**1.58**	1.69	**2.32**
3.5	1.48	1.68	-0.14	-0.14	0.77	1.23
4.0	0.93	1.07	-0.27	-0.29	0.42	0.65
4.5	0.89	1.01	-0.32	-0.35	0.35	0.54
5.0	0.93	1.06	-0.55	-0.59	0.29	0.42
5.5	1.53	**1.91**	-0.60	-0.69	0.64	1.03
6.0	1.28	1.58	-0.72	-0.77	0.48	0.76
6.5	0.80	0.91	-1.08	-1.51	0.05	0.06
7.0	0.74	0.99	-0.97	-1.44	0.01	0.02

Table 6.2: We show the ensemble mean of our MAP estimates of the cumulative signal amplitude α for increasing AP filter sizes. We compute the statistical significance as S/N=$\langle \alpha^{\varphi_f} \rangle_n / \sigma_\alpha$. The maximum S/N ratios for each case are highlighted. Note that, for readability we only list here results for $\theta_f = [3, 7]$ arcmin.

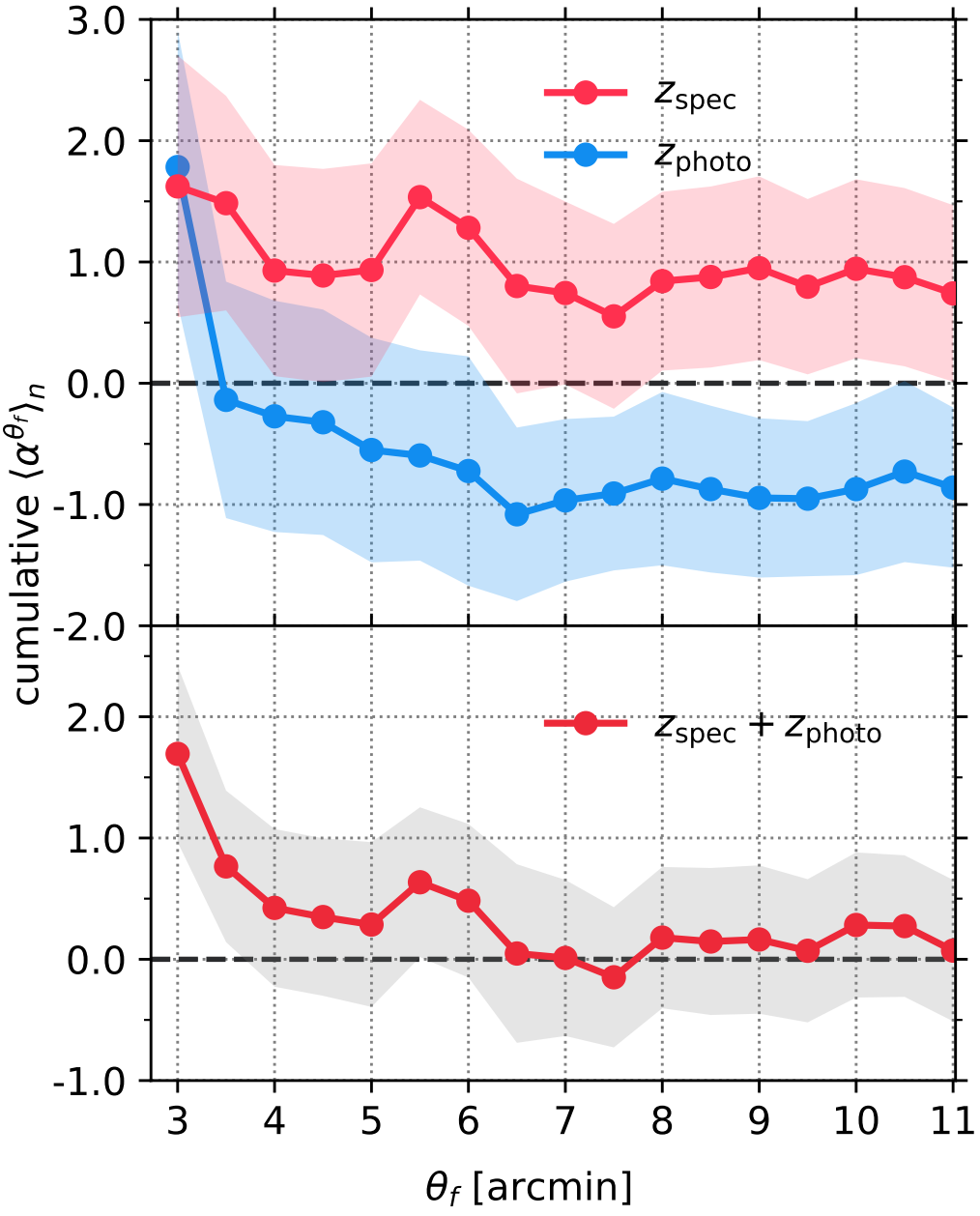

Figure 6.9: *Top panel:* The cumulative large-scale bulk-flow amplitude $\langle \alpha \rangle_n$ combining measurements at progressively larger AP filter sizes, $\theta_f = [3.0, 11.0]$ arcmin, using $z_{\rm spec}$ (red) and $z_{\rm photo}$ (blue) dataset. Going from left to right, each data point combines information from all previous points. The shaded regions denote the corresponding $1\,\sigma$ uncertainties, including both uncertainties in CMB anisotropies and the reconstructed velocity field. *Bottom panel:* Same as top panel but combining both datasets.

6.5 Null tests for systematics

In this section, we assert the significance of our measurement by performing two null tests on mock data in which we

1. shuffle the momentum of the clusters in our analysis, and
2. replace the SMICA2018 by a set of 300 SMICA-like mock maps taken from Planck.

Since the $z_{\rm photo}$ sample adds very little information, in the tests below, we will only include clusters from the $z_{\rm spec}$ sample.

6.5.1 Null tests: Individual-scale signal

For the case of individual-scale measurements using the AAP filter, we measure a large-scale bulk-flow amplitude consistent with zero for $z_{\rm spec}$ sample with cluster positions being shuffled, as can be seen in Figure 6.10. When applying our pipeline on the set of 300 SMICA2018 simulations (including CMB and instrumental noise) provided by Planck [44], we also recover S/N ratios consistent with zero-mean Gaussian distributions. We show in Figure 6.11 the histograms of S/N for increasing individual filter scale from $\varphi = 0.7$ to $\varphi = 1.5$. Note the histogram of $\varphi = 0.9$ (top right panel) shows that our reported S/N$\simeq 2$ for $z_{\rm spec}$ (cf. Table 6.1) is consistent with the 95% confidence interval.

6.5.2 Null tests: Multi-scale signal

For the case of cumulative multi-scale measurements using the AP filter, we also measure a cumulative large-scale bulk-flow amplitude consistent with zero for $z_{\rm spec}$ sample with cluster positions being shuffled, as shown in Figure 6.12. Our null test using the Planck simulations also recovers S/N ratios consistent with zero-mean Gaussian distributions. We show in Figure 6.13 the histograms of S/N for cumulative multi-scale from $\theta_f = 3.0 - 3.5$ arcmin to $\theta_f = 3.0 - 7.5$ arcmin. Note the histogram of $\theta_f = 3.0 - 5.5$ arcmin (middle row, center panel) shows that our reported S/N$\simeq 1.9$ for $z_{\rm spec}$ (cf. Table 6.2) is again consistent with the 95% confidence interval.

6.6 Discussion and conclusion

As its constraining power will improve with upcoming high-quality CMB and galaxy cluster data, the kSZ effect has the potential to become a key astrophysical and cosmological probe. It is thus important now to carefully model systematic effects that could potentially bias our kSZ measurements. So far, one of the often neglected systematics in kSZ measurements is that of the reconstructed (LOS) velocity. Using a *systematic-free* ensemble of inferred peculiar velocity field within the SDSS3-BOSS volume, we have presented measurements of the kSZ and large-scale bulk flow signal for a subsample of maxBCG clusters. Our

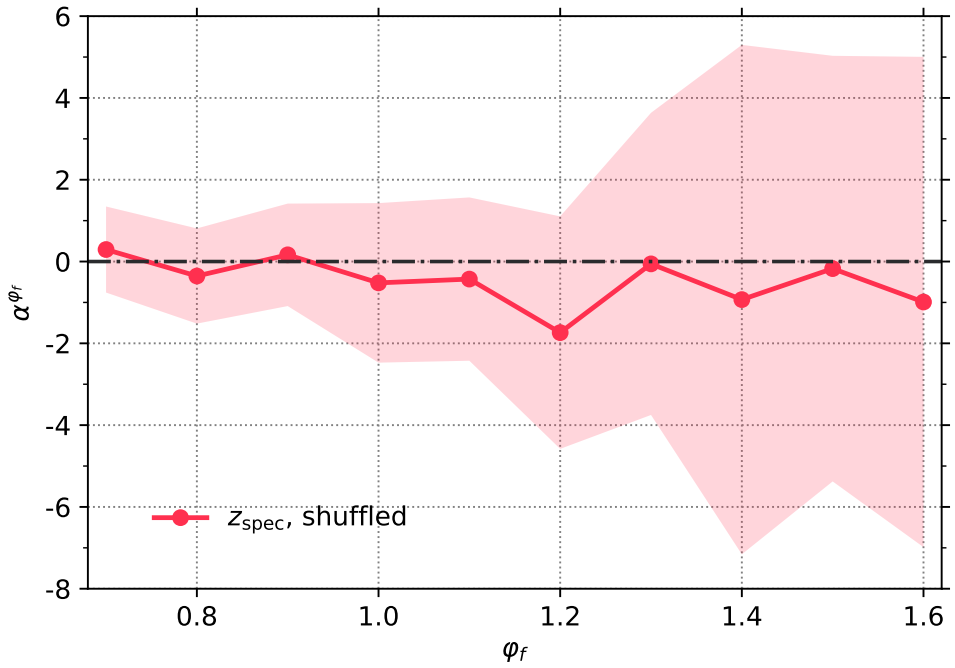

Figure 6.10: Individual-scale signal amplitude measured using $z_{\rm spec}$ sample but with sky positions of the clusters shuffled, plotted as a function of AAP filter scale.

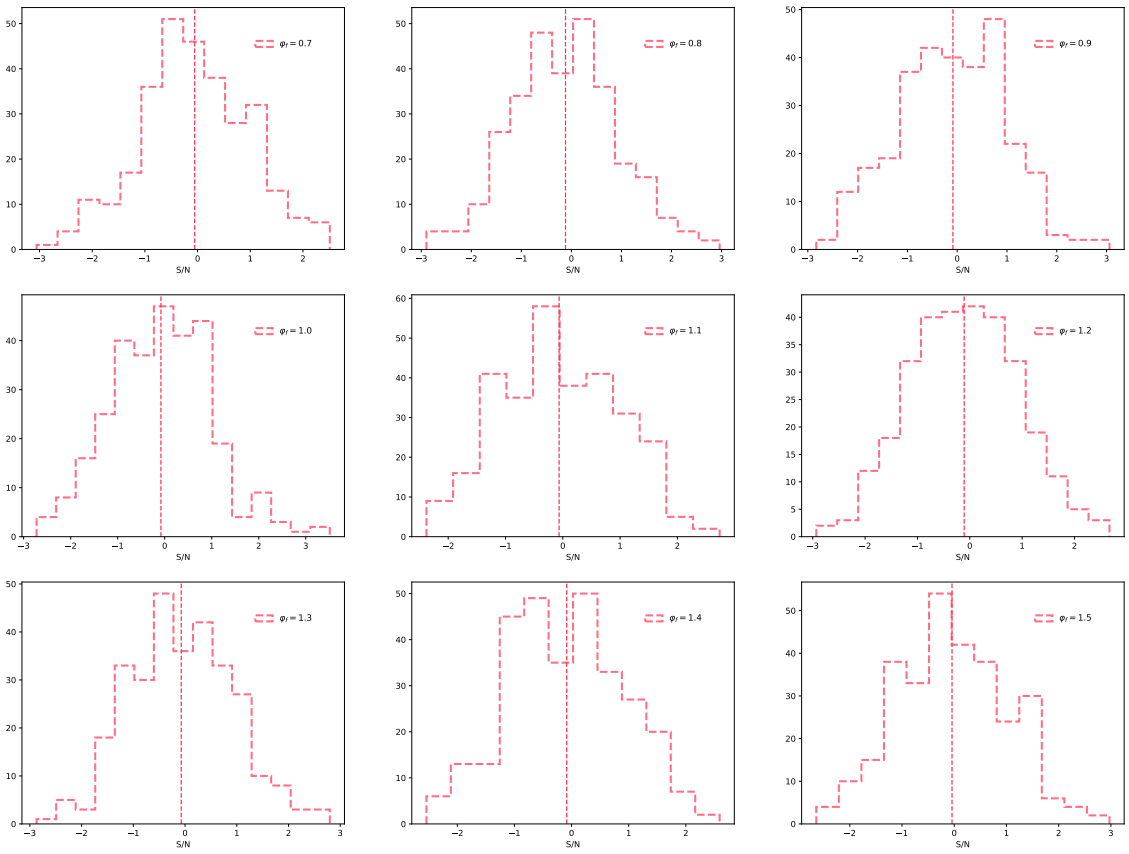

Figure 6.11: Histograms of the S/N of individual-scale signal for $z_{\rm spec}$ sample, measured at $\varphi_f = [0.7, 1.5]$, from 300 simulations of `SMICA2018` (including CMB and `SMICA`-like instrumental noise). The vertical dashed line presents the mean in each histogram.

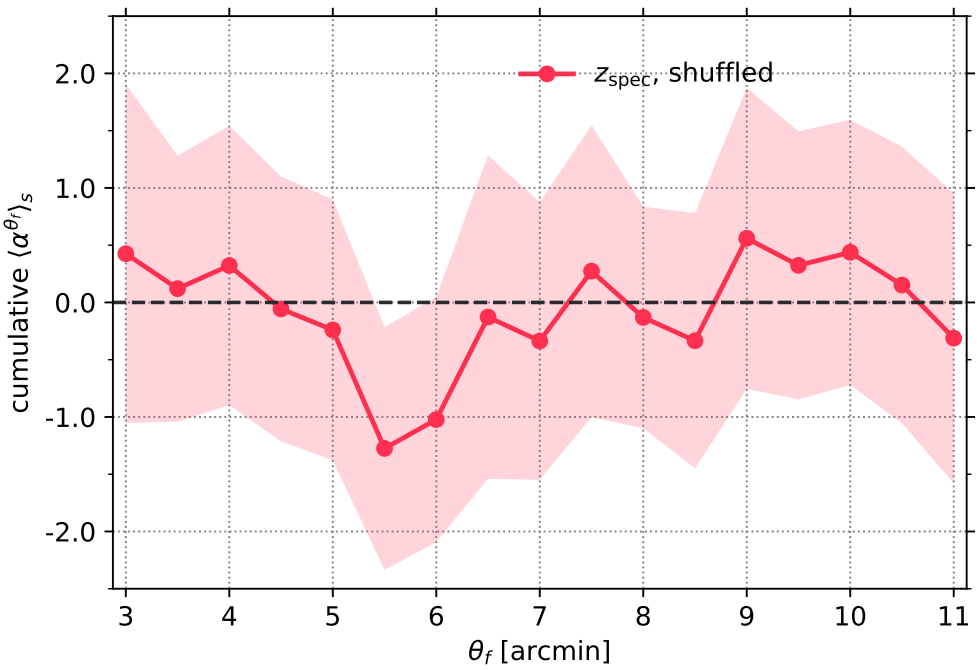

Figure 6.12: Multi-scale signal amplitude measured using $z_{\rm spec}$ sample but with sky positions of the clusters shuffled, plotted as a cumulative function of AP filter sizes.

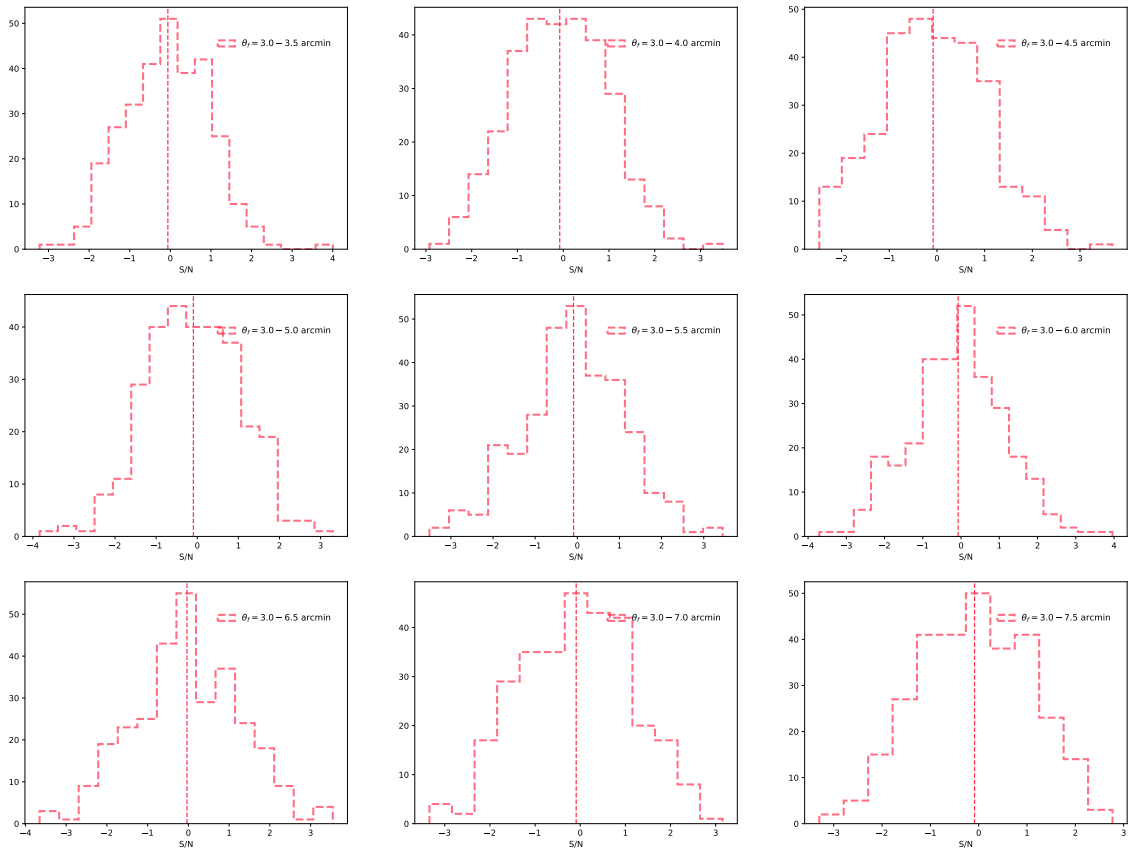

Figure 6.13: Histograms of the S/N of multi-scale signal for $z_{\rm spec}$ sample, measured cumulatively between $\theta_f = 3.0$ and $\theta_f = 7.5$ arcmin, from 300 simulations of `SMICA`2018 (including CMB and `SMICA`-like instrumental noise). The vertical dashed line presents the mean in each histogram.

Bayesian forward inference approach allows us to, for the first time, include uncertainties in the velocity reconstruction into the final uncertainty in the signal. As such, the significance of our kSZ measurement can be thought of as a rigorously conservative estimate. At the moment, this estimate seems to be mostly limited by the resolution and instrumental noise of the CMB experiment, since most information comes from the smaller scales. Although we have only observed a modest evidence of the kSZ signal at $\simeq 2\sigma$, we expect our results to be significantly improved with high-quality CMB data, either from future CMB experiments, for example, CMB-S4 [29], or from combined existing datasets such as SPT-Planck [185].

In addition, we have developed a robust kSZ likelihood that would be straightforward to optimize for specific studies, for example, by introducing specific cluster gas profiles and/or more sophisticated filtering techniques. We leave the study of cluster gas profile and these possibilities for future work. Another related advantage of this approach is that a prior on α can be easily introduced, for example, in cases where the cluster gas profile is known from complimentary measurements, e.g. X-ray, tSZ, etc.

For whatever methods of extracting the kSZ signal from the CMB map, the anisotropy estimates at positions of clusters will always be dominated by primary CMB anisotropies, tSZ, and instrumental noise. Consequently, a key factor for any kSZ measurement is the number of clusters with high-precision measurements of redshift. On one hand, this could be improved with either future high-resolution spectroscopic cluster catalogs or high-quality tSZ-cleaned CMB maps (see, e.g. [186]) which would allow for the inclusion of more massive clusters in the analysis. In this work, we have explored the possibility of including photometric clusters in the maxBCG catalog by sampling these clusters' redshift. While our approach arguably yields consistent, unbiased results from the considered $z_{\rm photo}$ sample from maxBCG catalog, a more detailed investigation would be needed to optimally combine information from the photometric dataset and that from spectroscopic dataset, given the improved precision of future data. On another hand, one could also think of simultaneously measuring both kSZ and tSZ signals so that there is no need for the removal of massive clusters. Nothing – in principle – prevents this simple extension of the data model in Eq. (6.16). In fact, this is also our plan for a follow-up study.

Chapter 7

Inference of cosmological parameters

This chapter presents our approach to the problem of cosmological parameter inference using Bayesian forward modeling of galaxy clustering at the field-level. As pointed out in Section 4.1, the key to this approach is a conditional probability or likelihood that is capable of capturing the stochasticity or scatter of local tracer density induced by small-scale modes integrated out in the EFT description of tracer bias (see Section 3.1). A somewhat similar attempt to ours, although focusing on the Alcock-Paczyński (AP) test and a different set of cosmological parameters $(\Omega_{\rm m}, w_0)$, was demonstrated on a SDSS3-like, model-consistent mock galaxy catalog in [127]. There, the mock galaxy catalog was generated by populating galaxies on top of an evolved matter field following the same bias model and likelihood used in the analysis. As shown in Chapter 5 (cf. Figure 5.4), for such self-consistency test, the main limiting factor is the tracer shot noise. As we will see below, our numerical tests in Section 7.5 and Section 7.6 are highly more stringent, since our tracers are DM halos identified in N-body simulations and the phases are always *fixed* to the true phases of the simulation. The results in this chapter were published in [123, 136], of which I am a co-author. For the sake of clarity and readability, we will briefly review the key results of [123] before focusing on the numerical application presented in [136]. Further, we relocate some relevant discussions and results in the main body of [123], for example, relation between our approach and BAO reconstruction, to Appendix M–Appendix O. It is worth mentioning that a rigorous derivation of our EFT-Fourier likelihood was later published in [187], of which I do not contribute directly.

The rest of this chapter is organized as follows. In Section 7.1, we introduce the EFT-inspired likelihood for cosmological parameter inference – focusing on the case of the clustering amplitude σ_8 – formulated in Fourier-space. In Section 7.3, we further analytically marginalizing over a subset of bias parameters to obtain a marginalized version of the likelihood in which the rest of bias parameters can be profiled out. The final result of this process is a profile likelihood of σ_8 whose performance is asserted at field-level in Section 7.5 and Section 7.6 using the suite of simulations described in Section 7.4. Finally, we summary our results and discuss future extensions of our approach in Section 7.7. Appendix J–Appendix O present further details, notably a derivation of the Poisson expectation for stochasticity.

Throughout this chapter, we again consider the general bias expansion in Eq. (3.1) up to second-order. This, for clarification, means that we include the following default set of operators in our bias expansion:

$$O \in \left\{\delta,\ \delta^2 - \left\langle\delta^2\right\rangle,\ (K^2_{ij}) - \left\langle(K_{ij})^2\right\rangle,\ \nabla^2\delta\right\},$$
$$\text{with coefficients} \quad \left\{b_1,\ \frac{b_2}{2},\ b_{K^2},\ c_{\nabla^2\delta}\right\}, \tag{7.1}$$

where $b_N \equiv N! b_{\delta^N}$ denote the bias coefficient corresponding to the N-th power of the matter density field (local-in-matter-density, LIMD). We denote the higher-derivative bias coefficient as $c_{\nabla^2\delta}$ rather than $b_{\nabla^2\delta}$, as it is an effective coefficient which also absorbs other contributions which depend on the chosen smoothing scale and cutoff. More generally, the bias parameters b_O correspond to well-defined physical bias parameters, while parameters denoted as c_O (so far, only $c_{\nabla^2\delta}$) are understood as effective coefficients which also absorb higher-order contributions. Eq. (7.1) corresponds to the complete set of operators up to second order at leading order in derivatives, and the leading higher-derivative operator ($\nabla^2\delta$). The significance of this choice will become clear later.

7.1 EFT approach to the conditional probability

The conditional probability Eq. (4.4) is written in real space, i.e. it relates the *empirically observed* $\vec{\delta}_h$ and *deterministically predicted* $\vec{\delta}_{h,\mathrm{det}}$ (cf. Eq. (3.32) and Eq. (5.4)) cell-by-cell in real space. Perturbative approaches, including the effective field theory of biased tracers, however naturally work in Fourier space. This is because the initial conditions are (approximately) a homogeneous Gaussian random field (see Section 2.1 and Section 4.1.1), so that their covariance is *diagonal* in Fourier space.

Let us thus instead consider the problem of deriving a conditional probability relating the halo density field in Fourier space, $\vec{\delta}_h \to \delta_h(\boldsymbol{k})$, to a linear combination $\delta_{h,\mathrm{det}}(\boldsymbol{k})$ of fields $O(\boldsymbol{k})$ constructed from the evolved matter field $\delta(\boldsymbol{k})$. Clearly, the field $\delta_{h,\mathrm{det}}(\boldsymbol{k})$ cannot be a perfect match to $\delta_h(\boldsymbol{k})$, but has noise. In the EFT approach, this noise formally arises because we are integrating out the small-scale modes of the density field, those above some maximum wavenumber k_{max}. An obvious question then is how the maximum wavenumber k_{max} should be chosen. A priori, the only guidance we have is that it should be less than the non-linear scale, $k_{\mathrm{NL}} \approx 0.3\,h\,\mathrm{Mpc}^{-1}$ at $z = 0$, but higher at higher redshifts (see Section 4.4 of [123] for a more precise prescription).

Importantly, the noise does not only affect δ_h, but in general the fields entering our predicted field $\delta_{h,\mathrm{det}}$ also have noise. We can thus write

$$\delta_h(\boldsymbol{k}) - \delta_{h,\mathrm{det}}(\boldsymbol{k}) = \varepsilon_h(\boldsymbol{k}) - \varepsilon_{\mathrm{model}}(\boldsymbol{k}) = \varepsilon_h(\boldsymbol{k}) - b_1\varepsilon_m(\boldsymbol{k})\,. \tag{7.2}$$

In the second equality, we have only kept the noise contribution from the matter density ε_m, which is multiplied by b_1 since this is how the matter density field enters $\delta_{h,\mathrm{det}}$. The

contributions from noise fields in the quadratic operators (specifically those that cannot be absorbed by ε_h) are higher order, as shown in Section 4.5 of [123], and can thus be dropped. Apart from this ranking, we have not used perturbative arguments so far. We will work in the continuum limit to keep the conceptual derivation clear, and move to a finite Fourier grid shortly.

Our goal is now to integrate out the noise fields ε_h and ε_m, since, by definition, we cannot predict them at the field level. For this, we assume that, on the scales $k < k_{\rm max}$ of interest, both fields can be approximated as Gaussian. On sufficiently large scales, this approximation is guaranteed to be accurate by the central limit theorem. We will return to this point in Appendix M. Further, we can use the fact that the power spectra of the noise fields have to be analytic in k on large scales. This is because they arise from interactions of modes that are of much smaller scale, and thus cannot involve the power spectrum on the scale k [97]. Finally, the noise in the matter density field has to satisfy $\lim_{k\to 0}\varepsilon_m(\boldsymbol{k})/k^2 = \text{const.}$, through mass and momentum conservation. We thus write, up to including k^4,

$$\begin{aligned}
\langle\varepsilon_h(\boldsymbol{k})\varepsilon_h(\boldsymbol{k})\rangle' &= P^{\varepsilon}_{hh}(k) = P^{\varepsilon,0}_{hh} + P^{\varepsilon,2}_{hh}k^2 + P^{\varepsilon,4}_{hh}k^4 \\
\langle\varepsilon_h(\boldsymbol{k})\varepsilon_m(\boldsymbol{k})\rangle' &= P^{\varepsilon}_{hm}(k) = P^{\varepsilon,2}_{hm}k^2 + P^{\varepsilon,4}_{hm}k^4 \\
\langle\varepsilon_m(\boldsymbol{k})\varepsilon_m(\boldsymbol{k})\rangle' &= P^{\varepsilon}_{mm}(k) = P^{\varepsilon,4}_{mm}k^4 \,.
\end{aligned} \tag{7.3}$$

The Cauchy-Schwartz inequality implies $|P^{\varepsilon,2}_{hm}| \leq \sqrt{P^{\varepsilon,0}_{hh}P^{\varepsilon,4}_{mm}}$. To summarize, the noise fields $\varepsilon_h, \varepsilon_m$ follow a multivariate normal distribution given by

$$\begin{aligned}
&\mathcal{P}\Big(\varepsilon_h(\boldsymbol{k}), \varepsilon_m(\boldsymbol{k})\Big) = |2\pi \mathrm{C}_\varepsilon|^{-1/2} \exp\left[-\frac{1}{2}(\varepsilon_h,\ \varepsilon_m)\mathrm{C}_\varepsilon^{-1}(\varepsilon_h,\ \varepsilon_m)^\dagger\right] , \\
&\text{where}\quad \mathrm{C}_\varepsilon = \mathrm{C}_\varepsilon(k^2) = \begin{pmatrix} P^{\varepsilon}_{hh}(k) & P^{\varepsilon}_{hm}(k) \\ P^{\varepsilon}_{hm}(k) & P^{\varepsilon}_{mm}(k) \end{pmatrix} ,
\end{aligned} \tag{7.4}$$

and $\dagger$ denotes the transpose and complex conjugate. We can now integrate out the stochastic fields in Eq. (7.2) to *derive* a likelihood for $\delta_h(\boldsymbol{k})$ given the known analytic scaling of the noise-field correlators on large scales:

$$\begin{aligned}
\mathcal{P}\Big(\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}(\boldsymbol{k})\Big) &= \int d\varepsilon_m(\boldsymbol{k})\, \mathcal{P}\Big(\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}(\boldsymbol{k}) - b_1\varepsilon_m(\boldsymbol{k}),\ \varepsilon_m(\boldsymbol{k})\Big) \\
&= (2\pi)^{-1/2}\left|P^{\varepsilon}_{hh}(k) + 2b_1P^{\varepsilon}_{hm}(k) + b_1^2P^{\varepsilon}_{mm}\right|^{-1/2} \\
&\quad\times \exp\left[-\frac{1}{2}\frac{|\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}(\boldsymbol{k})|^2}{P^{\varepsilon}_{hh}(k) + 2b_1P^{\varepsilon}_{hm}(k) + b_1^2P^{\varepsilon}_{mm}}\right] .
\end{aligned} \tag{7.5}$$

Given the diagonal covariance of the noise fields in Fourier space, we can then multiply the probabilities for the different wavenumbers $\boldsymbol{k}$, leading to the following conditional

probability for the halo field in Fourier space, up to an irrelevant normalization constant:

$$-\ln\mathcal{P}\left(\vec{\delta}_h\middle|\vec{\delta},\{b_O\},\{\lambda_a\}\right)=\int_{\boldsymbol{k}}^{k_{\rm max}}\left[\frac{1}{2}\ln\bar{\sigma}^2(k)+\frac{1}{2\bar{\sigma}^2(k)}\left|\delta_h(\boldsymbol{k})-\delta_{h,\rm det}[\vec{\delta},\{b_O\}](\boldsymbol{k})\right|^2\right], \tag{7.6}$$

where the scale-dependent variance is given by

$$\begin{aligned}\bar{\sigma}^2(k)&=P^{\varepsilon}_{hh}(k)+2b_1P^{\varepsilon}_{hm}(k)+b_1^2P^{\varepsilon}_{mm}(k)\\&=P^{\varepsilon,0}_{hh}+\left(P^{\varepsilon,2}_{hh}+2b_1P^{\varepsilon,2}_{hm}\right)k^2+\left(P^{\varepsilon,4}_{hh}+2b_1P^{\varepsilon,4}_{hm}+b_1^2P^{\varepsilon,4}_{mm}\right)k^4\,.\end{aligned} \tag{7.7}$$

Here, we have included terms up to order k^4, which are higher order, but ensure a positive variance. The integral in Eq. (7.6) is limited to wavenumbers up to some maximum value $k_{\rm max}$, which is a meta-parameter of the likelihood that can be varied. This cutoff ensures that the assumptions on the Gaussianity and analytic power spectra of the stochastic fields are fulfilled if $k_{\rm max}$ is chosen to be sufficiently small. From the EFT standpoint, Eq. (7.6) is thus the well-defined, unique likelihood for the large-scale biased-tracer field $\delta_h(\boldsymbol{k})$ at the field level. This derivation is based on two key EFT results: the fact that on the scales of interest, there is a finite number of fields O which describe the dependence of δ_h on large-scale perturbations; and that the residual is approximately Gaussian-distributed with a power spectrum of known form (analytic in k). The scales controlling the validity of the bias expansion are well known (see Sec. 4.1.3 of [65] for a discussion). The scale controlling the Gaussianity of the likelihood has not been investigated in detail so far. We turn to that in Appendix M.

A possible concern with a Fourier-space likelihood is that effects present in real world surveys such as masks are difficult to incorporate. Briefly, a mask would be incorporated by multiplying both $\delta_{h,\rm det}$ and ε_h with the mask in real space. This leads to a non-diagonal covariance in Eq. (7.6) which is given by a convolution of white noise with the mask. Fortunately, since the mask is fixed, the computation and inversion of this covariance has to be performed only once. In the following, we will continue to assume a trivial mask, as appropriate for applications to simulations with periodic boundary conditions in Section 7.4.

Let us now consider a finite volume, in particular a cubic box of side length $L_{\rm box}$, and move to the standard box normalization for the Fourier-space fields, i.e.

$$\delta(\boldsymbol{k})=\sum_i^{N_g^3}\delta(\boldsymbol{x}_i)e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i} \tag{7.8}$$

$$\delta(\boldsymbol{x})=\frac{1}{N_g^3}\sum_{\boldsymbol{k}}^{k_{\rm Ny}}\delta(\boldsymbol{k}_i)e^{i\boldsymbol{k}_i\cdot\boldsymbol{x}}\quad\text{where}\quad\boldsymbol{k}\in(n_x,n_y,n_z)k_F\,,\ k_F=\frac{2\pi}{L_{\rm box}}\,,$$

and $n_i\in\{-N_g/2,\cdots N_g/2\}$ while $k_{\rm Ny}\equiv N_gk_F/2$. We will often use

$$\sum_{\boldsymbol{k}\neq0}^{k_{\rm max}}\equiv\sum_{\{n_x,n_y,n_z\}\neq\{0,0,0\}}^{n_x^2+n_y^2+n_z^2\leq(k_{\rm max}/k_F)^2}\,. \tag{7.9}$$

Correlators of box-normalized fields obey

$$\langle X(\boldsymbol{n}k_F)Y(\boldsymbol{n}'k_F)\rangle = \frac{1}{L^3_{\rm box}}\delta_{\boldsymbol{n},-\boldsymbol{n}'}P_{XY}(\boldsymbol{n}k_F)$$
$$\langle X(\boldsymbol{n}k_F)Y(\boldsymbol{n}'k_F)Z(\boldsymbol{n}''k_F)\rangle = \frac{1}{L^6_{\rm box}}\delta_{\boldsymbol{n}+\boldsymbol{n}',-\boldsymbol{n}''}B_{XYZ}(\boldsymbol{n}k_F,\boldsymbol{n}'k_F)\,, \tag{7.10}$$

where

$$\delta_{\boldsymbol{n},\boldsymbol{n}'} \equiv \delta_{n_x n'_x}\delta_{n_y n'_y}\delta_{n_z n'_z}\,, \tag{7.11}$$

and P_{XY} (B_{XYZ}) are the cross-power spectrum (bispectrum) respectively. Note that we have neglected the averaging over the finite k-space bin of width $k_{\rm Ny}$ on the r.h.s. of Eq. (7.10).

Eq. (7.6) then becomes

$$\ln\mathcal{P}\left(\vec{\delta}_h\middle|\vec{\delta},\{b_O\},\{\lambda_a\}\right) = -\sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}}\left[\frac{1}{2}\ln\sigma^2(k) + \frac{1}{2\sigma^2(k)}\left|\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}[\vec{\delta},\{b_O\}](\boldsymbol{k})\right|^2\right]\,, \tag{7.12}$$

where $\sigma^2(k)$ is now a dimensionless variance given by

$$\sigma^2(k) \equiv L^{-3}_{\rm box}\bar{\sigma}^2(k) = V^{\varepsilon,0}_{hh} + \left(V^{\varepsilon,2}_{hh} + 2b_1V^{\varepsilon,2}_{hm}\right)k^2 + \left(V^{\varepsilon,4}_{hh} + 2b_1V^{\varepsilon,4}_{hm} + b_1^2V^{\varepsilon,4}_{mm}\right)k^4\,, \tag{7.13}$$

and we have defined the noise variance parameters

$$V^{\varepsilon,n}_{xy} \equiv L^{-3}_{\rm box}P^{\varepsilon,n}_{xy}\,. \tag{7.14}$$

It is worth pointing out the very different interpretation of the EFT conditional probability derived here as compared to standard forward modeling approaches which use a local, real-space conditional probability as in Eq. (4.4). Instead of approximating a likelihood of unknown shape that is localized in physical space, we are expanding the *perturbatively known non-local Fourier-space likelihood* up to a cutoff $k_{\rm max}$. This point will become more clear in the following. Note that one can generalize the diagonal covariance to a non-diagonal one, including the non-Gaussian contributions generated by non-linear evolution. Again, we will turn to this in Appendix M.

Finally, as long as errors in the forward model are captured by the likelihood Eq. (7.12), the full posterior Eq. (4.6) for the initial phases, cosmological parameters, and bias parameters becomes

$$\mathcal{P}\left(\vec{\delta}_{m,\rm ini},\theta,\{b_O\},\{\lambda_a\}\middle|\vec{\delta}_h\right) = \mathcal{N}_P\mathcal{P}_{\rm prior}(\vec{\delta}_{m,\rm ini}|\theta)\exp\left[\ln\mathcal{P}\left(\vec{\delta}_h\middle|\vec{\delta}_{\rm fwd}[\delta_{\rm in},\theta],\{b_O\},\{\lambda_a\}\right)\right]\,, \tag{7.15}$$

where

$$\{b_O\} = \{b_1\,,\ b_2\,,\ b_{K^2}\}\cup\{c_{\nabla^2\delta}\}$$
$$\{\lambda_a\} = \{V^{\varepsilon,0}_{hh}\,,\ V^{\varepsilon,2}_{hh}\,,\ V^{\varepsilon,2}_{hm}\,,\ V^{\varepsilon,4}_{hh}\,,\ V^{\varepsilon,4}_{hm}\,,\ V^{\varepsilon,4}_{mm}\}\,. \tag{7.16}$$

In practice, we can limit the set $\{\lambda_a\}$ to the first three parameters, as the others are higher order (see the next section and Appendix D of [123]).

7.1.1 Analytical maximum-likelihood point: bias parameters

Eq. (7.15) is clearly a highly complex, non-linear and nonlocal (in terms of $\vec{\delta}_{m,\mathrm{ini}}$) posterior. In order to make progress in our physical understanding, let us assume that we have fixed the initial density field $\vec{\delta}_{m,\mathrm{ini}}$ to the true field. This thought example is easy to realize when applying the forward model to the results of simulations, whose initial conditions are known. For now, we will also fix the cosmology parameters θ to their true values.

Let us then consider the maximum-likelihood point for the bias parameter b_O, keeping all other parameters fixed (we will turn to the parameters controlling the variance below). For a fixed operator O, the maximum-likelihood point is given by

$$-\frac{\partial}{\partial b_O} \ln \mathcal{P}\left(\vec{\delta}_h | \vec{\delta}, \{b_O\}, \{\lambda_a\}\right) = \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \frac{1}{\sigma^2(k)} O(\boldsymbol{k}) \left(\delta_h - \delta_{h,\mathrm{det}}[\vec{\delta}, \{b_O\}]\right)^*_{\boldsymbol{k}} = 0\,. \tag{7.17}$$

In the limit of infinite volume (e.g., a large number of simulation realizations), the products of fields in this relation approach their ensemble averages, i.e. their cross-power spectra. Then, the maximum-likelihood point becomes

$$\begin{aligned}\sum_{\boldsymbol{k}}^{k_\mathrm{max}} \frac{1}{\sigma^2(k)} \langle O(\boldsymbol{k}) \delta_h^*(\boldsymbol{k})\rangle = \sum_{O'} \hat{b}_{O'} \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \frac{1}{\sigma^2(k)} \langle O(\boldsymbol{k}) O'^*(\boldsymbol{k})\rangle \qquad \forall\ O \\ \overset{O=\delta}{+}\ V_{hm}^{\varepsilon,2} k^2 + V_{hm}^{\varepsilon,4} k^4 + b_1 V_{mm}^{\varepsilon,4} k^4\,.\end{aligned} \tag{7.18}$$

The last line, which comes from the b_1-dependence of $\sigma^2(k)$ [Eq. (7.13)] is only present when considering the operator δ. Eq. (7.18) corresponds to matching specific filtered (and weighted) moments in Fourier space, with a correction that scales analytically with k and takes into account the correlation of noise in the halo field and matter.

By restricting $k_\mathrm{max} < k_\mathrm{NL}$, the non-linear scale, these filtered moments can be kept under perturbative control. Recall that the operators O here are still constructed by taking non-linear transformations of the density field on the grid. Thus, there are now two filters involved: the kernel corresponding to the density assignment, and the sharp-k filter for the moments on the larger length scale $1/k_\mathrm{max}$. This is in fact closely related to the approach followed by [188, 189], who showed that one can efficiently obtain bias parameters up to cubic order using this approach.

Finally, note that including higher-derivative contributions in the bias expansion Eq. (4.3) is very simple in the EFT likelihood, by generalizing $\delta_{h,\mathrm{det}}$ in Fourier space to

$$\delta_{h,\mathrm{det}}[\vec{\delta}, \{b_O\}](\boldsymbol{k}) = \sum_O b_O O(\boldsymbol{k}) \longrightarrow \sum_O \left[b_O - c_{\nabla^2 O} k^2\right] O(\boldsymbol{k})\,, \tag{7.19}$$

introducing an additional effective higher-derivative bias parameter $c_{\nabla^2 O}$ which can be marginalized over to take into account improperly-modeled higher-order contributions. We continue to only include such a higher-derivative contribution $\propto c_{\nabla^2\delta}$ for the density

field, $O = \delta$. A detailed discussion on why this is sufficient can be found in Section 4.4 of [123].

In order to gain a more explicit understanding of Eq. (7.18), let us drop the second line, present if $O = \delta$, for the remainder of this section, and continue to take the infinite-volume limit. We then have

$$\sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \frac{1}{\sigma^2(k)} \langle O(\boldsymbol{k})\delta_h^*(\boldsymbol{k})\rangle = \sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \frac{1}{\sigma^2(k)} \sum_{O'} \hat{b}_{O'} \langle O(\boldsymbol{k})O'^*(\boldsymbol{k})\rangle \qquad \forall\ O\,, \tag{7.20}$$

which is to be solved for the bias parameters $\hat{b}_O$. Let us consider this equality at fixed $\boldsymbol{k}$. If the set of bias operators in Eq. (4.3) is of size N_O, Eq. (7.20) is a linear system of N_O equations that can be straightforwardly solved, yielding an estimator $\hat{\boldsymbol{b}}(\boldsymbol{k})$ for the set of coefficients (see also [188, 189]):

$$\hat{\boldsymbol{b}}(\boldsymbol{k}) = \boldsymbol{M}^{-1}(\boldsymbol{k}) \cdot \boldsymbol{H}(\boldsymbol{k})\,, \qquad \boldsymbol{M} = \{\langle OO'^*\rangle\}_{O,O'}\,; \quad \boldsymbol{H} = \{\langle \delta_h^* O\rangle\}_O\,. \tag{7.21}$$

Eq. (7.20) then corresponds to a weighted mean of the estimated bias parameters $\hat{\boldsymbol{b}}(\boldsymbol{k})$ over $\boldsymbol{k}$.

Similarly, taking the derivative of Eq. (7.12) with respect to $\sigma^2(k)$ yields maximum-likelihood values for the various components of $\sigma^2(k)$, as derived in Appendix D of [123]. In particular, the ML point for the constant part $V_{hh}^{\varepsilon,0} \equiv \lim_{k\to 0}\sigma^2(k)$ is given by, in the infinite-volume limit:

$$0 = \sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \frac{1}{\sigma^4(k)} \left[\frac{1}{2}\sigma^2(k) - \left\langle |\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}(\boldsymbol{k})|^2 \right\rangle\right]\,. \tag{7.22}$$

Thus, inaccuracies in the model directly contribute to the effective variance in the likelihood (assuming one allows σ^2 to vary). A deficient model thus lowers the amount of information that can be extracted, as expected. In order for the likelihood to be consistent however, it is of course necessary that the noise correlator $\langle|\delta_h - \delta_{h,\rm det}|^2(\boldsymbol{k})\rangle$ is in fact analytic in k, as assumed in Eq. (7.13). We will see in the next section why and under what conditions this holds. At the order we work in throughout this chapter, it is sufficient to include terms up to order k^2 in $\sigma^2(k)$, which corresponds to the first three parameters in Eq. (7.16). While the first, $V_{hh}^{\varepsilon,0}$ corresponds to the physical halo shot noise in the large-scale limit, the other two contributions are effective stochastic parameters which absorb residuals of the model.[1]

7.1.2 Analytical maximum-likelihood point: cosmological parameters

We now move on to the estimation of cosmological parameters. The most simple, and interesting, cosmological parameter in this context is the normalization of the primordial

[1] Indeed, beyond the large-scale limit there exists no unique definition of stochasticity (e.g., [190, 191]).

perturbations (or equivalently linear matter power spectrum). Following convention, we parametrize this through the variance $(\sigma_8)^2$ at redshift zero of the linear density field filtered with a real-space tophat kernel on the scale $8\,h^{-1}$Mpc. This parameter is interesting, since, at the linear level, it is exactly degenerate with the linear bias b_1. Thus, we need non-linear information in the measured density field $\vec{\delta}_h$ to break this degeneracy, which is a highly nontrivial test of the forward model.

In order to investigate this, we assume that the operators appearing in Eq. (4.3) [e.g., Eq. (7.1)] are split by perturbative order so that each scales homogeneously with σ_8. For convenience, we define $\alpha \equiv \sigma_8/\sigma_{8,\mathrm{fid}}$, where the fiducial value $\sigma_{8,\mathrm{fid}}$ is fixed. We can then write

$$O(\sigma_8) = \alpha^{n_{s,O}} O(\sigma_{8,\mathrm{fid}})\,, \tag{7.23}$$

so that, e.g., $n_{s,\delta^{(1)}} = 1, n_{s,\delta^{(2)}} = n_{s,\delta^2} = 2$, and so on. We then have

$$\vec{\delta}_{h,\mathrm{det}}(\{b_O\}, \sigma_8) = \sum_O b_O \alpha^{n_{s,O}} \vec{O}(\sigma_{8,\mathrm{fid}})\,. \tag{7.24}$$

In the following, all operators O will be assumed to be evaluated at $\sigma_{8,\mathrm{fid}}$, and we will omit this dependence for clarity. We obtain, again neglecting the terms in the second line of Eq. (7.18),

$$-\frac{\partial}{\partial b_O} \ln P\left(\vec{\delta}_h | \vec{\delta}, \{b_O\}, \{\lambda_a\}\right) = \sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \frac{1}{\sigma^2(k)} \alpha^{n_{s,O}} O(\boldsymbol{k}) \left(\delta_h - \delta_{h,\mathrm{det}}[\vec{\delta}, \{b_O\}]\right)^*_{\boldsymbol{k}} = 0 \tag{7.25}$$

$$-\frac{\partial}{\partial \ln \sigma_8} \ln P\left(\vec{\delta}_h | \vec{\delta}, \{b_O\}, \{\lambda_a\}\right) = \sum_O b_O n_{s,O} \alpha^{n_{s,O}} \sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \frac{1}{\sigma^2(k)} O(\boldsymbol{k}) \left(\delta_h - \delta_{h,\mathrm{det}}[\vec{\delta}, \{b_O\}, \alpha]\right)^*_{\boldsymbol{k}} = 0\,.$$

In the infinite-volume limit and at fixed $\boldsymbol{k}$, we thus have

$$\begin{aligned} \langle O \delta_h^* \rangle_{\boldsymbol{k}} &= \sum_{O'} b_{O'} \alpha^{n_{s,O'}} \langle O O'^* \rangle_{\boldsymbol{k}} \\ \sum_O b_O n_{s,O} \alpha^{n_{s,O}} \langle O \delta_h^* \rangle_{\boldsymbol{k}} &= \sum_{O,O'} b_O n_{s,O} b_{O'} \alpha^{n_{s,O}+n_{s,O'}} \langle O O'^* \rangle_{\boldsymbol{k}}\,. \end{aligned} \tag{7.26}$$

We now define scaled parameters

$$\beta_O \equiv b_O \alpha^{n_{s,O}}\,. \tag{7.27}$$

Eq. (7.26) then becomes

$$\begin{aligned} \langle O \delta_h^* \rangle_{\boldsymbol{k}} &= \sum_{O'} \beta_{O'} \langle O O'^* \rangle_{\boldsymbol{k}} \\ \sum_O n_{s,O} \beta_O \langle O \delta_h^* \rangle_{\boldsymbol{k}} &= \sum_{O,O'} n_{s,O} \beta_O \beta_{O'} \langle O O'^* \rangle_{\boldsymbol{k}}\,. \end{aligned} \tag{7.28}$$

Contracting the first line with $\sum_O n_{s,O}\beta_O$ immediately shows that the second line is linearly dependendent on the first line, which contains N_O equations. That is, Eq. (7.28) in fact contains (at most) N_O independent equations for N_O+1 unknowns. This is not surprising: if all operators have independent bias parameters, then any change in σ_8 can be absorbed by a change in the N_O bias parameters. Thus, we need to rely on some relations between bias parameters of different operators. In our fiducial application of the EFT likelihood, these are $\delta^{(1)}$ and $\delta^{(2)}$ which are both multiplied by b_1.

At fixed $\boldsymbol{k}$, we thus have for the maximum-likelihood point of $\boldsymbol{\beta}$:

$$\begin{aligned} &\boldsymbol{H}(\boldsymbol{k}) = \boldsymbol{M}(\boldsymbol{k}) \cdot \boldsymbol{\beta}(\boldsymbol{k}) \\ \Rightarrow \quad &\hat{\boldsymbol{\beta}}(\boldsymbol{k}) = \boldsymbol{M}^{-1}(\boldsymbol{k}) \cdot \boldsymbol{H}(\boldsymbol{k})\,. \end{aligned} \tag{7.29}$$

This relation is of course of the same form as derived above in Eq. (7.21) (without considering σ_8). Crucially, in order to break the degeneracy between bias parameters and σ_8, there must be at least two non-degenerate operators multiplied by the same bias parameters, which scale differently with σ_8. Physically, this is the case at second order in perturbations due to the displacement term inside $\delta^{(2)}$. Since large-scale galaxy displacements have to be the same as those of matter due to the equivalence principle (see Section 2.7 of [65] for a discussion), this displacement term has to be multiplied by the same bias b_1 as the linear density field. The estimation of σ_8 using the EFT likelihood is hence based only on this fundamental physical constraint.

Finally, the full maximum-likelihood point then corresponds to a weighted average over $\boldsymbol{k}$, following Eq. (7.25) [and, in practice, including the contributions in the second line of Eq. (7.18)].

7.2 Summary of the EFT-Fourier likelihood

Before marginalizing and profiling out the bias parameters to focus on σ_8, let us first summarize the main findings of [123] here. Readers who are interested in the technical details can find them in Section 4 of [123] (see also [187]). The key result is that Eq. (7.18) indeed yields *unbiased results for σ_8 and the bias parameters which are under perturbative control*, if the following conditions are met:

1. A sharp-k filter $W(\boldsymbol{k}) = W_\Lambda(\boldsymbol{k}) \equiv \Theta_\mathrm{H}(\Lambda - \|\boldsymbol{k}\|)$ (cf. Eq. (4.11) of [123]), where Λ is a cutoff scale, is used to obtain the smoothed density field out of which the operators O are constructed.

2. k_max is smaller than the cutoff Λ of this filter.

3. The operators $O \to [O]$ are constructed as *renormalized* operators, as explained in Appendix J, where counterterms that are relevant at the perturbative order we work in are subtracted.

4. The list of operators includes $\nabla^2\delta(\boldsymbol{x}) \leftrightarrow -k^2\delta(\boldsymbol{k})$, and the likelihood involves a specific scale-dependent variance term $\sigma^2(k)$ which scales analytically with k [Eq. (7.13)] and depends on b_1. Both of these ingredients are important to consistently absorb higher-order contributions.

For matter and halo density fields that satisfy conditions 1.-3. above, the conditional probability or "likelihood" for the halo density field given the evolved matter density in Fourier space can be written as

$$\ln \mathcal{P}\left(\vec{\delta}_h \middle| \vec{\delta}, \{b_O\}\right) = -\sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \left[\frac{1}{2}\ln\sigma^2(k) + \frac{1}{2\sigma^2(k)}\left|\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}[\vec{\delta}, \{b_O\}](\boldsymbol{k})\right|^2\right] . \tag{7.30}$$

Here, to shorten the notation we have included both deterministic and stochastic bias parameters in $\{b_O\}$; $\sigma^2(k)$ is a dimensionless variance that is analytic in k, specifically a power series in k^2 (cf. Eq. (7.13)). For our numerical implementation in Section 7.5, we will include terms up to order k^4, and parametrize $\sigma^2(k)$ as follows:

$$\sigma^2(k) = \left(\sigma_\varepsilon + k^2[\sigma_{\varepsilon,2} + b_1\sigma_{\varepsilon\varepsilon_m,2}]\right)^2 . \tag{7.31}$$

The parametrization is chosen so that $\sigma^2(k)$ is positive definite. σ^2_ε can be interpreted as the amplitude of halo stochasticity in the large-scale limit ($k \to 0$). In Appendix L we derive the expectation for σ^2_ε for a Poisson process, which will be useful for the interpretation of numerical results on this parameter. $\sigma^2_{\varepsilon,2}$ is the leading scale-dependent correction to the halo stochasticity (see Section 2.7 of [65] for a discussion). Finally, $\sigma_{\varepsilon\varepsilon_m,2}$ captures the cross-correlation of stochasticity in the halo and matter fields; the stochasticity in the matter field $\varepsilon_m(\boldsymbol{k})$ is constrained by mass- and momentum conservation to be of order k^2 on large scales.

The closed-form expression for the likelihood in Eq. (7.12) facilitates a straightforward way to derive maximum likelihood estimates for bias parameters, and, as shown in [123], for the cosmological parameter σ_8. This analytical approach was used in [123] to demonstrate unbiased parameter estimation from the same suite of simulations described in Section 7.4 and used in Section 7.5, Section 7.6. We summarize the result in Appendix O.

The degeneracy between b_1 and σ_8, which is perfect in linear theory, is broken when including non-linear information. In particular, the fact that the displacement term contained in the second-order matter density is also multiplied by b_1, coupled with the fact that the second-order matter density scales differently with σ_8 than the linear-order one. Thus, fundamentally, the possibility of estimating σ_8 in this way is due to the equivalence principle, which ensures that galaxies move on the same trajectories as matter on large scales, and thus requires that the second-order displacement term is multiplied by the same bias coefficient as the linear-order density field (see also Section 2 of [65]). At higher orders in perturbations, more such terms that are protected by the equivalence principle appear, and the EFT likelihood will consistently capture those as well once extended to higher order.

7.3 Marginalizing over bias parameters

In the approach summarized in the previous section, a number of nuisance parameters have been introduced to capture the uncertainties associated with some of the poorly understood physical processes describing halo and galaxy formation. In practical applications, these parameters would then be estimated alongside cosmological parameters in a statistical analysis of observational data. Interestingly, owing to the simple functional form of the likelihood in Eq. (7.12), it is easily possible to analytically marginalize over some of the bias parameters b_O. Here, we consider only those that do not appear in the variance $\sigma^2(k)$; given Eq. (7.31), this includes all bias parameters except b_1. While it might be possible to extend the analytical marginalization to parameters which appear in the variance, the marginalization performed here is sufficient for our purposes.

Let us thus write

$$\delta_{h,\mathrm{det}}(\boldsymbol{k}) = \mu(\boldsymbol{k}) + \sum_{O\in\mathcal{O}_\mathrm{marg}} b_O O(\boldsymbol{k})\,, \qquad \mu(\boldsymbol{k}) = \sum_{O\in\mathcal{O}_\mathrm{all}\setminus\mathcal{O}_\mathrm{marg}} b_O O(\boldsymbol{k})\,, \tag{7.32}$$

where $\mathcal{O}_\mathrm{marg}$ denotes the subset of operators, whose bias parameters we wish to marginalize over (we denote the cardinality of this set as n_marg). We can then write the likelihood Eq. (7.12) as

$$\begin{aligned}\mathcal{P}\left(\vec{\delta}_h\middle|\vec{\delta},\{b_O\}\right) = \frac{1}{\mathcal{N}}\exp\left[-\frac{1}{2}\sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}}\ln\sigma^2(k)\right] \\ \times\exp\left\{-\sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}}\left[\frac{|\delta_h(\boldsymbol{k})-\mu(\boldsymbol{k})|^2}{2\sigma^2(k)} - 2\sum_{O\in\mathcal{O}_\mathrm{marg}} b_O\frac{\Re\left[\delta_h(\boldsymbol{k})-\mu(\boldsymbol{k})\right]O^*(\boldsymbol{k})}{2\sigma^2(k)}\right.\right. \\ \left.\left. + \sum_{O,O'\in\mathcal{O}_\mathrm{marg}} b_O b_{O'}\frac{O(\boldsymbol{k})O'^*(\boldsymbol{k})}{2\sigma^2(k)}\right]\right\},\end{aligned} \tag{7.33}$$

where $\mathcal{N}$ is a normalization constant which is independent of all parameters. This expression can be more compactly written as

$$\begin{aligned}\mathcal{P}\left(\vec{\delta}_h\middle|\vec{\delta},\{b_O\}\right) = \frac{1}{\mathcal{N}}\exp\left[-\frac{1}{2}\sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}}\ln\sigma^2(k)\right] \\ \times\exp\left\{-\frac{1}{2}C + \sum_{O\in\mathcal{O}_\mathrm{marg}} b_O B_O - \frac{1}{2}\sum_{O,O'\in\mathcal{O}_\mathrm{marg}} b_O b_{O'} A_{OO'}\right\}\,,\end{aligned} \tag{7.34}$$

where

$$C = \sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \frac{1}{\sigma^2(k)} \left|\delta_h(\boldsymbol{k}) - \mu(\boldsymbol{k})\right|^2$$
$$B_O = \sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \frac{\Re\left[\delta_h(\boldsymbol{k}) - \mu(\boldsymbol{k})\right] O^*(\boldsymbol{k})}{\sigma^2(k)}$$
$$A_{OO'} = \sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \frac{O(\boldsymbol{k}) O'^*(\boldsymbol{k})}{\sigma^2(k)} \,. \tag{7.35}$$

Note that $A_{OO'}$ is a Hermitian and positive-definite matrix. The former is immediately obvious from its definition. The latter follows from the fact that $A_{OO'}$ is the zero-lag covariance matrix of a set of sharp-k-filtered real fields $O(\boldsymbol{x})$. Eq. (7.34) then allows us to perform the Gaussian integral over the b_O. Here, we will assume uninformative priors, although Gaussian priors can trivially be introduced by adding a prior covariance to $A_{OO'}$. The result is[2]

$$\begin{aligned}\mathcal{P}\left(\vec{\delta}_h \middle| \vec{\delta}, \{b_O\}_{\text{unmarg}}\right) &= \left(\prod_{O\in\mathcal{O}_{\text{marg}}} \int db_O\right) \mathcal{P}\left(\vec{\delta}_h \middle| \vec{\delta}, \{b_O\}\right) \\ &= \frac{(2\pi)^{n_{\text{marg}}/2}}{\mathcal{N}} \left|A_{OO'}\right|^{-1/2} \exp\left[-\frac{1}{2}\sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \ln\sigma^2(k)\right] \\ &\quad \times \exp\left\{-\frac{1}{2}C(\{b_O\}) + \frac{1}{2}\sum_{O,O'\in\mathcal{O}_{\text{marg}}} B_O(\{b_O\})(A^{-1})_{OO'} B_{O'}(\{b_O\})\right\} .\end{aligned} \tag{7.36}$$

We have thus reduced the parameter space from $\{b_O\}$ to $\{b_O\}_{\text{unmarg}}$. This marginalization applies to an arbitrary number of bias coefficients to be marginalized over. The price to pay is that we now need to compute the vector B_O and the matrix $A_{OO'}$. Further, we need the determinant of $A_{OO'}$ and its inverse. Note, however, that n_{marg} – and therefore the size of B_O and $A_{OO'}$ – will rarely become a large number in practical applications. More specifically, in the tests presented below, we marginalize over $c_{\nabla^2\delta}, b_2, b_{K^2}$ and so $n_{\text{marg}} = 3$. Further, $A_{OO'}$ is independent of the remaining, unmarginalized bias parameters. On the other hand, $A_{OO'}$ does depend on the parameters entering the variance $\sigma^2(k)$, Eq. (7.31).

At this point, let us briefly comment on the relation of this analytic marginalization to other approaches presented in the recent literature. In particular, Refs. [189, 192] perform

[2]Using the well-known Gaussian integral identity

$$\int d^n\vec{x} \exp\left[-\frac{1}{2}\vec{x}^\top A\vec{x} + \vec{B}^\top\cdot\vec{x}\right] = \frac{(2\pi)^{n/2}}{|A|^{1/2}} \exp\left[\frac{1}{2}\vec{B}^\top A^{-1}\vec{B}\right] .$$

a Gram-Schmidt orthogonalization on the fields entering the bias expansion Eq. (4.3), arguing that lower-order operators thus become independent of higher-order fields and make their corresponding bias parameters more robust to higher-order corrections. This approach is directly related to the analytic marginalization pointed out here. To see this, consider the case where an orthogonalization has been performed on the operators. In particular, this implies that $\langle \mu(\boldsymbol{k}) O(\boldsymbol{k}) \rangle = 0$ for all $O \in \mathcal{O}_{\rm marg}$. This in turn renders B_O independent of all unmarginalized bias parameters, so that it becomes a constant vector. Then, the marginalized likelihood reduces to the same form as the unmarginalized likelihood keeping only the terms involving unmarginalized bias parameters. In this sense, orthogonalization is equivalent to marginalizing over bias parameters. The computational cost of both approaches is expected to be essentially the same. However, the marginalization described here does offer the possibility of including prior information on the bias parameters that are marginalized over.

7.3.1 Estimating systematic errors

One of the crucial advantages of the rigorous perturbative approach pursued here is that it allows for an estimate of the systematic error due to imperfections in the likelihood. We can distinguish three principal sources of such systematic errors (see [187] for more details on Types 2 and 3 in particular):

Type 1: Errors in the forward model for the matter density field (and correspondingly the operators constructed from it);

Type 2: Higher-order bias terms neglected in the expansion in $\delta_{h,\rm det}$;

Type 3: Higher-order contributions to the variance $\sigma^2(k)$ as well as in the form of the likelihood itself.

The most rigorous way to evaluate the size of these contributions is to include the set of leading higher-order terms that have been neglected in the forward model, bias expansion, and likelihood, and evaluate the shift in resulting parameter values. In case of the forward model (Type 1), this can be tested by using the density field from N-body simulations instead of 2LPT to construct the bias operators. This will be presented in Section 7.6. In case of the bias expansion (Type 2), this test is not too difficult either, since the coefficients can be marginalized over analytically, as shown in the previous section. We defer an implementation of the higher-order contributions to future work however.

Let us here approximately estimate the size and scaling with $k_{\rm max}$ of the systematic error of Type 2. Note that strictly speaking we have two cutoffs: the cutoff Λ of the sharp-k filter, and $k_{\rm max} < \Lambda$. In practice, one will choose $k_{\rm max}$ as a fixed fraction of Λ (see Section 7.6); hence, it is sufficient to consider the dependence on Λ here. For simplicity, we evaluate the systematic shift in the bias parameters b_O. As described in [123], one can similarly evaluate the shift in σ_8 by introducing scaled bias parameters β_O. We will also count higher-derivative terms as higher order in perturbations, which assumes that the

scale controlling higher-derivative terms does not differ greatly from the non-linear scale. Thus, "higher-order contributions" include higher-derivative contributions in what follows.

Incorporating both sources of error described above, the *correct* likelihood can be written as

$$\begin{aligned} -2\ln\mathcal{P}(\delta_h|\delta) &= \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \left[\ln\sigma^2(k) + \frac{1}{\sigma^2(k)} \left| \delta_h(\boldsymbol{k}) + \delta_h^{\mathrm{h.o.}}(\boldsymbol{k}) - \sum_O b_O \left(O + O_\mathrm{err}\right)(\boldsymbol{k}) \right|^2 \right] \\ &= -2\ln\mathcal{P}_\mathrm{fid}(\delta_h|\delta) - 2\sum_O B_O^\mathrm{err} b_O + \sum_{O,O'} A_{OO'}^\mathrm{err} b_O b_{O'}\,, \end{aligned} \tag{7.37}$$

where

$$\begin{aligned} B_O^\mathrm{err} &= \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \frac{1}{\sigma^2(k)} \Re\left[O_\mathrm{err}(\boldsymbol{k}) \delta_h^*(\boldsymbol{k}')' + O^*(\boldsymbol{k}) \delta_h^{\mathrm{h.o.}*}(\boldsymbol{k}')' \right] \\ A_{OO'}^\mathrm{err} &= \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \frac{1}{\sigma^2(k)} O_\mathrm{err}(\boldsymbol{k}) O'^*_\mathrm{err}(\boldsymbol{k}')'\,. \end{aligned} \tag{7.38}$$

Here, $O_\mathrm{err}(\boldsymbol{k})$ denotes the error field in the operator O due to deficiencies in the forward model, while $\delta_h^{\mathrm{h.o.}}(\boldsymbol{k})$ denotes the higher-order bias contributions to the actual halo density field. Finally, $P_\mathrm{fid}(\delta_h|\delta)$ stands for the fiducial likelihood, which differs from the correct one due to the systematic error terms. In the second line of Eq. (7.37), we have dropped an irrelevant constant term which does not depend on the parameters being varied. Let us write the fiducial likelihood in analogy to Eq. (7.34) as

$$\begin{aligned} -2\ln\mathcal{P}_\mathrm{fid}(\delta_h|\delta) &= \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \ln\sigma^2(k) + \sum_{O,O'} b_O b_{O'} A_{OO'} - 2\sum_O B_O b_O\,, \quad \text{where} \\ A_{OO'} &= \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \frac{1}{\sigma^2(k)} O(\boldsymbol{k}) O'^*(\boldsymbol{k}')' \\ B_O &= \sum_{\boldsymbol{k}\neq 0}^{k_\mathrm{max}} \frac{1}{\sigma^2(k)} \Re O(\boldsymbol{k}) \delta_h^*(\boldsymbol{k}')\,. \end{aligned} \tag{7.39}$$

Under the assumption that the parameter shift Δb_O due to the systematic errors is small, one can immediately solve for this shift based on the maximum-likelihood points of the correct and fiducial likelihoods. One can then estimate the expected amplitude of the shift by taking the expectation values of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{A}^\mathrm{err}, \boldsymbol{B}^\mathrm{err}$. This is closely analogous to the "Fisher bias". Using bold-face to denote vectors in the n_O-dimensional vector space of operators considered, we obtain in matrix notation

$$\Delta\boldsymbol{B} \equiv \boldsymbol{B} - \boldsymbol{B}^\mathrm{fid} = \left(\boldsymbol{A}^{-1}\boldsymbol{A}^\mathrm{err}\right)\left(\boldsymbol{B}^\mathrm{fid}\right) - \boldsymbol{A}^{-1}\left(\boldsymbol{B}^\mathrm{err}\right)\,. \tag{7.40}$$

This expression involves the expectation value of the correlators in Eq. (7.38) and Eq. (7.39) which are straightforward to evaluate in perturbation theory. We begin by estimating at which order in perturbation theory the various correlators contribute.

First, the expectation values of $A_{OO'}$ and B_O are of order $P_{\rm L}(k) + P_{\text{1-loop}}(k)$ (in case of $A_{\delta\delta}$ and B_δ), or of order $P_{\text{1-loop}}(k)$ (all other elements). On the other hand, both $\boldsymbol{B}^{\rm err}$ and $\boldsymbol{A}^{\rm err}$ are of order $P_{\text{2-loop}}(k)$. To see this, notice that both $O^{\rm err}$ and $\delta_h^{\text{h.o.}}$ are at least of cubic order in the linear density field. This means that all correlators which involve these error fields are two-loop contributions, apart from the cross-correlation with $\delta^{(1)}$, which is at 1-loop order. The latter however only appears in $B_\delta^{\rm err}$, via $\left\langle \delta^{{\rm err},(3)}\delta_h^{(1)} \right\rangle$ and $\left\langle \delta_h^{(3)}\delta^{(1)} \right\rangle$. As we argue in App. C of [123], these particular 1-loop contributions are of very similar shape as that coming from the higher-derivative bias, and are thus largely absorbed by $c_{\nabla^2\delta}$.

Thus, without performing any detailed calculation, we can very roughly estimate that

$$\Delta b\Big|_{\rm loops} \sim \frac{\sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \sigma^{-2}(k) P_{\text{2-loop}}(k)}{\sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \sigma^{-2}(k) P_{\rm L}(k)} \,. \tag{7.41}$$

As an approximate estimate of the size of two-loop correlators, we will use the auto-correlation of $[\delta^3]$:

$$P_{\text{2-loop}}(k) \sim \left\langle [\delta^3](\boldsymbol{k})[\delta^3](\boldsymbol{k}') \right\rangle' = 6 \int d^3r \, [\xi_{\rm L}(r)]^3 e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \,, \tag{7.42}$$

where $\xi_{\rm L}(r)$ is the linear matter correlation function. We emphasize that this is a very rough estimate: in reality, $\Delta\boldsymbol{B}$ involves many different contributions with various order-unity coefficients, which could add up or partially cancel. The main prediction of Eq. (7.41) is the scaling with $k_{\rm max}$.

Finally, let us consider systematics of Type 3, i.e. higher-order terms in the likelihood itself. Similar to the bias expansion, these come in two forms: an expansion in powers of k, equivalent to spatial derivatives; and an expansion in powers of perturbations, in this case the error field $\varepsilon(\boldsymbol{k})$ whose variance is σ_ε^2. Beginning with the former, a naive counting following loop contributions to the power spectrum indicates that a term Ck^2, where C is a constant and which corresponds to the term $\propto \sigma_\varepsilon\sigma_{\varepsilon,2}$ in Eq. (7.31), is of 2-loop order (see, e.g., Section 4.1. of [65]). Hence, terms of order $C'k^4$ should have a negligible impact. This is corroborated by our numerical results (Section 7.6). The second type of higher-order stochasticity corresponds to non-Gaussian corrections such as the stochastic three-point function $\langle\varepsilon\varepsilon\varepsilon\rangle$ as well as coupling between stochasticity and the long-wavelength perturbations. These are briefly discussed in Section 5 of [123] (see also [187]). A derivation of the precise form of these contributions to the likelihood requires more theoretical investigation, and is left for future work. However, we can guess the approximate magnitude of these contributions by relating them to the terms we have kept here, which are non-linear in the long-wavelength modes that determine δ_Λ. The higher-

order stochastic contributions are expected to be suppressed at least by

$$\Delta b\Big|_{\text{stoch.}} \sim \frac{|\varepsilon(\boldsymbol{k})|}{|\delta(\boldsymbol{k})|}\Big|_{k_{\max}} \sim \sqrt{\frac{P_\varepsilon}{P_{\text{L}}(k_{\max})}} \overset{\text{Poisson}}{=} \left[\bar{n} P_{\text{L}}(k_{\max})\right]^{-1/2}, \tag{7.43}$$

where in the last equality we have assumed the Poisson expectation, $P_\varepsilon = 1/\bar{n}$ where $\bar{n}$ is the mean number density of halos, which is a reasonable first-order estimate for this purpose. While the proper result will involve a summation over $\boldsymbol{k}$ modes similar to Eq. (7.41), we conservatively evaluate the ratio at $k_{\max}$ here, as it is unclear what precise weighting should be employed for this type of higher-order contribution. Notice that Eq. (7.43) also approaches 0 as $k \to 0$, but depends sensitively on the abundance of halos. In particular, it becomes large for small number densities.

7.3.2 Profiling the likelihood

We now introduce the framework used in our numerical tests to obtain maximum likelihood values and confidence intervals for cosmological parameters. Below, we center our discussion around the normalization of the primordial power spectrum, described by σ_8. The reasoning for choosing σ_8 as parameter is that it can only be inferred by using information in the non-linear density field, as mentioned in Section 7.1. An unbiased inference thus means that the specific part of the information content in the non-linear density field that is robust has been properly isolated. In particular, non-linear information is explicitly necessary in order to break the degeneracy with the bias parameter b_1, rendering it the most direct test of our non-linear inference approach. Future work will consider other cosmological parameters as well.

Besides from being a function of σ_8, the marginalized likelihood Eq. (7.36) also depends on bias and other nuisance parameters (including the entire set of Fourier modes of the three-dimensional matter density field). Since the probabilistic inference of the initial matter density field from tracers like a halo catalog is numerically very expensive (see e.g. [127]), we instead constrain it to the actual initial conditions used in the simulations, evolved to low redshifts using either second-order Lagrangian perturbation theory (2LPT), or the N-body code directly. This forward evolution is then performed for a set of discrete σ_8 values around the fiducial $\sigma_8 = 0.85$ (see Section 7.5). We can then maximize the likelihood to simultaneously obtain best-fit values for cosmological and the remaining nuisance parameters. This is in fact the most stringent test possible for any systematic bias, since the absence of any flexibility in the phases means that there is less room for errors in the likelihood to be absorbed by changing the initial conditions.

On the other hand, by fixing the phases, the only way to obtain rigorous error estimates would be to analyze a large ensemble of large-volume simulations. Since these are costly, we here resort to a different method, allowing us to obtain error estimates from the likelihood itself: the profile likelihood method [193] provides estimates of confidence intervals for individual parameters of multivariate distributions within a frequentist approach. For a

probability distribution $P(a, \{b_i\})$, the profile likelihood for parameter a is defined as

$$\mathcal{P}^{\rm prof}(a) = \arg\max_{\{b_i\}} [\mathcal{P}(a, \{b_i\})] \,, \tag{7.44}$$

where the additional set of parameters $\{b_i\}$ has been profiled out. Constructing a full profile likelihood for σ_8 is still numerically expensive, since it formally requires a recomputation of the final matter density field each time the function argument is updated. To speed up the analysis, we instead interpolate the profile likelihood evaluated on a predefined grid in σ_8 centered about the fiducial value of the simulation. The details of this procedure will be described in Section 7.5.

7.4 Simulations

All numerical tests presented below are based on two GADGET-2 simulations from the same set of N-body simulations mentioned in Chapter 5. The cosmology and other details of the simulations thus follow the description in Section 5.2. In addition to the N-body runs with the fiducial $\sigma_8 = 0.85$, [140] also presented runs for two additional values of σ_8 bracketing the fiducial one in order to perform the derivative of the halo mass function with respect to σ_8 (for studies related to the scale-dependent bias induced by primordial non-Gaussianity). In this chapter, we further employ those runs, in addition to simulations for 4 additional σ_8 values – which were performed specifically for this work, as well.

The DM halo samples considered here consist of two mass ranges each at two redshifts, and are summarized in Table 7.1. We only consider halos above $10^{13} h^{-1} M_\odot$, corresponding to a minimum of 55 member particles. Note that differences in the number densities of the halo samples imply differences in the expected parameters as well as errors on the inferred σ_8, an important aspect in our validation of the inference framework.

7.5 Implementation

We now provide additional details of the setup and numerical implementation used in our tests. We take as given the halo catalogs described in the previous section, as well as a set

Redshift	Mass range $\log_{10} M_h[h^{-1}M_\odot]$	N_h (run 1)	N_h (run 2)	$\overline{n}_h$ $[(h^{-1}\mathrm{Mpc})^{-3}]$
0	$[13, 13.5]$	2807757	2803575	3.5×10^{-4}
0	$[13.5, 14]$	919856	918460	1.1×10^{-4}
1	$[13, 13.5]$	1507600	1506411	1.9×10^{-4}
1	$[13.5, 14]$	301409	302182	3.8×10^{-5}

Table 7.1: The halo samples used in our numerical tests. Throughout, masses $M \equiv M_{200m}$ are spherical-overdensity masses with respect to 200 times the background matter density.

of matter particles generated either via 2LPT or full N-body for a set of values. Since the matter density field is a function of σ_8, our main parameter of interest, mapping the profile likelihood as a continuous function of σ_8 would require to recompute the set of operators for each function evaluation. To expedite the analysis, we instead generate representations of the evolved matter density field for fixed initial phases and a discrete set of values for σ_8 (as well as redshifts $z = 0$ and $z = 1$ of the halo samples) given by

$$\begin{aligned}\sigma_8 &\in \{0.65,\ 0.75,\ 0.80,\ 0.83,\ 0.85,\ 0.87,\ 0.90,\ 0.95,\ 1.00,\ 1.10,\ 1.20\} \quad \text{(2LPT)}\,,\\ \sigma_8 &\in \{0.68,\ 0.78,\ 0.83,\ 0.85,\ 0.87,\ 0.92,\ 1.02\} \quad \text{(N-body)}\,.\end{aligned} \tag{7.45}$$

For a given halo sample at a given redshift, and a fixed value σ_8^i, the steps for evaluating the σ_8 profile likelihood are as follows:

1. The halos and matter particles are assigned to a 1024^3 grid using a cloud-in-cell density assignment. The high resolution is chosen to avoid leakage of the assignment kernel to the low wavenumbers of interest.

2. A sharp-k filter is applied to the matter and halo fields in Fourier space, such that modes with $|\boldsymbol{k}| > \Lambda$ are set to zero. The Fourier-space grids are subsequently restricted to 384^3, chosen such that the Nyquist frequency of each grid remains above 3Λ for all values of Λ considered here.

3. Quadratic and higher-derivative operators are constructed from the sharp-k filtered matter density field and held in memory. The quadratic operators are renormalized following Appendix J.

4. The maximum of the likelihood over the parameter space spanned by the remaining bias and stochastic variance parameters is then found via function minimizer MINUIT [194] (in practice we minimize $-2\ln\mathcal{L}$, i.e. the pseudo-χ^2). The operator fields do not need to be recomputed for each evaluation, as only their coefficients are varied.

More precisely, we employ the analytic marginalization described in Section 7.3 for the parameters b_2, $c_{\nabla^2\delta}$ and b_{K^2}, leaving only b_1 and the three stochastic amplitudes to be varied in the minimization. We have not found any significant impact of the term $\sigma_{\varepsilon\varepsilon_m,2}$ in Eq. (7.31), but a significant degeneracy with $\sigma_{\varepsilon,2}$. For this reason, we fix the former to zero in our default analysis. This leaves a three-dimensional parameter space to be searched in the minimization, which typically converges quickly. We have found the minimization robust to varying initialization points and number of successive MINUIT cycles. Our default choice for the maximum wavenumber in the likelihood $k_{\rm max}$ is $\Lambda/2$.

This procedure results in a set of values $\{\sigma_8^i,\ -2\ln\mathcal{P}^{\rm prof}(\sigma_8^i)\}_i$ which we find is fit well by a parabola in all cases (we disregard values of σ_8 where the minimization failed to converge). The best-fit value $\hat{\sigma}_8$ is given by the minimum point of the best-fit parabola, while the estimated 1σ error on $\hat{\sigma}_8$ is given by the inverse square-root of the curvature of the parabolic fit. We emphasize that this error does not include any residual cosmic variance, and is essentially purely governed by the halo stochasticity which appears in the variance of the likelihood.

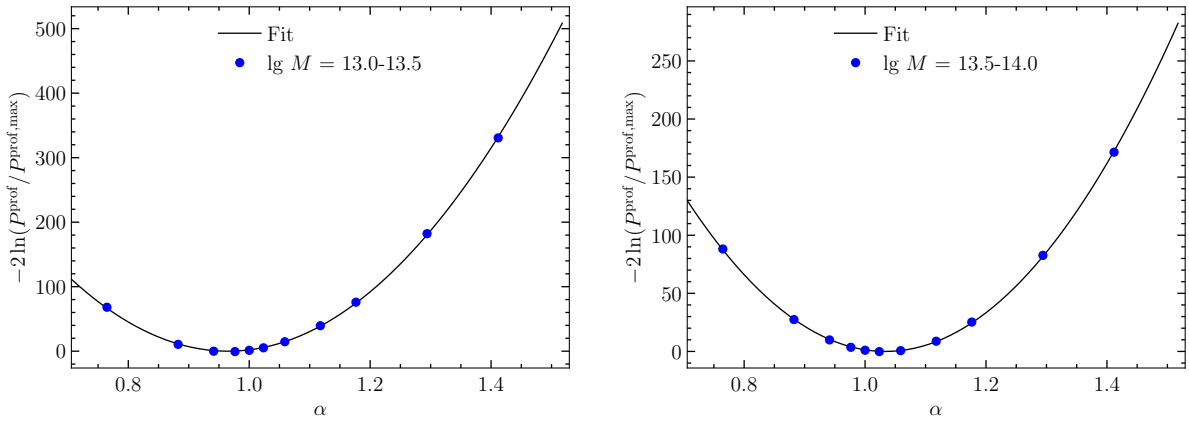

Figure 7.1: Examples of the σ_8 profile likelihood $-2\ln\mathcal{P}^{\rm prof}(\sigma_8)/\mathcal{P}^{\rm prof}(\hat{\sigma}_8)$, plotted as a function of $\alpha \equiv \sigma_8/\sigma_8^{\rm fid}$. In all cases, results are for run 1 and $\Lambda = 0.1\,h\,{\rm Mpc}^{-1}$ at $z = 0$. Also shown are the parabolic fits whose maximum point results in the value $\hat{\sigma}_8$ or equivalently $\hat{\alpha}$ listed in Table 7.2. The value for $-2\ln\mathcal{P}^{\rm prof}(\hat{\sigma}_8)$, which is subtracted for better readibility, is taken from the parabolic fit.

7.6 Results

In the following, we present results for the best-fit value $\hat{\sigma}_8$. For convenience, we phrase these in terms of the ratio to the fiducial value, introducing

$$\hat{\alpha} \equiv \frac{\hat{\sigma}_8}{\sigma_8^{\rm fid}}. \tag{7.46}$$

First, Figure 7.1 shows examples of the profile likelihood determined as described in the previous section, with the parabolic fit that is used to determine $\hat{\sigma}_8$. All panels are for $\Lambda = 0.1\,h\,{\rm Mpc}^{-1}$. Clearly, the log-profile likelihood is well approximated by a parabola, so that we expect maximum point and curvature to yield unbiased estimates of the maximum and 68%-level confidence intervals with respect to a full scan of the profile likelihood. The results for all halo samples and $\Lambda = 0.1\,h\,{\rm Mpc}^{-1}$, our fiducial choice, are summarized in Table 7.2. We find that an unbiased value of $\hat{\sigma}_8$ is recovered to within $\sim 2\sigma$ in most cases. Notice that the run-to-run variance is larger than the estimated error bars in several cases. This could be due to residual cosmic variance, which is not contained in the estimated error bars as discussed in the previous section, although possible issues with the minimizer in isolated cases also cannot be excluded. In order to investigate this, more realizations would be needed.

The remaining columns in Table 7.2 show the value of b_1 as well as the stochastic amplitudes, all corresponding to the maximum-likelihood point for the fiducial value $\sigma_8^{\rm fid}$. Recall that all other parameters are analytically marginalized over. The bias b_1 is essentially fixed by the cross-correlation of δ_h with δ. Correspondingly, we find that the combination $b_1\sigma_8$ is constant for all σ_8 values to within several percent. The stochastic amplitude σ_ε^2

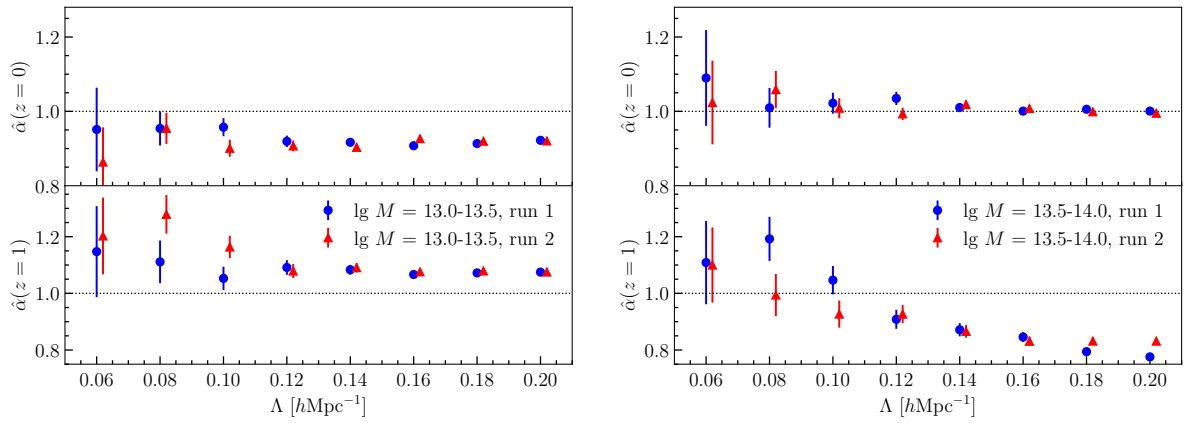

Figure 7.2: Best-fit values $\hat{\alpha} = \hat{\sigma}_8/\sigma_8^{\rm fid}$ as a function of the cutoff Λ (with $k_{\rm max} = \Lambda/2$ in each case) for the low- and intermediate-mass samples in the left and right panels, respectively. In each panel, the upper plot shows results at $z = 0$ while the lower plot shows $z = 1$.

is scaled to the Poisson expectation following Appendix L. Values greater (less) than one thus correspond to super- (sub-)Poisson stochasticity. We do find evidence for a smaller stochasticity for the rarer halo samples, in agreement with previous findings [148]. The last column shows the ratio of the higher-order (in k^2) stochastic parameter to the leading-order one. This gives a rough indication for the spatial length scale squared associated with the scale-dependent stochasticity. We thus find this length scale to be of order $(1-5)\,h^{-1}$Mpc. Notice however that this parameter is expected to also absorb various higher-order contributions not explicitly included in the likelihood, as discussed in [123], so that one cannot robustly infer a physical length scale from this value.

Allowing Λ to vary, we obtain the results shown in Figure 7.2. In each case, we show results for both simulation runs. While the differences in $\hat{\alpha}$ from unity are broadly consistent with being residual stochasticity and cosmic variance for $\Lambda \leq 0.1\,h\,{\rm Mpc}^{-1}$, this no longer holds for higher cutoff values which at $z = 1$ should still be under good perturbative control. On the other hand, the results appear remarkably stable toward higher values of Λ up to $\Lambda = 0.2\,h\,{\rm Mpc}^{-1}$ in many cases. Notice that the majority of the modes contributing to the profile likelihood with $\Lambda = 0.14\,h\,{\rm Mpc}^{-1}$, say, are not included in the likelihood with $\Lambda = 0.1\,h\,{\rm Mpc}^{-1}$, so that these are largely independent measurements.

Before turning to possible explanations for these trends, let us briefly comment on the choice of $k_{\rm max}/\Lambda$. We do not find strong trends with this parameter. Increasing $k_{\rm max}$ at fixed Λ yields similar trends as increasing Λ itself, which is shown in Figure 7.2. For this reason, we fix $k_{\rm max} = \Lambda/2$ throughout.

As a test of the systematics of Type 1, we use the N-body density field itself instead of 2LPT for the construction of the bias operators. The result is shown in Figure 7.3. We only find minor shifts in $\hat{\sigma}_8$. Indeed, the cross-correlation coefficient between the 2LPT and N-body density fields in Fourier space is better than 0.97 for all scales and redshifts considered here, so a large shift would be surprising. Although there is some improvement,

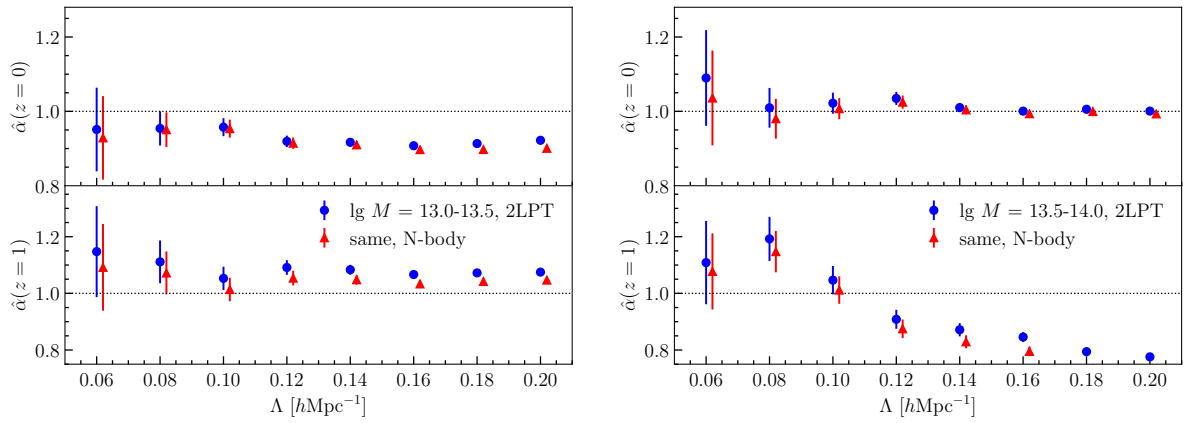

Figure 7.3: Same as Figure 7.2, but using the density field from the N-body simulation. Results for run 1 are shown.

we conclude that the 2LPT density field is not primarily responsible for the bias in $\hat{\sigma}_8$ found.

We next turn to systematics of Type 3, specifically the possible impact of the chosen implementation of $\sigma^2(k)$ in terms of two free parameters (since we fix $\sigma_{\varepsilon\varepsilon_m,2}$ to zero). Since the halo number density is smaller at higher redshift, and the stochasticity correspondingly larger, this could possibly explain the increased bias in $\hat{\sigma}_8$ at $z = 1$ compared to $z = 0$. Performing the profile likelihood analysis with varying $\sigma_{\varepsilon\varepsilon_m,2}$ on the one hand, and both the former parameter and $\sigma_{\varepsilon,2}$ fixed to zero on the other, leads to sub-percent shifts in $\hat{\sigma}_8$ (this can also be gleaned by the values shown in the right-most column of Table 7.2, which, when multiplied by $k^2_{\rm max}$ indicate the upper bound on the fractional contribution of the term $\sigma_{\varepsilon,2}k^2$ to the total variance). Thus, we conclude that the parametrization of $\sigma^2(k)$ is unlikely to be responsible for the bias in $\hat{\sigma}_8$ as well.

This leaves two possible sources of systematic error: our Type-2 systematic, i.e. higher-order terms in the bias expansion (both in perturbations and derivatives); and other systematics of Type-3, namely higher-order corrections to the form of the likelihood itself. Both types of terms are expected to be largest for the rarest and most highly biased halo samples. It is worth nothing that higher-order bias contributions are not necessarily smaller at higher redshift for fixed halo mass, since higher-redshift samples are more biased. Further, higher-derivative terms, which are expected to be tied to the Lagrangian radius of halos, most likely do not decrease toward higher redshift (see for example the results of [195]). Indeed, preliminary investigations show that incorporating higher-order terms in the derivative expansion, such as $(\nabla^2)^2\delta$ and $\nabla^2(\delta^2)$, in the set of operators does have an impact on the profile likelihood. We leave a detailed investigation of the impact of higher-order bias terms to upcoming work.

The remaining Type-3 systematics are expected to be controlled by the ratio in Eq. (7.43); this turns out to be of order unity or larger for the halo samples considered here, implying that higher-order stochastic corrections could be significant. At this point, lacking an

explicit expression for these terms, it is difficult to quantitatively evaluate their impact however.

Additional investigations have pointed to a likely cause for the bias in the inferred value of σ_8 which is related to a higher-order term that is enhanced on large scales. Specifically, the auto-correlation of the quadratic bias operators $A_{OO'}$ contains a contribution from the connected matter four-point function (trispectrum), which, for the likelihood and second-order bias expansion used here, corresponds to an error term $A^{\rm err}_{OO'}$ (Section 7.3.1). For small k, the dominant term from the particular trispectrum configuration involved scales as $\langle\delta^2_\Lambda\rangle^2 P_{\rm L}(k)$, where $P_{\rm L}(k)$ is the linear matter power spectrum. While this contribution is suppressed compared to the leading, disconnected term, the latter approaches a constant at small k, while the trispectrum contribution grows toward small k (assuming $k \gtrsim 0.02\,h\,{\rm Mpc}^{-1}$) due to the factor $P_{\rm L}(k)$. For this reason, it can bias the maximum-likelihood value for σ_8 even for very small values of Λ which correspondingly push k to small values (see the behavior of α for $\Lambda \lesssim 0.1\,h\,{\rm Mpc}^{-1}$ in Figure 7.2). We leave a more detailed investigation, and possible remedies, for future work.

7.7 Discussion and conclusion

We have derived an EFT-based, Fourier-space likelihood for the galaxy density field that allows for cosmological inference from galaxy clustering with rigorously controlled theoretical systematics, without the need for measuring arbitrarily higher order n-point functions. In our concrete application, we have restricted to a second-order bias expansion, including the leading higher-derivative bias contribution as well as scale-dependent stochasticity. While we have not included projection effects such as redshift-space distortions, and thus referred to the tracers as halos, the bias expansion is fully general and also holds for galaxies.

At this order, and when combined with a 2LPT forward model for matter, our posterior self-consistently combines the following sources of cosmological information in large-scale structure:

1. The leading- and next-to-leading order power spectrum, and leading-order bispectrum. In particular, this breaks the bias-amplitude degeneracy, which is perfect at linear order, using the second-order displacement term. Further, it allows for improved constraints on the slope (spectral index) of the linear power spectrum by extending the range in wavenumber k useable for robust constraints.

2. Fully resummed BAO reconstruction using 2LPT displacements, both at the power spectrum and bispectrum levels (see Appendix N for a detailed discussion on this topic).

3. Correct description of curvature and tidal effects on the local BAO scale. This effectively includes information from the 4-point function as well as higher-order statistics, through the highly nontrivial posterior in Eq. (4.6).

These probes translate into constraints on cosmological parameters. The first point allows for direct constraints on σ_8 and parameters which control the shape of the power spectrum, such as Ω_m, n_s, and the sum of neutrino masses. The second and third allow for constraints on the expansion history and thus dark energy equation of state.

We subsequently presented numerical results demonstrating the first application of this EFT-Fourier likelihood for LSS to DM halo catalogs from GADGET-2 gravity-only simulations. More precisely, the test case is to obtain an unbiased estimate of the amplitude of the linear power spectrum σ_8 (or equivalently normalization of scalar perturbations $\mathcal{A}_s$) purely based on the non-linear information in the halo density field that is protected by the equivalence principle. For this, we vary four bias parameters as well as two stochastic amplitudes. The reasoning for choosing σ_8 as parameter is that it can only be inferred by using information in the non-linear density field. An unbiased inference – which we have not completely achieved yet – would thus mean that the robust, protected non-linear information content has been isolated. We expect that other parameters, such as the BAO scale or the matter power spectrum shape, will then also be unbiased, as they rely to a lesser degree on purely non-linear information.

We have additionally presented a method to analytically marginalize over bias parameters, which we apply to three of the four bias parameters in our implementation. We expect that this analytical marginalization will prove extremely powerful when going to higher order in the bias expansion (both in orders of perturbations and derivatives): when using this technique, the cost of finding the maximum-likelihood point, or more generally, sampling from the likelihood, only increases quadratically with the number of bias terms (since the cost is dominated by the evaluation of the matrix $A_{OO'}$).[3] On the other hand, the computational cost would grow much more rapidly if one where to explicitly vary all bias parameters.

Our results indicate that σ_8 can be recovered with a systematic error under $\sim 10\%$ for a range of halo samples at different redshifts when using a cutoff value $\Lambda = 0.1\,h\,\mathrm{Mpc}^{-1}$. The residual systematic bias in σ_8, as shown in [196], can be accounted for by introducing the same Λ cut-off in the initial conditions. It would further be interesting to perform a joint inference from the different halo samples that we have considered separately here, which however requires a generalization of the likelihood to include the stochasticity cross-covariance between different tracers. We leave these developments to future work.

Our approach has some resemblance to what was recently presented in [152]. Instead of the sharp-k filter employed here (necessary for an unbiased inference following [123]), the authors of [192] used a Gaussian filter. More importantly, they allowed for the bias coefficients to be free functions of k, $b_O \to b_O(k)$, which removes the information on σ_8, and instead focused on the degree of cross-correlation of the field we call $\delta_{h,\mathrm{det}}$ here with the actual halo density field. It would be interesting to study the corresponding correlation coefficient for our, sharp-k-filtered field $\delta_{h,\mathrm{det}}$ given the best-fit bias parameters. We defer

[3] This scaling is based on the fact that the computational time of evaluating the likelihood is dominated by the determination of the quantities $A_{OO'}$ and B_O in the notation of Section 7.3. The cost of the matrix inversion is negligible for a realistic number of operators.

this to future study as well.

A natural next question is: what is the expected statistical error for the inferred σ_8 in the realistic case when the phases of the linear density field are unknown? In order to determine this, one unfortunately has to marginalize over those phases, which requires an implementation of the EFT likelihood into a sampling framework along the lines of, e.g. [39, 127]. We thus have to defer this question to future work as well. It is clear however that this uncertainty will be very sensitive to the value of the cutoff Λ. Regardless of the expected statistical precision, it should be pointed out that, by fixing the phases to their ground-truth values throughout, the test of unbiased cosmology inference presented here is the most stringent test possible.

For future improvements, it is worth emphasizing again that an EFT-based likelihood can be straightforwardly extended to include higher-order bias contributions. While this might mean that non-Gaussian corrections to the likelihood and more noise fields need to be considered, the fundamental approach remains the same, and the dimensionality of the inference problem hardly changes. Thus, this approach appears much more feasible than explicitly measuring ever higher n-point functions. Further, many additional types of cosmological physics can be included straightforwardly, such as the scale-dependent bias induced by primordial non-Gaussianity, as well as multiple tracers within the same volume.

For realistic applications to actual data, the main missing ingredients are nontrivial survey geometries (mask) and redshift-space distortions (see Section 3.2.1). The former leads to a non-diagonal noise covariance as discussed in Section 7.1, which however only needs to be determined once. It is then worth mentioning that, in light of a recent results in [197], it is possible to derive an equivalent likelihood in real-space if one allows for the dependence of the stochasticity covariance matrix on the evolved matter field.

Redshift-space distortions can be treated in the EFT approach as well [198, 199, 200], and can be included in forward modeling correspondingly.[4] We leave all of these developments to future work.

[4]Redshift-space distortions allow for a measurement of the growth rate in the linear regime. However, if line-of-sight-dependent selection effects are present, then this information is removed due to a degeneracy with a new bias parameter. Then, non-linear information is necessary to infer the growth rate [200], in close analogy with the discussion in Section 7.1.2 here.

Redshift	Mass range $\log_{10} M[h^{-1}M_\odot]$	$\hat{\alpha}$ (run 1)	$\hat{\alpha}$ (run 2)	b_1	σ^2_ε [Poisson]	$\sigma_{\varepsilon,2}/\sigma_\varepsilon$ $[(h^{-1}\mathrm{Mpc})^2]$
0	[13.0-13.5]	0.96 ± 0.02	0.90 ± 0.02	1.20	1.11	-25.5
0	[13.5-14.0]	1.02 ± 0.03	1.01 ± 0.03	1.61	0.96	-11.7
1	[13.0-13.5]	1.05 ± 0.04	1.16 ± 0.04	2.36	0.93	1.3
1	[13.5-14.0]	1.05 ± 0.05	0.93 ± 0.05	3.49	0.89	10.5

Table 7.2: Summary of results for $\Lambda = 0.1\,h\,\mathrm{Mpc}^{-1}$ and $k_{\max} = \Lambda/2$, with the likelihood settings described in the text. For the best-fit scaled σ_8 estimate $\hat{\alpha}$, results from run 1 and run 2 are shown individually with estimated 68% CL error bars. b_1 and stochastic amplitudes are reported for the fiducial $\sigma_8 = \sigma_8^{\mathrm{fid}}$ and averaged over both runs. The stochastic variance σ^2_ε is scaled to the Poisson expectation for the given halo sample, as described in Appendix L. The last column shows the ratio of the higher-derivative stochastic amplitude to the lowest-order one, indicating the scale associated with the expansion of $\sigma^2(k)$ in k.

Chapter 8

Conclusion and outlook

This thesis summarizes our attempts to improve and apply the Bayesian forward modeling and inference approach – motivated by the BORG framework – to galaxy clustering data, in order to robustly extract *unbiased* cosmological and astrophysical information.

On the side of cosmological inference, we have put this framework to some of the most stringent tests up to now. These tests were designed to study the precision and robustness of the main ingredients of Bayesian forward modeling, in particular, the deterministic bias model and the LSS likelihood. Specifically, in Chapter 5 we have implemented a non-local, second-order bias model in BORG and in Chapter 7, we have developed and rigorously tested a novel likelihood derived from EFT approach, in Fourier space. As pointed out at the end of Chapter 7, there is still a substantial amount of work required to ready our framework for cosmological parameter inference from real galaxy surveys. In fact, many works in this direction have been already carried out before and during the writing of this thesis [187, 197, 196]. The path, however, is clear. The EFT approach and Bayesian forward modeling framework together allow for a straightforward extension of galaxy bias modeling to any higher-order. Indeed, [196] has already extended the bias model studied in Chapter 7 to third-order. On another hand, [197] pointed out that the EFT Gaussian likelihood can also be implemented directly in real-space – with some modifications compared to the version presented in Chapter 7 – thus avoiding potential problems arising from non-trivial survey masks. For the next steps, we aim to implement and study the EFT-based likelihood in the BORG framework – where we are going to sample also the phases of initial density field – with the goal to recover the input cosmological parameters of the N-body simulations within the uncertainty provided by BORG ensemble. It would also be interesting to see how the rigorous EFT approach fares against the ad-hoc models studied in Chapter 5 (see also [127]), or a neural network approach as in [201].

On the side of astrophysical measurement, using BORG-SDSS3 reconstructions, we have demonstrated that our approach provides a consistent way to account for systematics induced by reconstructed velocity field in kSZ measurements. This is of course only the first step towards a completely *systematic-free* measurement of the kSZ signal, especially given that future CMB and redshift data will be highly more precise and any measurement will be more sensitive to residual systematics. However, precisely for this reason, it is

important to develop a fully Bayesian framework – like the one presented in Chapter 6 – that facilitates the combination of various datasets in a consistent and transparent way. Only by doing so, we could exploit the kSZ probes to their full potential in future. Indeed, using kSZ measurement to probe galaxy cluster gas profile and constrain baryonic fraction inside galaxy clusters and groups is another fascinating topic that we plan to explore as soon as higher-resolution CMB data allows us to resolve more clusters, especially the ones at the low-mass end.

Appendix A

Fourier transform

A.1 Dirac Delta function

With the notation of Fourier transform given in Table 1.2, we define the Dirac Delta function by demanding that the application of the inverse Fourier transform onto the Fourier transform of $f(\boldsymbol{x})$ must yield the exact same function

$$\begin{aligned} f(\boldsymbol{x}) &= \int_{\boldsymbol{k}} \left(\int_{\boldsymbol{x}'} f(\boldsymbol{x}') e^{-i\boldsymbol{k}\cdot\boldsymbol{x}'} \right) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \\ &= \int_{\boldsymbol{x}'} \left(\int_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \right) f(\boldsymbol{x}'). \end{aligned} \tag{A.1}$$

Compare Eq. (A.1) to the translation properties of the Dirac Delta function, i.e.

$$f(\boldsymbol{x}) = \int_{\boldsymbol{x}'} \delta_D(\boldsymbol{x}-\boldsymbol{x}') f(\boldsymbol{x}'), \tag{A.2}$$

we arrive at the definition

$$\delta_D(\boldsymbol{x}-\boldsymbol{x}') = \int_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}. \tag{A.3}$$

A.2 Filtering and convolution theorem

The application of any given filter $W(\boldsymbol{x})$, e.g. the AP filter in Chapter 6 or the sharp, low-pass filter in Chapter 7 onto a field $f(\boldsymbol{x})$, yields a filtered field $f_W(\boldsymbol{x})$

$$f_W(\boldsymbol{x}) \equiv (W * f)(\boldsymbol{x}) = \int_{\boldsymbol{x}'} W(\boldsymbol{x}') f(\boldsymbol{x}-\boldsymbol{x}'). \tag{A.4}$$

This operation could be represented as a multiplication in Fourier space where

$$\begin{aligned}
f_W(\boldsymbol{k}) &= \int_{\boldsymbol{x}} \left[\int_{\boldsymbol{x}'} W(\boldsymbol{x}') f(\boldsymbol{x}-\boldsymbol{x}') \right] e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \\
&= \int_{\boldsymbol{x}} \left[\int_{\boldsymbol{x}'} \left(\int_{\boldsymbol{x}''} W(\boldsymbol{x}') f(\boldsymbol{x}'') \delta_D(\boldsymbol{x}-\boldsymbol{x}'-\boldsymbol{x}'') \right) \right] e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \\
&= \int_{\boldsymbol{x}} \left[\int_{\boldsymbol{x}'} \left(\int_{\boldsymbol{x}''} W(\boldsymbol{x}') f(\boldsymbol{x}'') \int_{\boldsymbol{k}'} e^{i\boldsymbol{k}'\cdot(\boldsymbol{x}-\boldsymbol{x}'-\boldsymbol{x}'')} \right) \right] e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \\
&= \int_{\boldsymbol{k}'} \left[\left(\int_{\boldsymbol{x}} e^{i\boldsymbol{x}\cdot(\boldsymbol{k}'-\boldsymbol{k})} \right) \left(\int_{\boldsymbol{x}'} W(\boldsymbol{x}') e^{-i\boldsymbol{k}'\cdot\boldsymbol{x}'} \right) \left(\int_{\boldsymbol{x}''} f(\boldsymbol{x}'') e^{-i\boldsymbol{k}'\cdot\boldsymbol{x}''} \right) \right] \\
&= \int_{\boldsymbol{k}'} \left[\delta_D(\boldsymbol{k}'-\boldsymbol{k})\, W(\boldsymbol{k}')\, f(\boldsymbol{k}') \right] \\
&= W(\boldsymbol{k})\, f(\boldsymbol{k}).
\end{aligned} \tag{A.5}$$

We employ this property (together with the Fourier transform and its inverse) frequently throughout our work, whenever a computation involving filtered fields is encountered.

Below, we list the four symmetric 3D filters $W_R(x)$ commonly appeared in the works described by this thesis, and their Fourier transforms $W_R(k)$,

- Spherical sharp-k filter:

$$W_R(x) = \frac{3}{4\pi R^3} \left[\frac{3 j_1(x/R)}{(x/R} \right], \tag{A.6}$$

$$W_R(k) = \Theta_{\mathrm{H}}(1-kR). \tag{A.7}$$

- Spherical Gaussian filter:

$$W_R(x) = \frac{1}{(2\pi R^2)^{3/2}}\, e^{(-x^2/2R^2)}, \tag{A.8}$$

$$W_R(k) = e^{(-k^2R^2)}, \tag{A.9}$$

- Spherical top-hat filter:

$$W_R(x) = \frac{3}{4\pi R^3} H\left(1-\frac{x}{R}\right), \tag{A.10}$$

$$W_R(k) = 3\frac{j_1(kR)}{kR}, \tag{A.11}$$

- CIC filter:

$$W_R(x) = \frac{1}{R} \begin{cases} 1-|x|/R & |x|<R \\ 0 & \text{otherwise,} \end{cases} \tag{A.12}$$

$$W_R(k) = \left[\frac{\sin\left(\frac{\pi k}{2k_N}\right)}{\frac{\pi k}{2k_N}} \right]^2 , \tag{A.13}$$

where $\Theta_{\mathrm{H}}(x)$ denotes the Heaviside step function

$$\Theta_{\mathrm{H}}(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{A.14}$$

and $j_1(x)$ denotes the first-order spherical Bessel function

$$j_1(x) = \frac{\sin(x) - x\cos(x)}{x^2}, \tag{A.15}$$

while $k_N = \pi/R$ denotes the Nyquist frequency.

A.3 Numerical implementations of Fourier transform

For numerical computation, a discrete version of the Fourier transform (DFT) needs to be implemented. The DFTs in BORG are carried out with the help of the FFTW package[1]. For the processing of BORG output, we employ the Numpy's implementation of DFT[2].

[1] `http://www.fftw.org/`

[2] `https://docs.scipy.org/doc/numpy/reference/routines.fft.html`

Appendix B

Hamiltonian Monte-Carlo sampling

This appendix aims to provide only the background information to support the main text, readers interested in the HMC method might find a more in-depth review in [202]. An insightful review of inference using MCMC methods could be found in [203]. Below we will assume some prior knowledge on basic MCMC methods, for example, the Metropolis-Hasting algorithm. Readers who are new to this topic might find Part IV, and in particular, Section 29-30 of [**?**] helpful.

B.1 Overview

Hamiltonian Monte Carlo (HMC) – also known as Hybrid Monte Carlo as it was first introduced in [128] – is a MCMC method that improves upon the original Metropolis-Hasting algorithm [204, 130] by introducing a dynamical method to find and propose a new state during Metropolis updates, that is, by following the the Hamiltonian equations of motion:

$$\begin{aligned} \frac{d\vec{q}}{dt} &= \frac{\partial H(\vec{q},\vec{p})}{\partial \vec{p}}, \\ \frac{d\vec{p}}{dt} &= -\frac{\partial H(\vec{q},\vec{p})}{\partial \vec{q}}, \end{aligned} \tag{B.1}$$

where

$$H(\vec{q},\vec{p}) = U(\vec{q}) + K(\vec{p}) \tag{B.2}$$

for a system with potential $U(\vec{q})$ (see Section B.3.1) and kinetic energy $K(\vec{p})$ (see Section B.3.2). The reason behind this choice will become clear in the following sections, especially in Section B.2.1 and Section B.3.1.

The unoriginal name "Hamiltonian Monte Carlo" caught on not only because it retains the abbreviation while being more descriptive but also because it highlights the beauty of this method. Conservation of the *Hamiltonian* guarantees that a sampler following the trajectory of Hamiltonian dynamics – as opposed to that of a *random walk* like in the original Metropolis-Hasting algorithm – not only systematically explores a bigger volume

of the parameter space but also doing that with a high acceptance rate, even in the presence of numerical errors due to discretization of time in Eq. (B.1) in computer implementations.

B.2 Properties of Hamiltonian dynamics

Here we highlight some general properties of Hamiltonian dynamics which constitute to its success in MCMC application.

B.2.1 Hamiltonian conservation

Perhaps the most important feature for the success of HMC is the fact that the Hamiltonian Eq. (B.2) is *invariant* under time translation

$$\frac{dH}{dt} = \frac{\partial H}{\partial \vec{q}} \cdot \frac{d\vec{q}}{dt} + \frac{\partial H}{\partial \vec{p}} \cdot \frac{d\vec{p}}{dt} = \frac{\partial H}{\partial \vec{q}} \cdot \frac{\partial H}{\partial \vec{p}} + \frac{\partial H}{\partial \vec{p}} \cdot \left(-\frac{\partial H}{\partial \vec{q}} \right) = 0. \tag{B.3}$$

This means, technically, H is conserved along the trajectory of Hamiltonian dynamics, thus proposals found through this dynamical transition always have a unity acceptance rate during Metropolis updates. Practically, any HMC implementation requires numerically solving the equations of Hamiltonian dynamics Eq. (B.1). This implies discretization of time in Eq. (B.1) into small step sizes ϵ that introduces numerical errors which in turn results into H being only asymptotically invariant (in the limit of vanishing ϵ).

B.2.2 Reversibility

Any MCMC update must leave the target distribution invariant. This requirement is easily satisfied if the Markov chain transition is *reversible*, which is the case for the Hamiltonian dynamics, for if $t \to -t$ and $\vec{p} \to -\vec{p}$ – assuming Eq. (B.1) and $H(\vec{q}, -\vec{p}) = H(\vec{q}, \vec{p})$ – then

$$\begin{aligned} \frac{d\vec{q}}{d(-t)} &= -\frac{d\vec{q}}{dt} = -\frac{\partial H(\vec{q}, \vec{p})}{\partial(\vec{p})} = \frac{\partial H(\vec{q}, -\vec{p})}{\partial(-\vec{p})} \\ \frac{d(-\vec{p})}{d(-t)} &= \frac{d\vec{p}}{dt} = -\frac{\partial H(\vec{q}, \vec{p})}{\partial \vec{q}} = -\frac{\partial H(\vec{q}, -\vec{p})}{\partial \vec{q}} \end{aligned} \tag{B.4}$$

which assumes the same form of Eq. (B.1).

B.2.3 Volume preservation

The fact that the divergence of the *phase-space* velocity field of a Hamiltonian system vanishes

$$\frac{\partial}{\partial \vec{q}} \cdot \frac{d\vec{q}}{dt} + \frac{\partial}{\partial \vec{p}} \cdot \frac{d\vec{p}}{dt} = \frac{\partial}{\partial \vec{q}} \cdot \frac{\partial H}{\partial \vec{p}} + \frac{\partial}{\partial \vec{p}} \cdot \left(-\frac{\partial H}{\partial \vec{q}} \right) = 0 \tag{B.5}$$

can be used to show that the Hamiltonian flow reserves the phase-space volume [202]. This result is a direct consequence of Hamiltonian dynamics being *symplectic*, i.e. Hamiltonian

flow leaves the symplectic manifold of phase-space intact, as stated by the Liouville's theorem [205]. This property makes Hamiltonian dynamics even more suitable to find and propose a new state for Metropolis updates, as one needs not to account for the change in volume when computing the acceptance probability.

B.3 Anatomy of the Hamiltonian in HMC

B.3.1 Potential energy

The *potential* term $U(\vec{q})$ in the Hamiltonian Eq. (B.2) – inspired by the connection between the probability distribution of a canonical ensemble consisting of microstates $(\vec{q}, \vec{p})$ in statistical mechanics

$$P(\vec{q}, \vec{p}) = \frac{1}{Z} \exp\left[-H(\vec{q}, \vec{p})/T\right] = \frac{1}{Z} \exp\left[-U(\vec{q})/T\right] \exp\left[-K(\vec{p})/T\right] \tag{B.6}$$

and the posterior of model parameter $\vec{\theta}$ in statistical inference

$$P(\vec{\theta}\,|\vec{d}) = \frac{P(\vec{\theta}) \prod_i P(\vec{d}\,|\vec{\theta})}{P(\vec{d})} = \frac{1}{P(\vec{d})} \exp\left[\ln\left(P(\vec{\theta}) \prod_i P(\vec{d}\,|\vec{\theta})\right)\right], \tag{B.7}$$

for systems with a unity temperature, i.e., $T = 1$, and a partition function equal to the model evidence, i.e. $Z = P(\vec{d})$, – is defined in terms of our target distribution, i.e. the probability function we wish to sample [128, 203]

$$U(\vec{q}) = -\ln\left(P(\vec{q}) \prod_i P(\vec{d}\,|\vec{q})\right) \tag{B.8}$$

wherein the model parameters play the role of "position" variables, $\vec{q} = \vec{\theta}$. Note that this is fully consistent with setting $T = 1$ and $Z = P(\vec{d})$, as, by definition,

$$Z = \int\int e^{-H(\vec{q},\vec{p})} d\vec{q} d\vec{p} = \int\int \left(P(\vec{q}, \vec{p}) \prod_i P(\vec{d}\,|\vec{q}, \vec{p})\right) d\vec{q} d\vec{p} = P(\vec{d}). \tag{B.9}$$

So far, we have treated the "momentum" variables $\vec{p}$ as nuisance parameters to be marginalized over. This is indeed the case in all practices of HMC sampling, as the $\vec{p}$ and the associated kinetic term $K(\vec{p})$ are only introduced so that the Hamiltonian dynamics (cf. Eq. (B.1)) can operate properly. The exact form of $K(\vec{p})$, however, can affect the efficiency of the HMC sampler, as will be discussed in the following section.

B.3.2 Kinetic energy

The *kinetic* term $K(\vec{p})$ of the Hamiltonian the Hamiltonian Eq. (B.2) consists of auxiliary, i.e. *nuisance*, "momentum" variables $\vec{p}$, which are typically drawn from *independent* Gaussian distributions in each iteration of the HMC algorithm, and a *positive-definite* "mass

matrix" or "Hamiltonian mass" $\mathcal{M}$

$$K(\vec{p}) = \frac{1}{2}(\vec{p})^{\intercal} \cdot \mathcal{M}^{-1} \cdot \vec{p}, \tag{B.10}$$

which is normally chosen to be *diagonal*, for the sake of computing its inversion. In essence, the kinetic term, specifically the "Hamiltonian mass" $\mathcal{M}$, determines the inertia of our exploration of the parameter space with the HMC sampler, of which we wish to maximize the efficiency. A too-large "mass" choice would result into a slow moving sampler while a too-small "mass" choice would lead to instability of the numerical integrator and thus non-conserved H – that is, low acceptance rate.

A typical choice of $\mathcal{M}$ is setting $\mathcal{M} = \mathcal{C}^{-1}$ where $\mathcal{C}$ is the covariance matrix of $\vec{q}$ [206, 207]. For practical purposes, especially in cases where the dimensionality of the parameter space is so high that it is infeasible to store and invert $\mathcal{M}$, an analysis of the stability of the numerical integrator, given a specific form of the potential energy $U(\vec{q})$, must be performed in order to identify the optimal choice for $\mathcal{M}$ [207, 40, 39].

B.4 HMC algorithm

Each iteration of a typical HMC algorithm can be generally divided into two steps.

1) The momentum variables are updated, $\vec{p} \to \vec{p}'$, using new values drawn from their multivariate Gaussian distribution with covariance matrix $\mathcal{M}$.

1) From $(\vec{q}, \vec{p}')$, Hamiltonian dynamics is simulated for n_{step} steps of size ϵ, using a symplectic integrator, e.g. the leapfrog integrator described in the next section.

The algorithm essentially alternates between a random walk across Hamiltonian energy levels and deterministic Hamiltonian trajectory along each level.

B.4.1 Leapfrog scheme

Solving the Hamiltonian dynamics on a computer amounts to discretizing the system of continuous differential equations in Eq. (B.1) into small step sizes ϵ and evolving them for $\tau = n_{\text{step}}\,\epsilon$ at each step. A popular choice to approximate the solution for such a system is the *leapfrog* integration scheme[1] [202, 208, 39]

$$\vec{p}\left(t + \frac{\epsilon}{2}\right) = \vec{p}(t) - \frac{\epsilon}{2}\frac{\partial U}{\partial \vec{q}}\vec{q}(t) \tag{B.11}$$

$$\vec{q}(t + \epsilon) = \vec{q}(t) + \epsilon\,\frac{\vec{p}(t + \epsilon/2)}{\vec{m}} \tag{B.12}$$

$$\vec{p}(t + \epsilon) = \vec{p}\left(t + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2}\frac{\partial U}{\partial \vec{q}}\,\vec{q}(t + \epsilon) \tag{B.13}$$

[1] This particular form of the leapfrog integrator is referred to as the "kick-drift-kick" form.

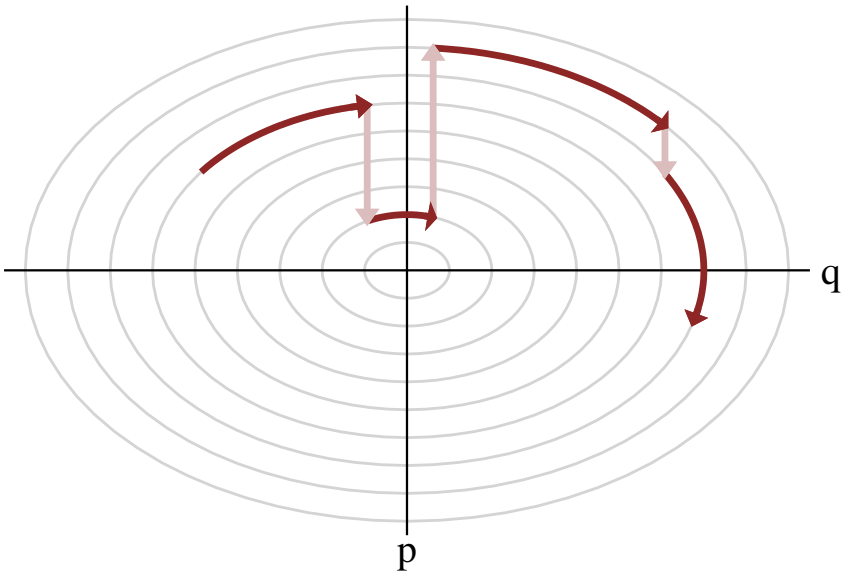

Figure B.1: Each HMC transition first lifts the current state onto a random Hamiltonian energy level (light red arrows) by (re-)sampling the momentum. In the next step, the system traverses the corresponding Hamiltonian trajectory for $L = n_{\text{step}}\,\epsilon$ (dark red arrows) before arriving at a new state which is then proposed to the Metropolis update. Diagram taken from [208].

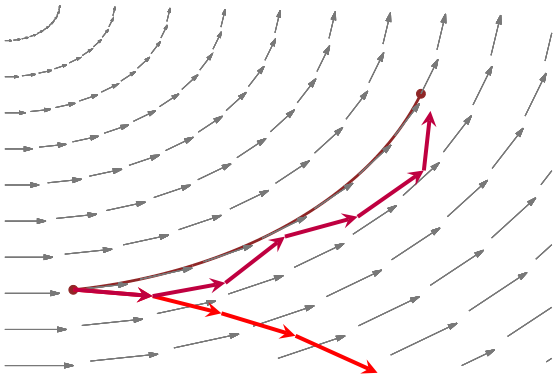

Figure B.2: With most numerical integrators, as the system evolves further and further from the initial point, numerical errors coherently add up and push the numerical trajectory (red line) away from the exact Hamiltonian trajectory. The family of symplectic integrators, including the leapfrog scheme described by Eqs. (B.11)–(B.12), generate an incompressible numerical trajectory, similar to the exact Hamiltonian trajectory they approximate. Consequently their numerical errors cancel out and the numerical trajectory (dark violet) oscillate around the exact one.

assuming a diagonal mass matrix $\mathcal{M}$ with $\vec{m} = \operatorname{diag} \mathcal{M}$.

Note that Eqs. (B.11)–(B.11) are all shear transformations. Consequently, the leapfrog scheme strictly preserves the phase-space volume. This incompressibility limits the deviation of the numerical trajectory from the exact Hamiltonian trajectory (which is also incompressible) such that the numerical trajectory keeps oscillate around the exact one. Additionally, Eqs. (B.11)–(B.11) are reversible by negating $\vec{p}$, applying the same number of steps, then negating $\vec{p}$ again.

The HMC system of BORG – as described in Chapter 4 – is evolved using the leapfrog integrator in the exact same form as Eqs. (B.11)–(B.13) wherein $\vec{q} \equiv \vec{d}_{\mathrm{in}}$ and $U(\vec{q}) \equiv \Psi(\vec{\delta}_{\mathrm{in}}) = -\ln \mathcal{P}\left(\vec{\delta}_{\mathrm{in}} \middle| \vec{d}_h, \mathcal{S}\right)$.

B.4.2 Step size and trajectory length

The leapfrog step size ϵ and number of leapfrog steps n_{step} taken during each Metropolis update together control the length of the Hamiltonian trajectory $L = n_{\mathrm{step}}\, \epsilon$, which in turn plays a crucial role in determining the performance of HMC sampling, i.e. the convergence rate of the MCMC chain and correlation between neighbor MCMC samples.

On one hand, if either ϵ or n_{step} is too large, L will be too long and the sampler might just end up near its starting point. Additionally, a larger step size ϵ introduces more numerical errors (hence instabilities) in the approximated Hamiltonian trajectory (see Figure B.3), which then typically leads to a low acceptance rate for states proposed by said trajectory.

On the contrary, choosing a too small value for either ϵ or n_{step} not only waste compu-

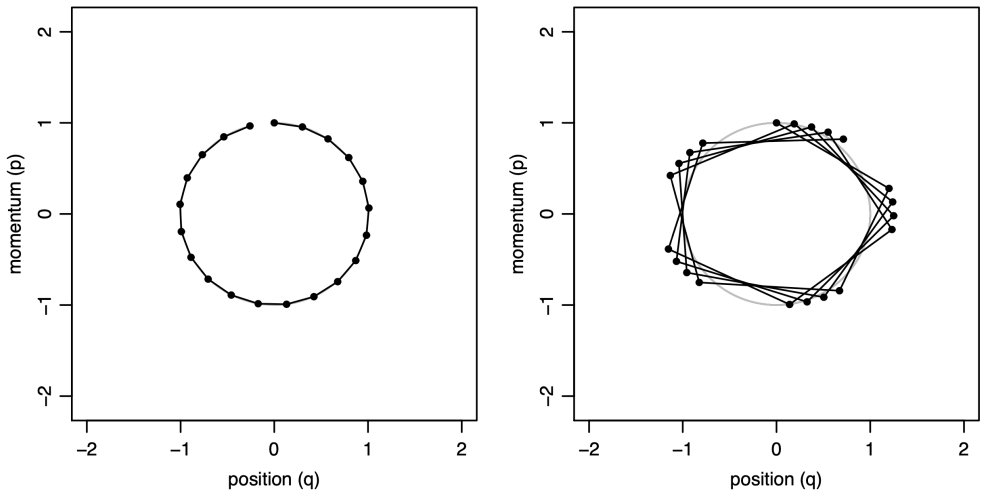

Figure B.3: The numerical trajectory computed with the leapfrog scheme using step sizes of $\epsilon = 0.3$ (left panel) and $\epsilon = 1.2$ (right panel) for the 1D Hamiltonian $H = q^2/2 + p^2/2$. With increasing ϵ, the difference between exact Hamiltonian trajectory and leapfrog-integrated trajectory becomes more and more visible. The latter, however, remains stable, only oscillating around the former. Plot taken from [202].

tational time but also potentially delay convergence of the resulting MCMC chain[2].

Finally, to avoid trajectory resonant, namely, the HMC sampler is trapped in some subsets of parameter space due to periodicity of the Hamiltonian [202], ϵ and n_{step} are randomly drawn from uniform distributions $[\epsilon_{\text{min}}], \epsilon_{\text{max}}]$ and $[n_{\text{min}}, n_{\text{max}}]$, respectively, at each step [39, 202].

B.4.3 Convergence and correlation

As mentioned in the previous section, even with the HMC, tuning of ϵ and n_{max} is still needed to boost the convergence rate of the MCMC chain. This step typically requires some preliminary runs during which the values of ϵ and n_{max} are varied while the evolutions of some performance-/convergence-indicative quantities are monitored with trace plots.

Another important aspect for any MCMC analysis is how to store and process samples in a chain in the most efficient way. In our analysis, we typically *thin* the MCMC chain by only storing and using 1-in-every-10 or 1-in-every-40 samples.

Below, in Appendix D, we will discuss these points in the specific context of BORG analysis.

[2]As the trajectory length $n_{\text{step}}\,\epsilon$ becomes too short, the HMC trajectory becomes similar to that of a random walk which is precisely what we try to avoid with HMC.

B.4.4 Scaling with dimensionality

The reason that HMC sampling is the crux of BORG is its scaling with dimensionality. Here we briefly summarize the discussions in [209, 202] without providing a detailed proof. We restrict our discussion to the sampling performance after the burn-in phase, i.e. when equilibrium has been reached.

Consider the problem of sampling in a d-dimensional parameter space, the key to quantify the performance of a MCMC sampling algorithm is the minimal amount of iterations or updates needed for it to reach a (nearly) independent point from the previous point, meanwhile, maintaining a reasonable acceptance rate. For the traditional random-walk Metropolis-Hasting algorithm, this number is simply proportional to d, which results into a total numerical cost that grows as $d\,d = d^2$. On the other hand, the HMC algorithm only requires a number proportional to $d^{1/4}$ of leapfrog updates to reach the next independent point while keeping the dimensional-average total error in H, and hence the acceptance rate, in check. This leads to a total numerical cost growing as $d\,d^{1/4} = d^{5/4}$. Let us consider the typical cases that BORG has to deal with wherein $d \sim 128^3 - 512^3$, the HMC sampling suppresses the numerical cost by a factor of $d^{3/4} \sim 38 - 64$.

It is worth emphasizing that the general discussion above only requires the acceptance rates of the two algorithms being reasonable, not necessarily the same. It can be rigorously shown that, when the computational cost to obtain an independent sample is minimized for each algorithm, an optimal rate of 23% and 65% can be achieved for Metropolis-Hasting and HMC algorithm, respectively (see [202] and references therein).

B.5 Advanced HMC methods

More advanced variations and extensions of the standard HMC algorithm are discussed, although not exhaustively, in [202], including adaptive HMC in which new step size is chosen for each HMC update based on results of previous updates, and trajectory tempering wherein the temperature T in Eq. (B.6) is varied[3] so that the HMC sampler can move more easily between distribution modes that are separated by regions with low probability.

[3] A higher T produces a more diffuse distribution, with the target distribution corresponding to $T = 1$.

Appendix C

Slice sampling

The *slice sampling* method [129] – employed for the sampling of bias parameters in BORG – exploits the idea that one can sample from a target distribution of a random variable $x \in \mathbb{R}^n$ by uniformly drawing samples from the $n+1$-dimensional region under the plot of its density function. One can then construct a Markov chain that converges to this uniform distribution by uniformly sample alternatively between the vertical and horizontal directions, hence the name. Let us briefly formulate this idea in an algorithmic format. For simplicity, we only consider here the case of $n = 1$. To sample from a multivariate distribution wherein $n > 1$, as in the case of the bias parameters b_O and likelihood parameters λ_a, such single-variable slice sampling updates can be sequentially applied to each variable while the other variables are fixed.

Suppose we want to sample x from $\mathcal{P}(x) \propto \mathcal{P}^*(x)$, we need to introduce an auxiliary variable y such that a transition of $x_0 \to x_1$ involves:

1. evaluate $\mathcal{P}^*(x_0)$;
2. $y \curvearrowleft \mathcal{U}(0, \mathcal{P}^*(x_0))$;
3. define a horizontal slice (x_l, x_r);
4. loop over step 5-8;
5. $x_1 \curvearrowleft \mathcal{U}(x_l, x_r)$;
6. evaluate $\mathcal{P}^*(x_1)$;
7. if $(\mathcal{P}^*(x_1) > y)$ then break from loop;
8. else modify (x_l, x_r).

Here $\mathcal{U}$ denotes a uniform distribution. Specifically, variations of slice sampling method differ by the details in step 3 and 8. The general requirement for any flavor is that it must satisfy detail balance so that it leaves the distribution $\mathcal{U}(x, y)$ under the curve $\mathcal{P}^*(x)$ invariant [129]. Below we illustrate step 3 and step 8 with the two procedures actually implemented in BORG, namely the "doubling" and "shrinkage" procedures.

C.1 "Doubling" procedure for defining (x_l, x_r)

We keep doubling the interval (x_l, x_r) until x_l and x_r satisfy $\mathcal{P}^*(x_l) < y$ and $\mathcal{P}^*(x_r) < y$, i.e. both left and right endpoints fall outside the slice:

1. $r \curvearrowleft \mathcal{U}(0,1)$;
2. $x_l = x - rw$, $x_r = x_l + w$;
3. while $(\mathcal{P}^*(x_l) > y)$ or $(\mathcal{P}^*(x_r) > y)$: $x_l = x_l - w$ or $x_r = x_r + w$.

The purpose of this step is to get an interval containing as much of the the slice as possible such that the new point x_1 would differ from x_0 as much as possible [129].

C.2 "Shrinkage" procedure for shrinking (x_l, x_r)

Whenever a x_1 drawn in step 5 leads to the (x_1, y) point lies above the $\mathcal{P}^*(x_0)$ curve, it becomes a new endpoint. So we effectively shrink the interval (x_l, x_r) such that x_0 is still enclosed in the new interval:

1. if $(x_1 > x_0)$ then $x_r = x_1$;
2. else $x_l = x_1$.

This is done to avoid a too large interval which would render subsequent sampling step inefficient [129].

Appendix D

BORG MCMC chain: burn-in phase and thinning

A MCMC chain generally needs a burn-in phase before it starts to converge to the target posterior distribution (see Section B.4.3). In general, a good indicator for the convergence of the chain generated by BORG is the convergence plot of the primordial power spectrum $P_{\rm ini}(k)$, of which we show an example in Figure D.1. For all BORG reconstructions shown below, we remove all MCMC samples generated during this burn-in phase.

Measuring the correlation between noise residuals in MCMC samples often offers a qualitative estimate on how much information is there in the chain, or on what usually known as the *effective* sample size. To do this, we first subtract the BORG ensemble mean from the inferred initial density field in each BORG sample, i.e.

$$\vec{\delta}^s_{\rm res} = \vec{\delta}^s_{m,\rm ini} - \left\langle \vec{\delta}_{m,\rm ini} \right\rangle_{s'} . \tag{D.1}$$

It is worth repeating here that $\langle\,\rangle_{s'}$ denotes the BORG ensemble average. We then measure the correlation of this residual for successive MCMC samples as

$$r_{\rm res}(s, s') = \frac{\mathrm{Cov}\left(\vec{\delta}^s_{\rm res},\, \vec{\delta}^{s'}_{\rm res}\right)}{\sigma_s \sigma_{s'}} \tag{D.2}$$

in which s and s' denote the MCMC identifiers and all implicit averages in Eq. (D.2) are now taken over the spatial grid. By definition, $r_{\rm res}$ varies between -1 and 1, with 0 implies the residuals are totally uncorrelated. Thus, for our purpose, we would select only samples separated by a distance of $\Delta s = s' - s$ such that $r_{\rm res}(\Delta s) = 0$. We refer to Δs as the *thinning factor* of the MCMC chain. Our measurements of $r_{\rm res}$ for all reconstructions included in this analysis indicate that, once the chains have passed their burn-in phases, our effective sample size is $\sim N_{\rm samples}/40$, regardless of the setup (e.g. tracer density, grid resolution, etc.). We show in the right panel of Figure D.2, $r_{\rm res}$ for four of the reconstructions.

For each reconstruction, after accounting for burn-in phase and thinning the rest of the chain by a factor of 40, we include 751 BORG samples in the analysis.

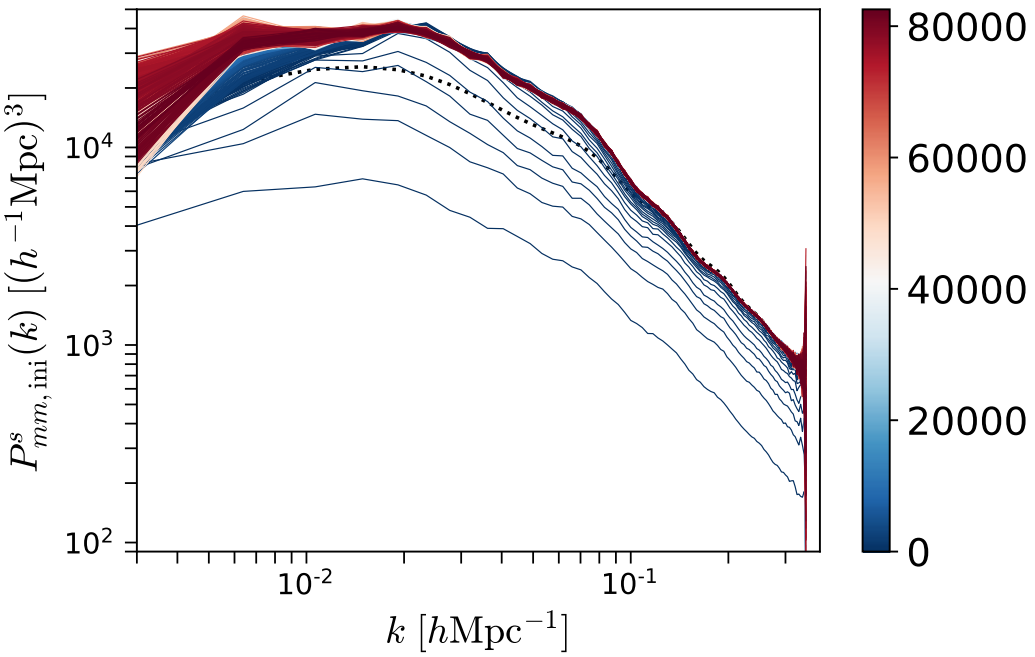

Figure D.1: An example of the convergence plot of the initial matter power spectrum in one of our BORG reconstructions. The color bar indicates the MCMC identifier. The plot illustrates the burn-in phase (deep blue) in which the power spectrum drifts towards the theoretical initial power spectrum specified by the given cosmology. This specific chain appears to converge after ~ 800 samples.

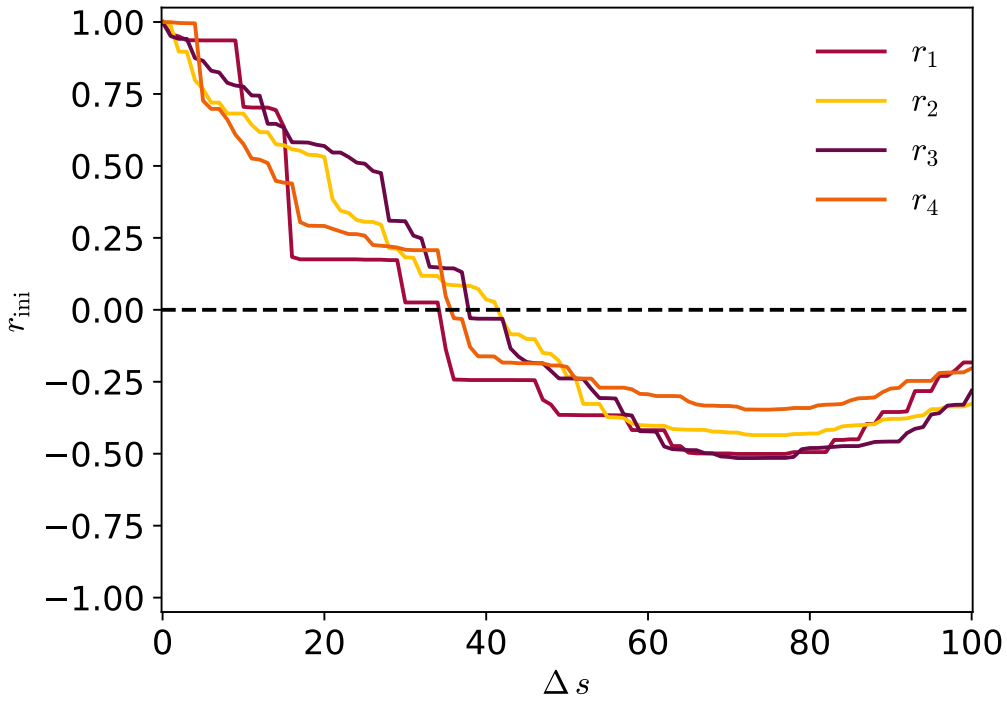

Figure D.2: Correlation plot of the fluctuating part in the initial matter density field in four of our BORG reconstructions. The correlation coefficient $r_{\rm ini}$ is measured as a function of the separation between MCMC samples Δs. The horizontal dashed line indicates a value of 0 which implies no correlation between the corresponding pair of MCMC samples.

Appendix E

Profile likelihood

The profile likelihood method [193, 210] – employed in Chapter 7 – provides not only a numerically efficient MLE of the marginalized likelihood but also the error/confidence intervals of such estimate, especially for cases with a small sample size like ours [193]. This method exploits the idea that the likelihood function of multiple model parameters, $\mathcal{P}(\{\theta\})$ wherein $\{\theta\} \equiv (a, \{b\}) \in \mathbb{R}^n$, can be reduced to a function of one single variable of interest, $\mathcal{P}^{\text{prof}}(a)$, with the rest of parameters, $\{b\}$, being treated as nuisance parameters and "profiled" out, i.e. maximized over as [210]

$$\mathcal{P}(\{\theta\}) \equiv \mathcal{P}(a, \{b\}) \to \mathcal{P}^{\text{prof}}(a) = \max_{\{\theta\} \in \mathbb{S}(a)} \mathcal{P}(a, \{b\}), \tag{E.1}$$

where $\mathbb{S}(a)$ denotes the hyperplanes in $\mathbb{R}^n$ such that $\Big\{\{\theta\} \in \mathbb{S} : \theta_0 = a\Big\}$. $\mathcal{P}^{\text{prof}}(a)$ is then an isometric profile of the original likelihood function $\mathcal{P}(\{\theta\})$ for a given $\theta_0 = a$, hence often referred to as the profile likelihood.

The error on our MLE estimate $\hat{a}$ of a is then simply given by the variance of the log-likelihood ratio $-2\ln\left[\mathcal{P}^{\text{prof}}(a)/\mathcal{P}^{\text{prof}}(\hat{a})\right]$ [193, 210], i.e. the 68% confidence interval corresponds to where $-2\ln\left[\mathcal{P}^{\text{prof}}(a)/\mathcal{P}^{\text{prof}}(\hat{a})\right]$ increases by 1 from the minimum value.

Appendix F

tSZ contamination and cluster mass cut

In [174], it has been tested that the tSZ contamination becomes important for rare, massive clusters with total mass larger than $10^{14} M_{\odot}$. Due to the limited number of clusters in the maxBCG catalog after redshift and survey mask cuts, we have found that, specifically for our sub-sample of maxBCG clusters and the Planck `SMICA` CMB map, it is necessary to remove all rare clusters heavier than $8.5 \times 10^{13} M_{\odot}$. This is demonstrated in Figure F.1 below, where we show the average AAP filter output, as a function of filter scale, at locations of clusters in each of five equi-log M_{200} bins, up to $M_{200} = 2.5 \times 10^{14} M_{\odot}$. We especially check for the cancellation of tSZ signal (and other possible foreground contaminations) by comparing the average AAP filter output to the typical amplitude of the kSZ signal of clusters in each corresponding mass bin. For simplification, we assume a universal cluster LOS velocity $v^{\rm LOS} = 300\,{\rm kms}^{-1}$.

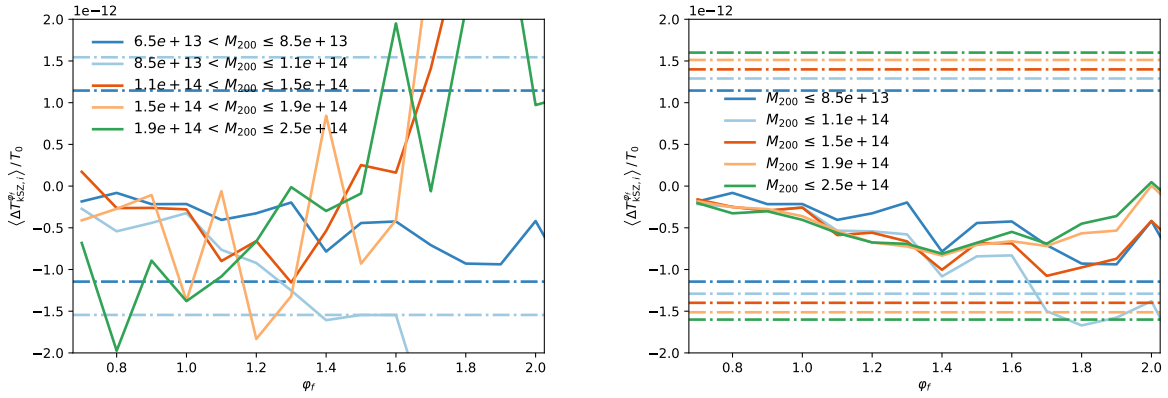

Figure F.1: The average AAP filter output from Planck `SMICA` CMB map at locations of maxBCG clusters below $2.2 \times 10^{14} M_{\odot}$, binned in M_{200} (middle) and cumulative (right). The dot-dashed lines represent the typical kSZ signal amplitude of clusters in the M_{200} bin with the same color.

Appendix G

The kSZ Likelihood: Mixture weights, MAP mean and variance of the kSZ signal amplitude

First, we derive the individual mixture weight λ_s for each component of the Gaussian mixture distribution of large-scale bulk-flow amplitude α_s in Section 6.3. Let us denote

$$A_i = \Delta T_{\mathrm{kSZ},i}/T_0, \tag{G.1}$$

and

$$B_i^s = -\tau_i v_{L,i}^{\mathrm{LOS},s}/c. \tag{G.2}$$

Then, the likelihood on the r.h.s. of Eq. (6.21) can be re-written as

$$\begin{aligned}
\ln \mathcal{P}\left(\{A_i\}|\alpha,\, \{B_i^s\}\right) &= -\frac{1}{2}\sum_i \frac{\left(A_i - \alpha\, B_i^s\right)^2}{\sigma_i^2} - \frac{1}{2}\sum_i \ln(\sigma_i^2) \\
&= -\frac{1}{2}\sum_i \left(\frac{A_i}{\sigma_i}\right)^2 + \alpha \sum_i \frac{A_i\, B_i^s}{\sigma_i^2} - \frac{\alpha^2}{2}\sum_i \left(\frac{B_i^s}{\sigma_i}\right)^2 - \frac{1}{2}\sum_i \ln(\sigma_i^2) \\
&= -\frac{1}{2}\sum_i \left(\frac{A_i}{\sigma_i}\right)^2 - \frac{1}{2}\left(\sum_i \left(\frac{B_i^s}{\sigma_i}\right)^2\right)\left(\alpha^2 - 2\alpha \frac{\sum_i \frac{A_i\, B_i^s}{\sigma_i^2}}{\sum_i \left(\frac{B_i^s}{\sigma_i}\right)^2}\right) - \frac{1}{2}\sum_i \ln\left(\sigma_i^2\right).
\end{aligned} \tag{G.3}$$

We further denote

$$\mu_s = \frac{\sum_i \left[\left(A_i\, B_i^s\right)/\sigma_i^2\right]}{\sum_i \left(B_i^s/\sigma_i\right)^2}, \tag{G.4}$$

$$\sigma_s^2 = 1/\left[\sum_i \left(\frac{B_i^s}{\sigma_i}\right)^2\right], \tag{G.5}$$

$$\gamma = \sum_i \left(\frac{A_i}{\sigma_i}\right)^2, \tag{G.6}$$

and

$$\delta = \sum_i \ln\left[\sigma_i^2\right], \tag{G.7}$$

so that we can shortened Eq. (G.3) to

$$\begin{aligned}\ln\mathcal{P} &= -\frac{1}{2}\gamma - \frac{1}{2\,\sigma_s^2}\left(\alpha^2 - 2\alpha\mu_s\right) - \frac{1}{2}\delta \\ &= -\frac{1}{2}\gamma - \frac{1}{2\,\sigma_s^2}\left((\alpha-\mu_s)^2 - \mu_s^2\right) - \frac{1}{2}\delta \\ &= -\frac{1}{2}\gamma + \frac{\mu_s^2}{2\,\sigma_s^2} - \frac{1}{2}\delta - \frac{1}{2\,\sigma_s^2}\,(\alpha-\mu_s)^2\,.\end{aligned} \tag{G.8}$$

The first and third terms on the RHS of Eq. (G.8) neither vary between BORG-SDSS3 samples nor depend on α, thus they can be dropped from Eq. (G.8), such that

$$\ln\mathcal{P} \propto \frac{1}{2\,\sigma_s^2}\left[\mu_s^2 - (\alpha-\mu_s)^2\right]. \tag{G.9}$$

We can now simply write

$$\mathcal{P}\left(\alpha|\{\Delta T_{\mathrm{kSZ},i}/T_0\},\{\tau_i v_{L,i}^{\mathrm{LOS},s}/c\}\right) \propto \mathcal{P}(\alpha)\frac{1}{N}\sum_s^N \mathrm{e}^{\omega_s}\,\sqrt{2\pi\,\sigma_s^2}\frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2\sigma_s^2}\right)}{\sqrt{2\pi\,\sigma_s^2}} \tag{G.10}$$

where $\omega_s \equiv \frac{\mu_s^2}{2\sigma_s^2}$. The *normalized* version will then be

$$\begin{aligned}\mathcal{P}\left(\alpha|\{\Delta T_{\mathrm{kSZ},i}/T_0\},\{\tau_i v_{L,i}^{\mathrm{LOS},s}/c\}\right) &= \mathcal{P}(\alpha)\frac{\frac{1}{N}\sum_s^N \mathrm{e}^{\omega_s}\,\sqrt{2\pi\,\left(\sigma_s\right)^2}\frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\,(\sigma_s)^2}}}{\frac{1}{N}\sum_s^N \mathrm{e}^{\omega_s}\,\sqrt{2\pi\,\left(\sigma_s\right)^2}} \\ &= \mathcal{P}(\alpha)\sum_s^N \lambda_s \frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\,\left(\sigma_s\right)^2}},\end{aligned} \tag{G.11}$$

where

$$\lambda_s = \frac{\mathrm{e}^{\omega_s + \frac{1}{2}\ln\left[2\pi\,(\sigma_s)^2\right]}}{\sum_s^N \mathrm{e}^{\omega_s + \frac{1}{2}\ln\left[2\pi\,(\sigma_s)^2\right]}}. \tag{G.12}$$

We can easily derive the mean of this distribution as follows, assuming a uniform prior, i.e. $\mathcal{P}(\alpha) = 1$,

$$\begin{aligned}\langle\alpha\rangle_s &= \int d\alpha\, \alpha \sum_s^N \lambda_s \frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\ (\sigma_s)^2}} \\ &= \sum_s^N \lambda_s \int d\alpha\, \alpha \frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\ (\sigma_s)^2}} \\ &= \sum_s^N \lambda_s \mu_s, \end{aligned} \tag{G.13}$$

which is precisely Eq. (6.25).

Similarly, we can explicitly work out the variance of the Gaussian mixture distribution.

$$\begin{aligned}\sigma_\alpha^2 &= \left\langle (\alpha - \langle\alpha\rangle_s)^2 \right\rangle \\ &= \int d\alpha \mathcal{P}\left(\alpha | \{\Delta T_{\mathrm{kSZ},i}/T_0\}, \{\tau_i v_{L,i}^{\mathrm{LOS},s}/c\}\right)\ (\alpha - \langle\alpha\rangle_s)^2 \\ &= \sum_s^N \lambda_s \int d\alpha \frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\ (\sigma_s)^2}}\ (\alpha - \langle\alpha\rangle_s)^2 \\ &= \sum_s^N \lambda_s \int d\alpha \frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\ (\sigma_s)^2}}\ \left[(\alpha)^2 - 2\,\alpha\,\langle\alpha\rangle_s + \langle\alpha\rangle_s^2\right] \\ &= \langle\alpha\rangle_s^2 - 2\,\langle\alpha\rangle_s^2 + \sum_s^N \lambda_s \int d\alpha \frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\ (\sigma_s)^2}}\,\alpha^2 \\ &= -\langle\alpha\rangle_s^2 + \sum_s^N \lambda_s \int d\alpha \frac{\exp\left(-\frac{(\alpha-\mu_s)^2}{2(\sigma_s)^2}\right)}{\sqrt{2\pi\ (\sigma_s)^2}}\ \left[(\alpha-\mu_s)^2 + 2\ (\alpha-\mu_s)\ \mu_s + (\mu_s)^2\right] \\ &= -\langle\alpha\rangle_s^2 + \sum_s^N \lambda_s \left[(\sigma_s)^2 + 0 + (\mu_s)^2\right] \\ &= -\langle\alpha\rangle_s^2 + \sum_s^N \lambda_s \left[(\sigma_s)^2 + (\mu_s)^2\right] \\ &= \sum_s^N \lambda_s\,(\sigma_s)^2 + \sum_s^N \lambda_s\ \left[(\mu_s - \langle\alpha\rangle_s)^2\right]. \end{aligned} \tag{G.14}$$

So the result is the sum of the average noise variances and the variance of the mean

estimate (cf. Eq. (6.26)). The second term clearly shows that our uncertainties on α include uncertainties from the velocity reconstruction.

Let us next try to compute $\ln \mathcal{P}\left(\{\boldsymbol{A_i}\}|\alpha, \{\boldsymbol{B_i^s}\}\right)$ for measurements at all θ_f scales in a similar fashion to how we arrived at Eq. (G.11).

$$\begin{aligned}
\ln \mathcal{P}\left(\{\boldsymbol{A_i}\}|\alpha, \{\boldsymbol{B_i^s}\}\right) &= -\frac{1}{2}\sum_i \left[\boldsymbol{A_i} - \alpha \boldsymbol{B_i^s}\right]^\intercal \left(\boldsymbol{C}_i\right)^{-1} \left[\boldsymbol{A_i} - \alpha \boldsymbol{B_i^s}\right] - \frac{1}{2}\sum_i \ln |\boldsymbol{C}_i| \\
&\propto -\frac{\alpha^2}{2}\sum_i \left(\boldsymbol{B_i^s}\right)^\intercal \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s} + \alpha \sum_i \boldsymbol{A_i^\intercal} \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s} \\
&\propto -\frac{1}{2}\sum_i \left(\boldsymbol{B_i^s}\right)^\intercal \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s} \left[\alpha^2 - 2\alpha \frac{\sum_i \boldsymbol{A_i^\intercal} \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s}}{\sum_i \left(\boldsymbol{B_i^s}\right)^\intercal \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s}}\right] \qquad \text{(G.15)}
\end{aligned}$$

where we have omitted terms that do not vary between BORG samples. Eq. (G.15) is similar to Eq. (G.8) with

$$\mu_s = \frac{\sum_i \boldsymbol{A_i^\intercal} \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s}}{\sum_i \left(\boldsymbol{B_i^s}\right)^\intercal \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s}} \qquad \text{(G.16)}$$

and

$$\sigma_s^2 = 1/\left[\sum_i \left(\boldsymbol{B_i^s}\right)^\intercal \left(\boldsymbol{C_i}\right)^{-1} \boldsymbol{B_i^s}\right] \qquad \text{(G.17)}$$

Appendix H

CMB contribution to covariance matrix of multi-scale kSZ measurement

In this appendix, we provide a detailed derivation of the CMB covariance matrix term (cf. Eq. (6.34)) that contributes to the covariance matrix of multi-scale kSZ measurement described in Section 6.3.2. Let us plug Eq. (6.31) into the first term on the r.h.s. of Eq. (6.30), we get

$$
\begin{aligned}
\mathcal{C}^{\theta_f\theta'_f}_{\mathrm{CMB},i} &= \left(\frac{\pi\theta_f\theta'_f}{T_0}\right)^2 \Bigg\langle \int \frac{d\boldsymbol{\ell}}{(2\pi)^2} \int \frac{d\boldsymbol{\ell}'}{(2\pi)^2} \\
&\qquad\qquad \exp\left[i\left(\boldsymbol{\ell}-\boldsymbol{\ell}'\right)\cdot\boldsymbol{\theta_i}\right] \hat{W}\left(\ell\theta_f\right)\, \Delta T^{\mathrm{obs}}_{\mathrm{CMB}}(\boldsymbol{\ell})\, \hat{W}\left(\ell'\theta'_f\right)\, \Delta T^{\mathrm{obs}}_{\mathrm{CMB}}(\boldsymbol{\ell}') \Bigg\rangle \\
&= \left(\frac{\pi\theta_f\theta'_f}{T_0}\right)^2 \int \frac{d\boldsymbol{\ell}}{(2\pi)^2} \int \frac{d\boldsymbol{\ell}'}{(2\pi)^2} \\
&\qquad\qquad \exp\left[i\left(\boldsymbol{\ell}-\boldsymbol{\ell}'\right)\cdot\boldsymbol{\theta_i}\right] \hat{W}\left(\ell\theta_f\right) \hat{W}\left(\ell'\theta'_f\right)\, \left\langle \Delta T^{\mathrm{obs}}_{\mathrm{CMB}}(\boldsymbol{\ell}) \Delta T^{\mathrm{obs}}_{\mathrm{CMB}}(\boldsymbol{\ell}') \right\rangle \\
&= \left(\frac{\pi\theta_f\theta'_f}{T_0}\right)^2 \int \frac{d\boldsymbol{\ell}}{(2\pi)^2} \int \frac{d\boldsymbol{\ell}'}{(2\pi)^2} \\
&\qquad\qquad \exp\left[i\left(\boldsymbol{\ell}-\boldsymbol{\ell}'\right)\cdot\boldsymbol{\theta_i}\right] \hat{W}\left(\ell\theta_f\right) \hat{W}\left(\ell'\theta'_f\right)\, (2\pi)^2\boldsymbol{\delta}(\boldsymbol{\ell}-\boldsymbol{\ell}')\, C^{\mathrm{CMB}}_\ell \\
&= \left(\frac{\pi\theta_f\theta'_f}{T_0}\right)^2 \int \frac{d\boldsymbol{\ell}}{(2\pi)^2}\, \hat{W}\left(\ell\theta_f\right) \hat{W}\left(\ell\theta'_f\right)\, C^{\mathrm{CMB}}_\ell \\
&= \left(\frac{\pi\theta_f\theta'_f}{T_0}\right)^2 \int_0^\infty \frac{d\ell\,\ell}{2\pi} \int_0^{2\pi} \frac{d\phi_\ell}{2\pi}\, \hat{W}\left(\ell\theta_f\right) \hat{W}\left(\ell\theta'_f\right)\, C^{\mathrm{CMB}}_\ell \\
&= \frac{\pi\theta_f^2(\theta'_f)^2}{2T_0^2} \int_0^\infty d\ell\,\ell\, \hat{W}\left(\ell\theta_f\right)\, \hat{W}\left(\ell\theta'_f\right)\, C^{\mathrm{CMB}}_\ell .
\end{aligned}
\tag{H.1}
$$

We further validate the analytical computation of the CMB covariance matrix in

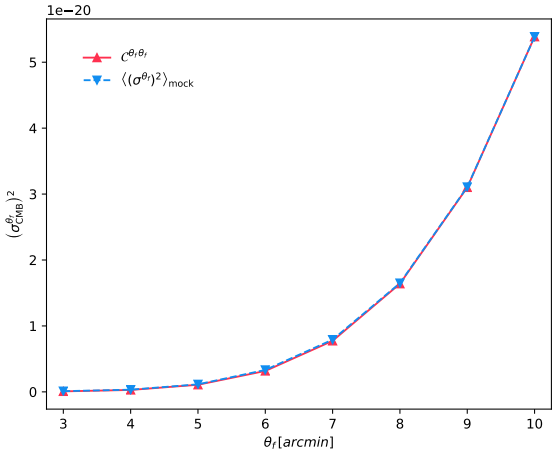

Figure H.1: Comparison between analytical (red) and numerical (blue) estimates of the CMB contribution to the diagonal of multi-scale kSZ signal covariance matrix.

Eq. (H.1) by comparing its diagonal elements with the $\sigma^2\left(\Delta T^{\theta_f}_{\rm CMB}(\boldsymbol{\theta_i})/T_0\right)$ numerically computed at 1000 random points in each of 1000 SMICA2018-like CMB mocks in Figure H.1. The analytical estimate using the Planck 2018 best-fit ΛCDM power spectrum is in extremely good agreement with the numerical estimate using SMICA2018-like CMB realizations.

Appendix I

GADGET-2 simulation of the BORG-SDSS3 volume and mock kSZ signal templates

As mentioned in Chapter 6, to generate the mock templates of kSZ signal within the BORG-SDSS3 volume, we use DM particles and halos from a GADGET-2 [83] simulation with DM-only at a very high resolution of $N_{\rm part} = 2048^3$. The initial conditions for the simulation is taken from BORG-SDSS3 sample $s = 9000$. The halos we use are main halos identified by the Rockstar halo finder algorithm[1] [181, 182] with a minimum number of 20 particles per halo. The cosmology and box size of this simulation agree with those of our BORG-SDSS3 reconstruction, while the initial conditions are taken from one MCMC sample of our BORG-SDSS3 reconstruction. The high resolution allows us to achieve a correct halo mass function (HMF) down to $M_h = 2 \times 10^{13} h^{-1} M_\odot$. Below we show the matter power spectrum $P_{mm}(k)$ and the DM halo mass function obtained from our simulation at redshift $z = 0.23$ corresponding to the median redshift of clusters in the original maxBCG catalog. For each halo i, we model its gas profile with the Gaussian profile previously considered in [174, 177]

$$n_{e,i}(\theta) = \frac{N_{e,i}}{\sqrt{2\pi\theta_i^2}} \exp\left(-\frac{\theta^2}{2\theta_i^2}\right) \tag{I.1}$$

where $\theta_i^2 = \theta_{200,i}^2 + \theta_{\rm bream}^2$ and $N_{e,i} = (f_b M_{200,i})/(\mu_e m_p)$. The LOS velocity of each particle and halo is interpolated from the BORG-PMCIC simulation using the same initial conditions used as input in our GADGET-2 simulation. We then generate a kSZ template using all DM particles within the same region analyzed in our work. The profiles shown in Figure 6.4 are measured by applying the AP filter on this mock kSZ template at locations of the DM halos.

[1] `https://bitbucket.org/gfcstanford/rockstar`

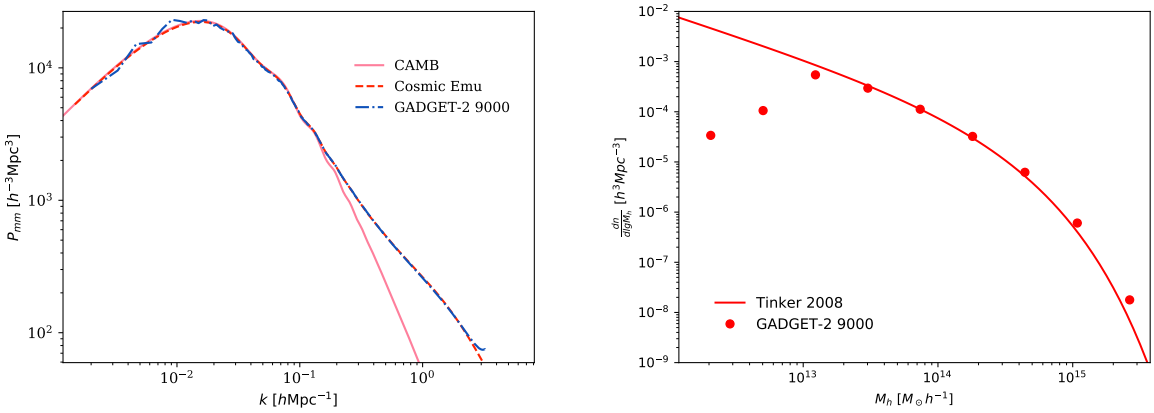

Figure I.1: *Left panel:* CDM power spectrum measured in our GADGET-2 simulation of BORG sample 9000, compared to the CAMB[1] [211, 212] linear and non-linear CDM power spectrum, emulated by Cosmic Emu[2] [213], at redshift $z = 0.23$. *Right panel:* DM halo number density found in our simulation as compared to the Tinker 2008 halo mass function [150] at redshift $z = 0.23$.

[1] `https://camb.info`

[2] `https://github.com/lanl/CosmicEmu`

Appendix J

Operator correlators and renormalization

When including nonlinear operators in the bias expansion, such as density squared δ^2, in the context of the general perturbative bias expansion, it is important to construct these operators in such a way that their cross-correlations are not sensitive to small-scale perturbations and the bias parameters associated to them are not sensitive to the smoothing scale (see [65] for a detailed discussion).

Technically, this is achieved by employing *renormalized operators* $[O]$, which are obtained from the *bare* operators O by subtracting counterterms. We follow the general approach presented in [138], who derived the following renormalization conditions:

$$\left\langle [O](\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}_1)\cdots\delta^{(1)}(\boldsymbol{k}_n)\right\rangle \overset{\{k_i\}\to 0}{\simeq} \left\langle O(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}_1)\cdots\delta^{(1)}(\boldsymbol{k}_n)\right\rangle_{\rm LO} , \tag{J.1}$$

where $n = 0, 1, 2, ...$ Here, $\langle\cdot\rangle_{\rm LO}$ stands for a correlator evaluated at leading order (LO) in perturbation theory. Eq. (J.1) ensures that all large-scale cross-correlations between the renormalized operators $[O]$ are independent of small-scale modes.

The $n = 0$ condition simply reads $\langle[O](\boldsymbol{k})\rangle = 0$, which we have already enforced in Eq. (7.1). Beyond this, at the perturbative order which we work in, it is sufficient to only consider the conditions for $n = 1$ for the quadratic operators δ^2 and K^2.[1] As shown in details in Appendix B of [123], the renormalization conditions can be enforced by subtracting counterterms in Fourier space as follows:

$$\begin{aligned}
[\delta](\boldsymbol{k}) &= \delta(\boldsymbol{k}) \\
\left[\delta^2\right](\boldsymbol{k}) &= (\delta^2)(\boldsymbol{k}) - \Sigma^2_{1-3}(k)\delta(\boldsymbol{k}) \quad \text{and} \quad [\delta^2](\boldsymbol{k}=0) = 0 \\
\left[K^2\right](\boldsymbol{k}) &= (K^2)(\boldsymbol{k}) - \frac{2}{3}\Sigma^2_{1-3}(k)\delta(\boldsymbol{k}) \quad \text{and} \quad [K^2](\boldsymbol{k}=0) = 0\,,
\end{aligned} \tag{J.2}$$

[1] Here, there is a subtlety if the quadratic operators are constructed from sharp-k filtered fields. This is discussed in Appendix B of [123].

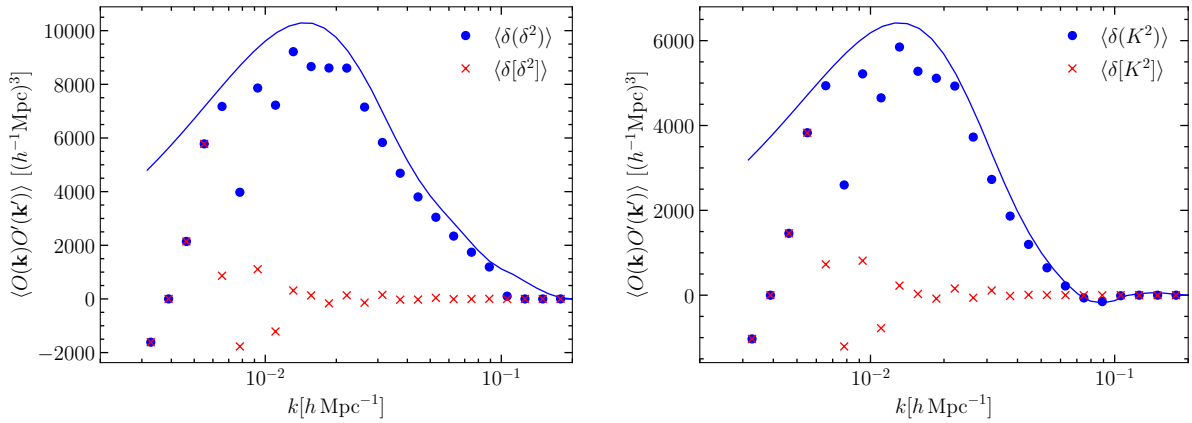

Figure J.1: *Left panel:* Measured operator correlator $\langle\delta_\Lambda(\delta_\Lambda)^2\rangle$ (i.e. before renormalization) and $\langle\delta_\Lambda[\delta_\Lambda]^2\rangle$ (after renormalization) using a 2LPT density field at $z = 0$ with $\Lambda = 0.1\,h\,\mathrm{Mpc}^{-1}$. The line shows the tree-level standard perturbation theory prediction. *Right panel:* Same as left panel, but for $\langle\delta_\Lambda(K^2_\Lambda)\rangle$ and $\langle\delta_L[K_\Lambda]^2\rangle$.

where

$$\Sigma^2_{1-3}(k) = 4\int_{\boldsymbol{p}} W_\Lambda(\boldsymbol{p})W_\Lambda(\boldsymbol{k}-\boldsymbol{p})F_2(-\boldsymbol{k},\boldsymbol{p})P_\mathrm{L}(p)\,. \tag{J.3}$$

Here, the $W_\Lambda(k)$ are sharp filter functions defined in Fourier space. Finally,

$$F_2(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{5}{7} + \frac{2}{7}\frac{(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)^2}{k_1^2k_2^2} + \frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\,. \tag{J.4}$$

Since this calculation is only valid at leading order, we use a numerical renormalization procedure in our likelihood implementation instead. Specifically, we measure

$$P_{\delta O^{[2]}}(k) = \left\langle\delta(\boldsymbol{k})O^{[2]}(\boldsymbol{k}')\right\rangle' \tag{J.5}$$

on a linear grid in k (we choose 100 bins between the fundamental and Nyquist frequencies). The same is done for the density field itself, yielding $P_{\delta\delta}(k)$. Then, for each mode $\boldsymbol{k}$, we renormalize through

$$[O^{[2]}](\boldsymbol{k}) = O^{[2]}(\boldsymbol{k}) - \frac{P_{\delta O^{[2]}}(|\boldsymbol{k}|)}{P_{\delta\delta}(|\boldsymbol{k}|)}\delta(\boldsymbol{k}), \tag{J.6}$$

where a cubic-spline interpolation is used to obtain the power spectra at each value of $\boldsymbol{k}$. Figure J.1 shows the cross-correlation of δ and the two quadratic operators δ^2 and K^2 before and after renormalization. For the k values that matter most in the likelihood, $k \gtrsim 0.02\,h\,\mathrm{Mpc}^{-1}$, the cross-correlation is removed to high accuracy by the renormalization procedure. Also shown is the tree-level perturbation-theory prediction for the correlator before renormalization, which matches the measurement reasonably well, although not perfectly even at low k, since modes near the cutoff Λ contribute to this cross-correlation.

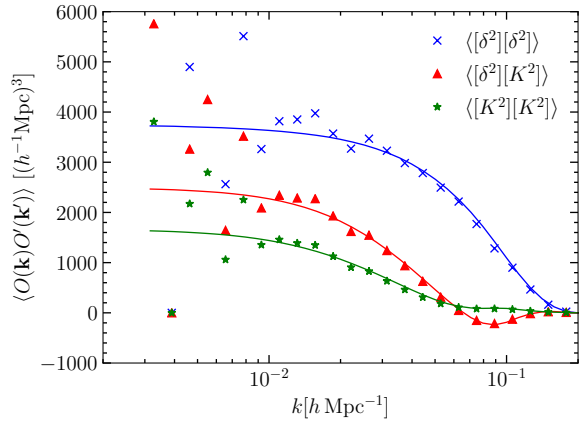

Figure J.2: Measured renormalized operator correlators $\left\langle [O^{[2]}][O'^{[2]}]\right\rangle$ for $O, O' \in \{(\delta_\Lambda)^2, (K_\Lambda)^2\}$. All operators are constructed from a 2LPT density field at $z = 0$ with $\Lambda = 0.1\,h\,\mathrm{Mpc}^{-1}$. The lines again show the tree-level standard perturbation theory predictions.

Figure J.2 shows the cross-correlation of the quadratic operators among each other. As argued in [123], the renormalization also removes the dominant higher-order (trispectrum) contribution to these correlators. Indeed, the leading perturbation-theory prediction matches the cross-correlation of the quadratic operators well. We have verified that the good agreement also holds for other values of redshift and Λ, and that the deviations from the perturbation-theory prediction show the expected scaling, with agreement improving toward higher redshift and for smaller values of the cutoff Λ. We conclude that the operator cross-correlations, which form the practical basis of the EFT likelihood as discussed in detail in [123], are well understood.

Appendix K

The EFT Gaussian likelihood: MLE for variance

Let us write $\sigma^2(k)$ in Eq. (7.12) as

$$\sigma^2(k) = \sum_{n=0,2,4} \sigma_n^2 k^n \,. \tag{K.1}$$

The maximum-likelihood points for the parameters σ_n^2 are then given by

$$0 = \sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} k^n \left[\frac{1}{2}\hat{\sigma}^{-2}(k) - \hat{\sigma}^{-4}(k)\left\{\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}(\boldsymbol{k})\right\}^2\right], \qquad n = 0, 2, 4\,. \tag{K.2}$$

That is, $\sigma^2(k)$ is attempting to match the correlator

$$\left\langle \left|\delta_h(\boldsymbol{k}) - \delta_{h,\rm det}(\boldsymbol{k})\right|^2 \right\rangle = 2 \sum_{n=0,2,4,\cdots} \hat{\sigma}_n^2 k^n \,, \tag{K.3}$$

where $\hat{\sigma}_n$ are the maximum-likelihood values for σ_n. Note that the expansion on the right-hand side only holds if all relevant contributions to $\delta_h(\boldsymbol{k})$ from modes $k < \Lambda$ are included in $\delta_{h,\rm det}(\boldsymbol{k})$. It is easy to see that this holds for the likelihood derived in Chapter 7: Eq. (K.3) corresponds to the mean squared residuals of the maximum-likelihood equations (cf. Eq. (4.5) of [123]). There are two sources of such residuals: actual noise in the halo density field, described by the statistics of the noise field ϵ_h, and residuals due to the imperfect forward modeling. As we have shown in [123], the latter are of the form Ak^2 and $ck^2 P_{\rm L}(k)$. The second type of residual is proportional to the long-wavelength modes themselves, and is thus removed by including the term $c_{\nabla^2\delta} k^2 \delta(\boldsymbol{k})$ in $\delta_{h,\rm det}(\boldsymbol{k})$. The first type corresponds to true stochastic residuals which contribute to the effective noise. We can thus write

$$\hat{\sigma}_n^2 = V_{hh}^{\epsilon,n} + V_{\rm eff}^{\epsilon,n} \,, \tag{K.4}$$

where $V_{\rm eff}^{\epsilon,0} = 0$, while, from Eq.(C.6) of [123], we expect $V_{\rm eff}^{\epsilon,2} \sim (k^2/\Lambda^2) k_{\rm NL}^{-3}$. Note that $V_{hh}^{\epsilon,0}$ corresponds to the true large-scale halo shot noise, while higher-order noise variances

are to be considered effective parameters as they absorb stochastic errors in the model. Further, by including both $V_{hh}^{\epsilon,2}$ and $V_{hm}^{\epsilon,2}$ in the likelihood, we guarantee that the mean as well as the residuals of the MLE equations can be absorbed consistently. To the order we work in throughout this thesis, the contributions of order k^4 and higher to the variance can be neglected.

Appendix L

Interpreting the variance $\boldsymbol{\sigma_\epsilon^2}$

In this appendix we derive the expectation for the variance parameter σ_ϵ^2 for Poisson noise. Neglecting long-wavelength perturbations, let $n_i \equiv n(\boldsymbol{x}_i)$ denote the number density of halos in the grid cell centered around $\boldsymbol{x}_i$. Assuming this is Poisson distributed, we obtain

$$\lambda \equiv \langle n_i \rangle = \frac{N_h}{N_g^3} \quad \text{and} \quad \langle n_i n_j \rangle = \lambda \delta_{ij} \,, \tag{L.1}$$

where N_h is the total number of halos in the box. The noise in the fractional halo density perturbation δ_h is then given by $\epsilon_i = n_i/\lambda$, where we neglect the subtraction of the mean here since it is irrelevant for modes of finite $\boldsymbol{k}$. The noise field obeys, under the Poisson assumption,

$$\langle \epsilon_i \epsilon_j \rangle = \frac{1}{\lambda} \delta_{ij} \,. \tag{L.2}$$

Finally, its power spectrum is given by

$$\left\langle |\epsilon(\boldsymbol{k})|^2 \right\rangle = \sum_{i,j} \langle \epsilon_i \epsilon_j \rangle \, e^{i\boldsymbol{k}\cdot(\boldsymbol{x}_i - \boldsymbol{x}_j)} = \frac{N_g^3}{\lambda} = \frac{N_g^6}{\bar{n}_h L_{\rm box}^3} \,, \tag{L.3}$$

where $\overline{n}_h = N_h/L_{\rm box}^3$ is the number density of halos. Notice that the value depends on the grid resolution adopted, which in our implementation is $N_g = 384$. The values of σ_ϵ^2 given in Table 7.2 are divided by this Poisson expectation. Values greater (less) than one thus correspond to super- (sub-)Poisson stochasticity.

Appendix M

Beyond the EFT Gaussian likelihood

Let us now consider the limitations of the Gaussian likelihood derived in Section 7.1. For this, we derive the expected size of the leading non-Gaussian corrections in the EFT approach. We use the box normalization in this section, since the precise counting of modes is relevant. Let us begin with the halo noise field, and assume that it has a nonzero three-point function:

$$\langle \epsilon_h(\boldsymbol{n}k_F)\epsilon_h(\boldsymbol{n}'k_F)\epsilon_h(\boldsymbol{n}''k_F)\rangle = \frac{1}{L^6_{\rm box}}\delta_{\boldsymbol{n}+\boldsymbol{n}',-\boldsymbol{n}''}B_{\epsilon_h}(\boldsymbol{n}k_F,\boldsymbol{n}'k_F)\,. \tag{M.1}$$

Strictly speaking, we need to consider the bivariate distribution for $\{\vec{\epsilon}_h,\vec{\epsilon}_m\}$ and integrate out $\vec{\epsilon}_m$ following Eq. (7.5). However, we are mostly interested in the scaling of the leading correction to the Gaussian likelihood. Let us thus use the Edgeworth expansion for ϵ_h, so that the likelihood of the ϵ_h field including the leading non-Gaussian correction becomes (again up to an irrelevant normalization)

$$-2\ln\mathcal{P}(\vec{\epsilon}_h) = \sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}}\Bigg[\ln\sigma^2_{\epsilon_h}(k) + \frac{|\epsilon_h(\boldsymbol{k})|^2}{\sigma^2_{\epsilon_h}(k)} - \frac{1}{3}\sum_{\boldsymbol{k}'\neq 0}^{k_{\rm max}}\frac{B_{\epsilon_h}(\boldsymbol{k},\boldsymbol{k}')}{L^6_{\rm box}\sigma^2_{\epsilon_h}(k)\sigma^2_{\epsilon_h}(k')\sigma^2_{\epsilon_h}(|\boldsymbol{k}+\boldsymbol{k}'|)}\epsilon_h(\boldsymbol{k})\epsilon_h(\boldsymbol{k}')\epsilon_h(-\boldsymbol{k}-\boldsymbol{k}')\Bigg]\,,$$

where $\sigma^2_{\epsilon_h}(k) = P^{\epsilon}_{hh}(k)/L^3_{\rm box}$. As an approximation to the proper marginalization over the stochastic fields, we now insert $\epsilon_h(\boldsymbol{k}) = \delta_h(\boldsymbol{k}) - \delta_{h,\rm det}(\boldsymbol{k})$ to obtain the leading correction to the likelihood $\ln\mathcal{P}(\vec{\delta}_h|\vec{\delta},\{b_O\},\{\lambda_a\})$. Taking the expectation value of the derivative of

this likelihood with respect to b_O, we obtain

$$\sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \frac{1}{\sigma^2(k)} \langle [O](\boldsymbol{k})\delta_h(\boldsymbol{k})\rangle = \sum_{O'} \hat{b}_{O'} \sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \frac{1}{\sigma^2(k)} \langle [O](\boldsymbol{k})[O'](\boldsymbol{k})\rangle \qquad \forall\ O$$
$$\overset{O=\delta}{+} \left[V^{\epsilon,2}_{hm}k^2 + V^{\epsilon,4}_{hm}k^4 + b_1 V^{\epsilon,4}_{mm}k^4\right]$$
$$+ \sum_{\boldsymbol{k}\neq 0}^{k_{\rm max}} \sum_{\boldsymbol{k}'\neq 0}^{k_{\rm max}} \frac{B_{\epsilon_h}(\boldsymbol{k},\boldsymbol{k}')B_{\epsilon_h\epsilon_h O}(\boldsymbol{k},\boldsymbol{k}')}{L^{12}_{\rm box}\sigma^2_{\epsilon_h}(k)\sigma^2_{\epsilon_h}(k')\sigma^2_{\epsilon_h}(|\boldsymbol{k}+\boldsymbol{k}'|)}\,, \tag{M.2}$$

where the first two lines come from the Gaussian part and are identical to Eq. (4.5) of [123], while the third line is the approximate expression for the leading non-Gaussian correction. The latter involves the continuum-normalized cross-bispectrum between ϵ_h and O:

$$B_{\epsilon_h\epsilon_h O}(\boldsymbol{k},\boldsymbol{k}') \equiv \langle \epsilon_h(\boldsymbol{k})\epsilon_h(\boldsymbol{k}')O(-\boldsymbol{k}-\boldsymbol{k}')\rangle'\,. \tag{M.3}$$

Note that the only operator for which this contribution can be valid on large scales is $O=\delta$, since, for any quadratic or higher-order operator, Eq. (M.3) starts at 1-loop order (i.e. involves at least 3, rather than 2 power spectra), and thus has to be suppressed on large scales. The bispectrum $B_{\epsilon_h\epsilon_h\delta}(\boldsymbol{k},\boldsymbol{k}')$ has a straightforward interpretation: it corresponds to the modulation of the halo noise amplitude by a density perturbation $\delta(|\boldsymbol{k}+\boldsymbol{k}'|)$.

In order to gain quantitative insight, let us thus consider $O=\delta$ and assume that ϵ_h follows a Poisson distribution. We then have (see e.g. [96] and Sec. 4.1 of [65])

$$P^{\epsilon}_{hh}(k) = \frac{1}{\overline{n}_h} \quad \Rightarrow \quad \sigma^2_{\epsilon_h}(k) = \frac{1}{\overline{n}_h L^3_{\rm box}}$$
$$B_{\epsilon_h}(\boldsymbol{k},\boldsymbol{k}') = \frac{1}{\overline{n}_h^2}$$
$$B_{\epsilon_h\epsilon_h\delta}(\boldsymbol{k},\boldsymbol{k}') = \frac{b_1}{\overline{n}_h} P_{\rm L}(|\boldsymbol{k}+\boldsymbol{k}'|)\,, \tag{M.4}$$

where $\overline{n}_h$ is the mean comoving number density of halos. Crucially, while the exact amplitude is not expected to match the actual noise field of halos, the leading scaling with k of all these contributions holds regardless of the Poisson assumption. The final ingredient needed for the estimate is the number of Fourier modes included in the likelihood. This is, approximately,

$$\frac{4\pi}{3}\frac{k^3_{\rm max}}{k^3_F} = \frac{4\pi}{3(2\pi)^3}(L_{\rm box}k_{\rm max})^3\,. \tag{M.5}$$

We then obtain the following upper limit on the size of the non-Gaussian contribution to the maximum-likelihood equality Eq. (M.2) at fixed $\boldsymbol{k}$:

$$\left|\sum_{\boldsymbol{k}'\neq 0}^{k_{\rm max}} \frac{B_{\epsilon_h}(\boldsymbol{k},\boldsymbol{k}')B_{\epsilon_h\epsilon_h O}(\boldsymbol{k},\boldsymbol{k}')}{L^{12}_{\rm box}\sigma^2_{\epsilon_h}(k)\sigma^2_{\epsilon_h}(k')\sigma^2_{\epsilon_h}(|\boldsymbol{k}+\boldsymbol{k}'|)}\right| = b_1 \sum_{\boldsymbol{k}'\neq 0}^{k_{\rm max}} \frac{(\overline{n}_h L^3_{\rm box})^3}{L^{12}_{\rm box}\overline{n}_h^3} P_{\rm L}(|\boldsymbol{k}+\boldsymbol{k}'|)$$
$$\lesssim \frac{b_1}{6\pi^2} k^3_{\rm max} P_{\rm L}(k_{\rm max})\,. \tag{M.6}$$

Again, the prefactor can differ by order unity, as it relies on the Poisson assumption. Crucially, the scaling with $k_{\rm max}$ is robust. Eq. (M.6) states that the non-Gaussianity of the halo noise field can be safely neglected as long as $k_{\rm max}^3 P_{\rm L}(k_{\rm max}) < \sigma_\Lambda^2$ is much less than one. We see that this is the same condition which is required for a reliable EFT likelihood in general.

Finally, let us turn to the non-Gaussianity of the matter noise field $\epsilon_m(\boldsymbol{k})$. As in the case of its variance, its non-Gaussian correlators are suppressed by powers of k. We roughly expect (see Section 4.5 of [123]) that $B_{\epsilon_m}(\boldsymbol{k}, \boldsymbol{k}') \sim k^2 k'^2 |\boldsymbol{k} + \boldsymbol{k}'|^2 / k_{\rm NL}^{12}$. The dominant contribution on large scales is expected to be the cross-correlation with the halo noise, whose magnitude can be roughly bounded to

$$|B_{\epsilon_m \epsilon_h \epsilon_h}(\boldsymbol{k}, \boldsymbol{k}')| \lesssim \frac{k^2}{k_{\rm NL}^2} B_{\epsilon_h}(\boldsymbol{k}, \boldsymbol{k}') \,. \tag{M.7}$$

We thus expect that the contributions involving ϵ_m to the maximum-likelihood point Eq. (M.2) (whose precise form we have not derived here), are correspondingly suppressed by powers of $(k/k_{\rm NL})^2$, and hence smaller than the halo noise contributions (unless one were to consider an extremely dense halo sample, or matter itself as tracer). This is the same conclusion as we have reached in the Gaussian case, where the leading contribution to the variance is P_{hh}^ϵ, while P_{hm}^ϵ is only relevant at NLO.

Appendix N

Relation between Bayesian forward inference and BAO reconstruction

The linear matter power spectrum contains an oscillatory feature, the baryon acoustic oscillation (BAO) feature, induced by sound waves in the baryon-photon fluid before recombination [214]. Since the physical scale corresponding to this feature, the sound horizon at recombination r_s, is known, a measurement of its apparent scale in the clustering of galaxies allows for direct estimates of the angular diameter distance and Hubble rate as functions of redshift.

The BAO feature in the power spectrum of the evolved density field is broadened due to the nonlinear growth of structure. This broadening degrades the precision with which the scale of the BAO feature can be determined in galaxy clustering. The dominant source of this broadening are displacements induced by large-scale modes [215]. Fortunately, since galaxy displacements are unbiased at lowest order in derivatives (on large scales), a fact which is ensured by the equivalence principle, the broadening obtained in a given realization of the density field can be predicted robustly.

For this reason, BAO reconstruction approaches have been developed. Generally, these work by first estimating the displacement field using the galaxy density smoothed on a large scale (via Eq. (N.4), in case of the Zel'dovich approximation), and then moving galaxies back to their initial positions using this estimated displacement. Since the first implementations of the method [216], many refined versions have been presented [217, 218, 219, 220, 221, 152, 222, 223, 153]. What all current reconstruction methods have in common is the backward-modeling approach and the presence of a smoothing scale.

Now let us consider BAO reconstruction from the perspective of a perturbative Bayesian forward model. The Eulerian position $\boldsymbol{x}$ at which a given galaxy is observed can be related to the corresponding Lagrangian position, i.e. position in the initial conditions, through the displacement $\boldsymbol{s}$:

$$\boldsymbol{x} = \boldsymbol{x}_{\text{fl}}(\boldsymbol{q}, \tau) = \boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q}, \tau)\,. \tag{N.1}$$

Since standard reconstruction methods are based on inferring large-scale displacements by assuming a linear bias relation, let us do the same here; we return to this issue below.

Assuming a deterministic forward model, as we do throughout, the final posterior Eq. (4.6) then becomes

$$\mathcal{P}\left(\vec{\delta}_{m,\mathrm{ini}}, \theta, b_1, \{\lambda_a\} \middle| \vec{\delta}_h\right) = \mathcal{N}_P \mathcal{P}_{\mathrm{prior}}(\vec{\delta}_{m,\mathrm{ini}}|\theta) \mathcal{P}\left(\vec{\delta}_h - b_1 \vec{\delta} \middle| \{\lambda_a\}\right) . \tag{N.2}$$

Due to mass conservation of matter, the density field is directly related to the Lagrangian displacement. In Fourier space, this relation becomes (e.g., [224])

$$\delta(\boldsymbol{k}) = \int_{\boldsymbol{q}} \exp\left[-i\boldsymbol{k}\cdot(\boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q}))\right] , \quad k > 0. \tag{N.3}$$

Note that this relation does not ensure that $\delta(\boldsymbol{k} = 0) = 0$. However, this is not relevant here as we use a Fourier-space likelihood that does not include $\boldsymbol{k} = 0$. Using the latter, Eq. (N.2) becomes

$$\begin{aligned} -\ln \mathcal{P}\left(\vec{\delta}_{m,\mathrm{ini}}, \theta, b_1, \{\lambda_a\} \middle| \vec{\delta}_h\right) = \sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \Bigg[& \frac{|\delta_{\mathrm{in}}(\boldsymbol{k})|^2}{2P_{\mathrm{L}}(k|\theta)} \\ & + \frac{1}{2\sigma^2(k)} \left| \delta_h(\boldsymbol{k}) - b_1 \int_{\boldsymbol{q}} \exp\left[-i\boldsymbol{k}\cdot\left(\boldsymbol{q} + \boldsymbol{s}[\vec{\delta}_{m,\mathrm{ini}}](\boldsymbol{q})\right)\right] \right|^2 \Bigg] . \end{aligned}$$

Here, we have assumed for simplicity that the range in wavenumber space covered by the prior on δ_{in} is the same as that of the likelihood involving δ_h, although this does not have to be the case in practice (and is not in our actual implementation; Appendix O).

In order to gain further analytic insight into how the posterior derived here works, let us assume the Zel'dovich approximation (ZA). We stress that in practice, our forward model never consists of this Zel'dovich approximation, but always more sophisticated forward models; for example, 2LPT or full N-body simulations. However, the ZA allows for a particularly simple illustration for how BAO reconstruction works when using a field-level inference approach. In the ZA, the Lagrangian displacement is evaluated at linear order:

$$\boldsymbol{s}(\boldsymbol{p}) = \boldsymbol{s}_{\mathrm{ZA}}(\boldsymbol{p}) \equiv -\frac{i\boldsymbol{p}}{p^2}\delta_{\mathrm{in}}(\boldsymbol{p}) . \tag{N.4}$$

We then obtain

$$\begin{aligned} -\ln \mathcal{P}\left(\vec{\delta}_{m,\mathrm{ini}}, \theta, b_1, \{\lambda_a\} \middle| \vec{\delta}_h\right) \overset{\mathrm{ZA}}{=} \sum_{\boldsymbol{k}\neq 0}^{k_{\max}} \Bigg[& \frac{|\delta_{\mathrm{in}}(\boldsymbol{k})|^2}{2P_{\mathrm{L}}(k|\theta)} \\ & + \frac{1}{2\sigma^2(k)} \left| \delta_h(\boldsymbol{k}) - b_1 \int_{\boldsymbol{q}} \exp\left[-i\boldsymbol{k}\cdot\left(\boldsymbol{q} + \int_{\boldsymbol{p}} \frac{-i\boldsymbol{p}}{p^2}\delta_{\mathrm{in}}(\boldsymbol{p}) e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\right)\right] \right|^2 \Bigg] . \end{aligned} \tag{N.5}$$

In order to obtain cosmological constraints, we now need to marginalize over the initial phases $\delta_{\mathrm{in}}(\boldsymbol{k})$ as well as b_1. Clearly, despite the very simple forward model, Eq. (N.5) is a

highly complex non-Gaussian and non-separable posterior, and the result of the marginalization is far from obvious.

Nevertheless, if one expands the exponential on the right-hand side up to linear order in $\delta_{\rm in}$, the non-prior part of the posterior reduces to a sum over $[\delta_h(\boldsymbol{k}) - b_1\delta_{\rm in}(\boldsymbol{k})]^2/2\sigma^2(k)$, i.e. precisely the expected Gaussian, linear-bias likelihood in the large-scale limit. We can thus assume that, after marginalization over b_1, the posterior will peak around the correct large-scale modes of the density field.[1] Let us thus assume that marginalizing over the initial phases with $|\boldsymbol{k}| < k_c$, where k_c will be determined below, fixes the amplitudes $\delta_{\rm in}(\boldsymbol{k})$ to their true values.

The BAO feature is an oscillatory feature in the linear power spectrum,

$$P_{\rm L}(k|\theta = \{\theta', r_s\}) = P_{\rm L}^{\rm smooth}(k|\theta')\left[1 + A_{\rm BAO}\sin(kr_s)\right] , \tag{N.6}$$

where r_s is the sound horizon, and $P_{\rm L}^{\rm smooth}(k)$ is the smooth (non-wiggle) part of the power spectrum, while $A_{\rm BAO}$ is the amplitude of the BAO feature which is not important here. We are interested in the constraint on r_s obtained from the posterior Eq. (N.5). Again, we cannot explicitly marginalize over $\delta_{\rm in}$. We however see that, upon taking a derivative of the posterior with respect to r_s, in order to obtain the maximum posterior and the curvature around it, we obtain a sum over wavenumbers weighted by an oscillatory function of k with frequency r_s.

The modes which dominate the BAO broadening effect are on much larger scales than the BAO feature. Hence, we choose $k_c \ll 1/r_s$, and marginalize over the modes with $|\boldsymbol{k}| < k_c$. Following the discussion above, we then obtain

$$\begin{aligned} -r_s\frac{\partial}{\partial r_s}\ln\mathcal{P}\left(\vec{\delta}_{m,\rm ini}, \theta, b_1, \{\lambda_a\}\middle|\vec{\delta}_h\right) &\overset{\rm ZA,marg.}{=} \sum_{k_c<|\boldsymbol{k}|<k_{\rm max}} (kr_s)\cos(kr_s) \\ &\times\left[\frac{|\delta_{\rm in}(\boldsymbol{k})|^2}{2P_{\rm L}(k|\theta)} + \frac{1}{2\sigma^2(k)}\left|\delta_h(\boldsymbol{k}) - b_1\int_{\boldsymbol{q}}\exp\left[-i\boldsymbol{k}\cdot\left(\boldsymbol{q} + \int_{\boldsymbol{p}}^{k_c}\frac{-i\boldsymbol{p}}{p^2}\delta_{\rm in}(\boldsymbol{p})e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\right)\right]\right|^2\right] . \end{aligned} \tag{N.7}$$

Note that, due to the oscillatory nature of the BAO feature, the displacement terms on the right-hand side are enhanced by $k/p \sim 1/(r_s p)$, and thus significant. Eq. (N.7) precisely corresponds to constraining the BAO scale by evaluating a Gaussian, linear-bias likelihood for the linear density field displaced to the Eulerian position predicted by the large-scale displacements $(-i\boldsymbol{p}/p^2)\delta_{\rm in}(\boldsymbol{p})$ which are reconstructed from the large-scale halo field $\delta_h(\boldsymbol{p})$. Thus, our posterior, combined with a Lagrangian forward model, naturally includes BAO reconstruction. The key difference from standard reconstruction approaches is that Eq. (N.7), in keeping with the entire Bayesian inference approach, employs a *forward model* of the BAO displacements, where the data are compared to the displaced initial density field. The former approaches on the other hand attempt to *move the data back* to

[1] If the set of cosmological parameters includes the power spectrum normalization σ_8, then there is a perfect degeneracy of b_1 and σ_8 in the linear regime. However, as discussed in the previous sections, this degeneracy is broken when the forward model and likelihood are consistently extended beyond linear order. Current reconstruction approaches instead adopt a prior on σ_8.

the initial positions. We stress that Eq. (N.7) is a rough approximation intended to isolate the relevant aspects of the inference approach; in practice, the actual posterior for r_s is obtained after proper marginalization over all initial phases.

Moreover, unlike the case in reconstruction approaches applied to the data, which explicitly introduce a smoothing scale $\sim 1/k_c$ used to reconstruct the displacement field from the observed galaxy density field, the posterior presented here self-consistently uses information from all scales $< k_{\rm max}$ to infer the displacements (again, assuming the proper marginalization over all initial phases is performed). In our numerical implementation, described in detail in [136], we go beyond the Zel'dovich approximation and use 2LPT. This improves the accuracy of the displacement field $\boldsymbol{s}[\vec{\delta}_{m,\rm ini}]$. Further, the posterior correctly includes the subleading effects of large-scale modes on the BAO feature, which correspond to over-dense regions effectively behaving like a curved universe with correspondingly different expansion history [225]. The Zel'dovich approximation does not correctly predict the amplitude of this effect. We emphasize again that the approach presented here is not restricted to the 2LPT forward model, and full N-body simulations could be used instead. Finally, since the same forward model is used for all operators, the posterior presented here also includes BAO reconstruction at the bispectrum level.

Appendix O

Results from the analytical MLE of the EFT likelihood

In Chapter 7, we already show results from numerical tests of σ_8 inference using the *marginalized* likelihood, where we marginalize over bias parameters (see Section 7.3). Here, we focus on the analytical maximum-likelihood estimates of both bias parameters and σ_8 from the *full* likelihood derived in Section 7.1 and [123, 187].

In what follows, our input data consists of two N-body simulations coming from the same suite of simulations used in Chapter 5 and Chapter 7. Their description thus follows that in Section 5.2 with $\sigma_8 = 0.85$. We use AHF halos identified at $z = 0$ as tracers to verify if our method allows to recover unbiased estimates of the input parameters.

To comply with the requirements summarized in Section 7.2, we first project the halo catalog, linearly evolved matter density, and nonlinear matter density fields onto high-resolution grids with 512^3 cells using CIC assignment. Here, the non-linear matter density field is constructed from a forward evolution using 2LPT, which can also be used for efficient sampling of initial phases. Using the density field constructed from the N-body particles results in similar results, since our analysis is using only fairly large scales. We subsequently apply a low-pass filter $W_\Lambda(\boldsymbol{k})$ to the data in Fourier space, removing modes for which $k > \Lambda = 0.1\,h$ Mpc^{-1}. Starting from the filtered density fields, we are then able to compute (renormalized) representations of the operators O as well as evaluate their Fourier space correlators $\langle OO'^* \rangle$ numerically. With a particular focus on the estimation of cosmological parameters, we can proceed to derive maximum likelihood estimates for the scaled bias parameters β_O and σ_8 given by Eq. (4.5) of [123], where we truncated the scale-dependent variance at order k^2. While this set of equations is non-linear in the parameters (in particular in $\alpha \equiv \sigma_8/\sigma_{8,\mathrm{fid}}$), explicit solutions can be easily found by means of standard computer algebra systems.

In Figure O.1 we show the results of our analysis, plotted as a function of wavevector k in order to illustrate any systematic trends. Here, we are using halos with masses $10^{13}\,h^{-1}M_\odot \leq M \leq 10^{13.5}\,h^{-1}M_\odot$ (left panel) and $10^{14}\,h^{-1}M_\odot \leq M \leq 10^{14.5}\,h^{-1}M_\odot$ (right panel) as biased tracers of the matter density field from two independent simulation runs. Averaged over scales up to $k_{\mathrm{max}} = 0.05\,h$ Mpc^{-1}, we obtain numerical values for the

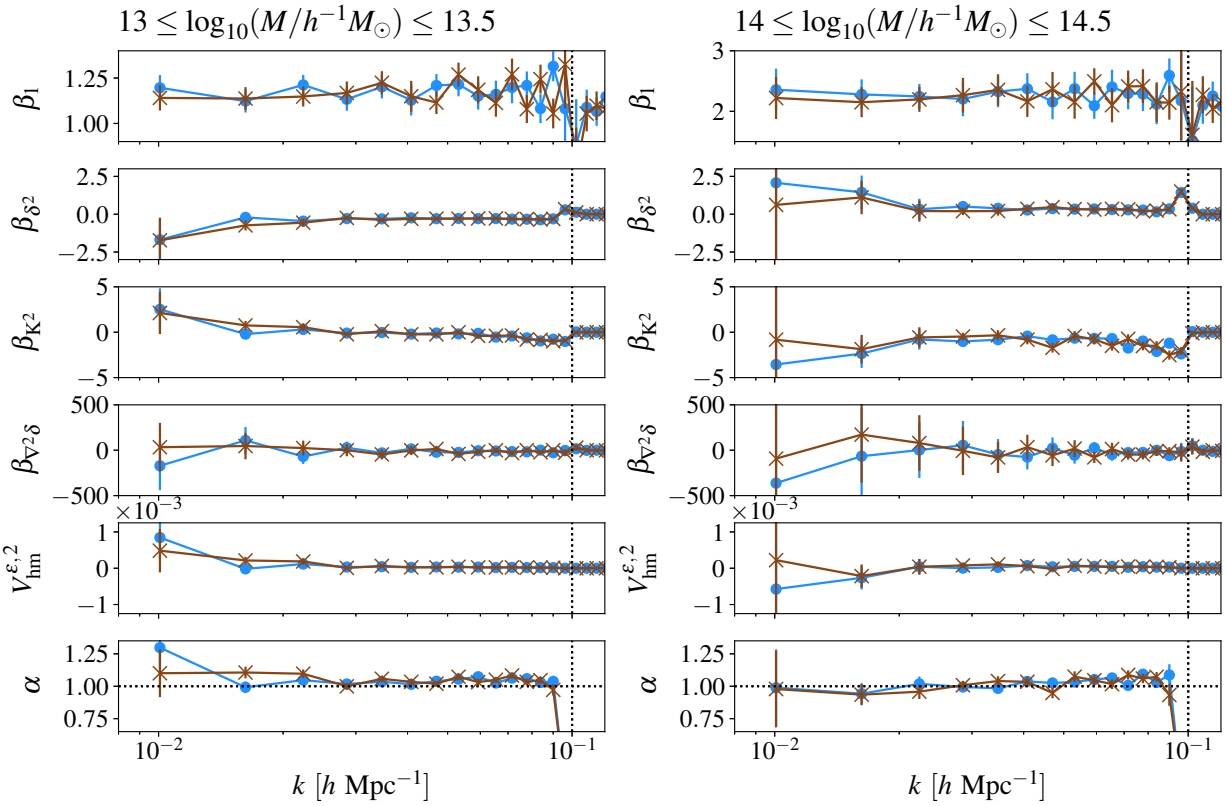

Figure O.1: The EFT-based likelihood presented here allows for an unbiased inference of bias and cosmological parameters like σ_8. For halos in the mass range $10^{13}\,h^{-1}M_\odot \leq M \leq 10^{13.5}\,h^{-1}M_\odot$ (*left panel*) and $10^{14}\,h^{-1}M_\odot \leq M \leq 10^{14.5}\,h^{-1}M_\odot$ (*right panel*), we show maximum likelihood estimates of the scaled bias parameters $\beta_1, \beta_{\delta^2}, \beta_{K^2}, \beta_{\nabla^2\delta}$, the variance $V_{hm}^{\epsilon,2}$, and $\alpha \equiv \sigma_8/\sigma_{8,\rm fid}$ (*from top to bottom*) for two different simulations as a function of k. The filtering scale is indicated by vertical dotted lines while the horizontal dotted line in the bottom panels corresponds to an unbiased σ_8 estimate that matches the input value of the simulation. Error bars are 2-σ bootstrap estimates over halo subsamples, which do not include residual cosmic variance.

scaling parameters $\alpha^{13-13.5} = 1.03,\ 1.04$, and $\alpha^{14-14.5} = 1.01,\ 0.99$ for the two realizations, respectively, indicating that we can consistently recover the input parameters from modes below the smoothing scale used in this example. The statistical uncertainty in the inferred value of α is not simple to estimate, since we have fixed the phases in our inference to the true values, and hence cancel cosmic variance to the largest possible degree. The differences between the two simulation realizations give a rough indication of the error in the MLE parameters. Further, as a lower limit on the errors associated with the estimation, we also show uncertainty estimates from bootstrap samples of the halo catalog in Figure O.1 but stress that they fail to fully capture the cosmic variance contribution. Clearly, σ_8 is correctly recovered up to a few percent accuracy, with some indication for a small positive bias.

We stress again that the tests discussed here are in some sense the most stringent possible probes of theoretical systematics, since a perfect knowledge of the underlying matter density field has been assumed. In real-world applications, however, the phases have to be inferred from the data and are subject to uncertainties that in general result in larger parameter error bars, potentially making remaining deficiencies in the theoretical modeling insignificant compared to statistical uncertainties. We defer a detailed investigation of this question to follow-up works where a similar EFT likelihood is implemented in the BORG framework which should allow for full, efficient sampling of all parameters, including these phases.

Bibliography

[1] M. G. Hauser and P. J. E. Peebles, *Statistical Analysis of Catalogs of Extragalactic Objects. II. the Abell Catalog of Rich Clusters*, *ApJ* **185** (Nov, 1973) 757–786.

[2] P. J. E. Peebles and M. G. Hauser, *Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs*, *ApJS* **28** (Nov, 1974) 19.

[3] P. J. E. Peebles, *Statistical Analysis of Catalogs of Extragalactic Objects.IV. Cross-Correlation of the Abell and Shane-Wirtanen Catalogs*, *ApJS* **28** (Nov, 1974) 37.

[4] P. J. E. Peebles and E. J. Groth, *Statistical analysis of catalogs of extragalactic objects. V. Three-point correlation function for the galaxy distribution in the Zwicky catalog.*, *ApJ* **196** (Feb, 1975) 1–11.

[5] P. J. E. Peebles, *Statistical analysis of catalogs of extragalactic objects. VI. The galaxy distribution in the Jagellonian field.*, *ApJ* **196** (Mar, 1975) 647–651.

[6] J. Tonry and M. Davis, *A survey of galaxy redshifts. I. Data reduction techniques.*, *AJ* **84** (Oct, 1979) 1511–1525.

[7] S. J. Maddox, G. Efstathiou, W. J. Sutherland and J. Loveday, *Galaxy correlations on large scales.*, *MNRAS* **242** (Jan, 1990) 43.

[8] P. J. E. Peebles, *Statistical Analysis of Catalogs of Extragalactic Objects. I. Theory*, *ApJ* **185** (Oct, 1973) 413–440.

[9] J. R. Bond, R. Crittenden, R. L. Davis, G. Efstathiou and P. J. Steinhardt, *Measuring cosmological parameters with cosmic microwave background experiments*, *PhRvL* **72** (Jan, 1994) 13–16, [astro-ph/9309041].

[10] M. Davis, J. Huchra, D. W. Latham and J. Tonry, *A survey of galaxy redshifts. II. The large scale space distribution.*, *ApJ* **253** (Feb, 1982) 423–445.

[11] D. J. Baumgart and J. N. Fry, *Fourier Spectra of Three-dimensional Data*, *ApJ* **375** (Jul, 1991) 25.

[12] C. Park, I. Gott, J. R. and L. N. da Costa, *Large-Scale Structure in the Southern Sky Redshift Survey*, *ApJL* **392** (Jun, 1992) L51.

[13] M. Davis, G. Efstathiou, C. S. Frenk and S. D. M. White, *The evolution of large-scale structure in a universe dominated by cold dark matter*, *ApJ* **292** (May, 1985) 371–394.

[14] G. Efstathiou, J. R. Bond and S. D. M. White, *COBE background radiation anisotropies and large-scale structure in the universe*, *MNRAS* **258** (Sep, 1992) 1P–6P.

[15] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess et al., *Structure in the COBE Differential Microwave Radiometer First-Year Maps*, *ApJL* **396** (Sep, 1992) L1.

[16] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *AJ* **145** (Jan, 2013) 10, [1208.0022].

[17] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson et al., *The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *AJ* **151** (Feb, 2016) 44, [1508.04473].

[18] M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein et al., *The DESI Experiment, a whitepaper for Snowmass 2013*, *arXiv e-prints* (Aug, 2013) arXiv:1308.0847, [1308.0847].

[19] L. Amendola, S. Appleby, D. Bacon, T. Baker, M. Baldi, N. Bartolo et al., *Cosmology and Fundamental Physics with the Euclid Satellite*, *Living Reviews in Relativity* **16** (Sep, 2013) 6, [1206.1225].

[20] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman et al., *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, *ApJ* **873** (Mar., 2019) 111, [0805.2366].

[21] M. Takada, R. S. Ellis, M. Chiba, J. E. Greene, H. Aihara, N. Arimoto et al., *Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph*, *PASJ* **66** (Feb, 2014) R1, [1206.0737].

[22] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard et al., *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*, *ApJS* **208** (Oct, 2013) 20, [1212.5225].

[23] Planck Collaboration, *Planck 2013 results. I. Overview of products and scientific results*, *A&A* **571** (Nov, 2014) A1, [1303.5062].

[24] Planck Collaboration, *Planck 2015 results. I. Overview of products and scientific results*, *A&A* **594** (Sep, 2016) A1, [1502.01582].

[25] Planck Collaboration, *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *arXiv e-prints* (Jul, 2018) arXiv:1807.06205, [1807.06205].

[26] J. E. Austermann, K. A. Aird, J. A. Beall, D. Becker, A. Bender, B. A. Benson et al., *SPTpol: an instrument for CMB polarization measurements with the South Pole Telescope*, vol. 8452 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 84521E. 2012. 10.1117/12.927286.

[27] R. J. Thornton, P. A. R. Ade, S. Aiola, F. E. Angilè, M. Amiri, J. A. Beall et al., *The Atacama Cosmology Telescope: The Polarization-sensitive ACTPol Instrument*, *ApJS* **227** (Dec, 2016) 21, [1605.06569].

[28] S. W. Henderson, R. Allison, J. Austermann, T. Baildon, N. Battaglia, J. A. Beall et al., *Advanced ACTPol Cryogenic Detector Arrays and Readout*, *Journal of Low Temperature Physics* **184** (Aug, 2016) 772–779, [1510.02809].

[29] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold et al., *CMB-S4 Science Book, First Edition*, *arXiv e-prints* (Oct, 2016) arXiv:1610.02743, [1610.02743].

[30] G. D'Amico, J. Gleyzes, N. Kokron, D. Markovic, L. Senatore, P. Zhang et al., *The Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure*, *arXiv e-prints* (Sep, 2019) arXiv:1909.05271, [1909.05271].

[31] M. M. Ivanov, M. Simonović and M. Zaldarriaga, *Cosmological Parameters from the BOSS Galaxy Power Spectrum*, *arXiv e-prints* (Sep, 2019) arXiv:1909.05277, [1909.05277].

[32] S. Joudaki, C. Blake, A. Johnson, A. Amon, M. Asgari, A. Choi et al., *KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering*, *MNRAS* **474** (Mar, 2018) 4894–4924, [1707.06627].

[33] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen et al., *Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing*, *PhRvD* **98** (Aug, 2018) 043526, [1708.01530].

[34] T. Tröster, A. G. Sánchez, M. Asgari, C. Blake, M. Crocce, C. Heymans et al., *Cosmology from large-scale structure. Constraining ΛCDM with BOSS*, *A&A* **633** (Jan, 2020) L10, [1909.11006].

[35] C. L. Bennett, A. J. Banday, K. M. Gorski, G. Hinshaw, P. Jackson, P. Keegstra et al., *Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results*, *ApJL* **464** (Jun, 1996) L1, [astro-ph/9601067].

[36] Planck Collaboration, *Planck 2018 results. VII. Isotropy and Statistics of the CMB*, *arXiv e-prints* (Jun, 2019) arXiv:1906.02552, [1906.02552].

[37] P. Coles and B. Jones, *A lognormal model for the cosmological mass distribution*, *MNRAS* **248** (Jan, 1991) 1–13.

[38] J. Carron, *On the Incompleteness of the Moment and Correlation Function Hierarchy as Probes of the Lognormal Field*, *ApJ* **738** (Sep, 2011) 86, [1105.4467].

[39] J. Jasche and B. D. Wandelt, *Bayesian physical reconstruction of initial conditions from large-scale structure surveys*, *MNRAS* **432** (Jun, 2013) 894–913, [1203.3639].

[40] J. Jasche and F. S. Kitaura, *Fast Hamiltonian sampling for large-scale structure inference*, *MNRAS* **407** (Sep, 2010) 29–42, [0911.2496].

[41] B. Reid, S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, R. Tojeiro et al., *SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues*, *MNRAS* **455** (Jan., 2016) 1553–1573, [1509.06529].

[42] B. P. Koester et al., *A MaxBCG Catalog of 13,823 Galaxy Clusters from the Sloan Digital Sky Survey*, *ApJ* **660** (May, 2007) 239–255, [astro-ph/0701265].

[43] G. Lavaux, J. Jasche and F. Leclercq, *Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data*, *arXiv e-prints* (sep, 2019) arXiv:1909.06396, [1909.06396].

[44] Planck Collaboration, *Planck 2018 results. IV. Diffuse component separation*, *arXiv e-prints* (Jul, 2018) arXiv:1807.06208, [1807.06208].

[45] F. Bernardeau, S. Colombi, E. Gaztañaga and R. Scoccimarro, *Large-scale structure of the Universe and cosmological perturbation theory*, *PhR* **367** (Sep, 2002) 1–248, [astro-ph/0112551].

[46] S. Weinberg, *Cosmology*. 2008.

[47] S. Dodelson, *Modern cosmology*. 2003.

[48] A. P. S. Yadav and B. D. Wandelt, *Primordial Non-Gaussianity in the Cosmic Microwave Background*, *Advances in Astronomy* **2010** (Jan., 2010) 565248, [1006.0275].

[49] D. Pietrobon, P. Cabella, A. Balbi, G. de Gasperis and N. Vittorio, *Constraints on primordial non-Gaussianity from a needlet analysis of the WMAP-5 data*, *MNRAS* **396** (July, 2009) 1682–1688, [0812.2478].

[50] Planck Collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *arXiv e-prints* (May, 2019) arXiv:1905.05697, [1905.05697].

[51] J. M. Bardeen, J. R. Bond, N. Kaiser and A. S. Szalay, *The Statistics of Peaks of Gaussian Random Fields*, *ApJ* **304** (May, 1986) 15.

[52] L. ISSERLIS, *On a formula for the product-moment coefficient of any order of a normal*
frequency distribution in any number of variables, *Biometrika* **12** (11, 1918) 134–139, [https://academic.oup.com/biomet/ARTICLE-pdf/12/1-2/134/481266/12-1-2-134.pdf].

[53] G. C. Wick, *The Evaluation of the Collision Matrix*, *Physical Review* **80** (Oct., 1950) 268–272.

[54] P. J. E. Peebles, *Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations*, *ApJL* **263** (Dec., 1982) L1–L5.

[55] J. R. Bond, A. S. Szalay and M. S. Turner, *Formation of Galaxies in a Gravitino-Dominated Universe*, *PhRvL* **48** (June, 1982) 1636–1640.

[56] G. R. Blumenthal, H. Pagels and J. R. Primack, *Galaxy formation by dissipationless particles heavier than neutrinos*, *Nature* **299** (Sept., 1982) 37–38.

[57] G. R. Blumenthal, S. M. Faber, J. R. Primack and M. J. Rees, *Formation of galaxies and large-scale structure with cold dark matter.*, *Nature* **311** (Oct., 1984) 517–525.

[58] M. H. Goroff, B. Grinstein, S. J. Rey and M. B. Wise, *Coupling of modes of cosmological mass density fluctuations*, *ApJ* **311** (Dec., 1986) 6–14.

[59] N. Makino, M. Sasaki and Y. Suto, *Analytic approach to the perturbative expansion of nonlinear gravitational fluctuations in cosmological density and velocity fields*, *PhRvD* **46** (July, 1992) 585–602.

[60] B. Jain and E. Bertschinger, *Second-Order Power Spectrum and Nonlinear Evolution at High Redshift*, *ApJ* **431** (Aug., 1994) 495, [astro-ph/9311070].

[61] F. R. Bouchet, *Introductory Overview of Eulerian and Lagrangian Perturbation Theories*, in *Dark Matter in the Universe* (S. Bonometto, J. R. Primack and A. Provenzale, eds.), p. 565, Jan., 1996, astro-ph/9603013.

[62] D. J. Heath, *The growth of density perturbations in zero pressure Friedmann-Lemaître universes.*, *MNRAS* **179** (May, 1977) 351–358.

[63] H. Martel and W. Freudling, *Second-Order Perturbation Theory in Omega not = 1 Friedmann Models*, *ApJ* **371** (Apr., 1991) 1.

[64] R. Takahashi, *Third-Order Density Perturbation and One-Loop Power Spectrum in Dark-Energy-Dominated Universe*, *Progress of Theoretical Physics* **120** (Sept., 2008) 549–559, [0806.1437].

[65] V. Desjacques, D. Jeong and F. Schmidt, *Large-scale galaxy bias*, *PhR* **733** (Feb, 2018) 1–193, [1611.09787].

[66] D. Huterer, D. Kirkby, R. Bean, A. Connolly, K. Dawson, S. Dodelson et al., *Growth of cosmic structure: Probing dark energy beyond expansion*, *AstropARTICLE Physics* **63** (Mar., 2015) 23–41, [1309.5385].

[67] T. Buchert, *A class of solutions in Newtonian cosmology and the pancake theory*, *A&A* **223** (Oct., 1989) 9–24.

[68] F. Moutarde, J. M. Alimi, F. R. Bouchet, R. Pellat and A. Ramani, *Precollapse Scale Invariance in Gravitational Instability*, *ApJ* **382** (Dec., 1991) 377.

[69] T. Buchert, *Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies and the 'Zel'dovich approximation'*, *MNRAS* **254** (Feb., 1992) 729–737.

[70] F. R. Bouchet, R. Juszkiewicz, S. Colombi and R. Pellat, *Weakly Nonlinear Gravitational Instability for Arbitrary Omega*, *ApJL* **394** (July, 1992) L5.

[71] T. Buchert and J. Ehlers, *Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies – second-order approach: an improved model for non-linear clustering*, *MNRAS* **264** (Sept., 1993) 375–387.

[72] T. Buchert, A. L. Melott and A. G. Weiss, *Testing higher-order Lagrangian perturbation theory against numerical simulations I. Pancake models*, *A&A* **288** (Aug., 1994) 349–364, [astro-ph/9309056].

[73] F. R. Bouchet, S. Colombi, E. Hivon and R. Juszkiewicz, *Perturbative Lagrangian approach to gravitational instability*, *A&A* **296** (Apr., 1995) 575, [astro-ph/9406013].

[74] Y. B. Zel'Dovich, *Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations.*, *A&A* **500** (Mar., 1970) 13–18.

[75] A. G. Doroshkevich, *The space structure of perturbations and the origin of rotation of galaxies in the theory of fluctuation.*, *Astrofizika* **6** (Jan., 1970) 581–600.

[76] A. L. Melott, T. Buchert and A. G. Weiss, *Testing higher-order Lagrangian perturbation theory against numerical simulations. II. Hierarchical models.*, *A&A* **294** (Feb., 1995) 345–365, [astro-ph/9404018].

[77] A. G. Weiss, S. Gottlöber and T. Buchert, *Optimizing higher order Lagrangian perturbation theory for standard CDM and BSI models*, *MNRAS* **278** (Feb., 1996) 953–964, [astro-ph/9505113].

[78] M. McQuinn and M. White, *Cosmological perturbation theory in 1+1 dimensions*, *JCAP* **2016** (Jan., 2016) 043, [1502.07389].

[79] M. Crocce, S. Pueblas and R. Scoccimarro, *Transients from initial conditions in cosmological simulations*, *MNRAS* **373** (Nov., 2006) 369–381, [astro-ph/0606505].

[80] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, *arXiv e-prints* (July, 2018) arXiv:1807.06209, [1807.06209].

[81] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, *MNRAS* **470** (Sept., 2017) 2617–2652, [1607.03155].

[82] F. Leclercq, J. Jasche, H. Gil-Marín and B. Wandelt, *One-point remapping of Lagrangian perturbation theory in the mildly non-linear regime of cosmic structure formation*, *JCAP* **2013** (Nov., 2013) 048, [1305.4642].

[83] V. Springel, *The cosmological simulation code GADGET-2*, *MNRAS* **364** (Dec., 2005) 1105–1134, [astro-ph/0505010].

[84] R. Scoccimarro, *The Bispectrum: From Theory to Observations*, *ApJ* **544** (Dec., 2000) 597–615, [astro-ph/0004086].

[85] J. Silk, A. Di Cintio and I. Dvorkin, *Galaxy Formation*, in *Proceedings of the International School of Physics 'Enrico Fermi' Course 186 'New Horizons for Observational Cosmology' Vol. 186*, vol. 186, pp. 137–187, Dec., 2014, 1312.0107, DOI.

[86] R. S. Somerville and R. Davé, *Physical Models of Galaxy Formation in a Cosmological Framework*, *ARA&A* **53** (Aug., 2015) 51–113, [1412.2712].

[87] T. Naab and J. P. Ostriker, *Theoretical Challenges in Galaxy Formation*, *ARA&A* **55** (Aug., 2017) 59–109, [1612.06891].

[88] N. Kaiser, *On the spatial correlations of Abell clusters*, *ApJL* **284** (Sept., 1984) L9–L12.

[89] V. Desjacques, D. Jeong and F. Schmidt, *Non-Gaussian Halo Bias Re-examined: Mass-dependent Amplitude from the Peak-Background Split and Thresholding*, *PhRvD* **84** (Sept., 2011) 063512, [1105.3628].

[90] F. Schmidt, D. Jeong and V. Desjacques, *Peak-background split, renormalization, and galaxy clustering*, *PhRvD* **88** (July, 2013) 023515, [1212.0868].

[91] D. Jeong and E. Komatsu, *Perturbation Theory Reloaded: Analytical Calculation of Nonlinearity in Baryonic Oscillations in the Real-Space Matter Power Spectrum*, *ApJ* **651** (Nov., 2006) 619–626, [astro-ph/0604075].

[92] J. N. Fry, *The Evolution of Bias*, *ApJL* **461** (Apr., 1996) L65.

[93] M. Tegmark and P. J. E. Peebles, *The Time Evolution of Bias*, *ApJL* **500** (June, 1998) L79–L82, [astro-ph/9804067].

[94] P. Catelan, F. Lucchin, S. Matarrese and C. Porciani, *The bias field of dark matter haloes*, *MNRAS* **297** (July, 1998) 692–712, [astro-ph/9708067].

[95] V. Desjacques, M. Crocce, R. Scoccimarro and R. K. Sheth, *Modeling scale-dependent bias on the baryonic acoustic scale with the statistics of peaks of Gaussian random fields*, *PhRvD* **82** (Nov., 2010) 103529, [1009.3449].

[96] F. Schmidt, *Towards a self-consistent halo model for the nonlinear large-scale structure*, *PhRvD* **93** (Mar., 2016) 063512, [1511.02231].

[97] A. Akbar Abolhasani, M. Mirbabayi and E. Pajer, *Systematic renormalization of the effective theory of Large Scale Structure*, *JCAP* **2016** (May, 2016) 063, [1509.07886].

[98] M. Hendry, *Dynamical methods for reconstructing the large scale galaxy density and velocity fields*, in *The Restless Universe* (B. A. Steves and A. J. Maciejewski, eds.), pp. 191–216, Jan., 2001.

[99] R. van de Weygaert and W. Schaap, *The cosmic web: Geometric analysis*, *Lecture Notes in Physics* (2008) 291–413.

[100] P. L. Schechter, *Mass-to-light ratios for elliptical galaxies.*, *AJ* **85** (July, 1980) 801–811.

[101] M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro et al., *The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey*, *ApJ* **606** (May, 2004) 702–740, [astro-ph/0310725].

[102] J. Jasche and G. Lavaux, *Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe*, *A&A* **625** (may, 2019) A64, [1806.11117].

[103] J. C. Jackson, *A critique of Rees's theory of primordial gravitational radiation*, *MNRAS* **156** (Jan., 1972) 1P, [0810.3908].

[104] L. Samushia, B. A. Reid, M. White, W. J. Percival, A. J. Cuesta, G.-B. Zhao et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring growth rate and geometry with anisotropic clustering*, *MNRAS* **439** (Apr., 2014) 3504–3519, [1312.4899].

[105] A. Taruya, T. Nishimichi and S. Saito, *Baryon acoustic oscillations in 2d: Modeling redshift-space power spectrum from perturbation theory*, *Physical Review D* **82** (Sep, 2010) .

[106] A. Taruya, S. Saito and T. Nishimichi, *Forecasting the cosmological constraints with anisotropic baryon acoustic oscillations from multipole expansion*, *Physical Review D* **83** (May, 2011) .

[107] Y. Zheng and Y.-S. Song, *Study on the mapping of dark matter clustering from real space to redshift space*, *Journal of Cosmology and AstropARTICLE Physics* **2016** (Aug, 2016) 050–050.

[108] Y.-S. Song, Y. Zheng, A. Taruya and M. Oh, *Hybrid modeling of redshift space distortions*, *Journal of Cosmology and AstropARTICLE Physics* **2018** (Jul, 2018) 018–018.

[109] S. W. Allen, A. E. Evrard and A. B. Mantz, *Cosmological Parameters from Observations of Galaxy Clusters*, *ARA&A* **49** (Sept., 2011) 409–470, [1103.4829].

[110] E. Rozo, R. H. Wechsler, E. S. Rykoff, J. T. Annis, M. R. Becker, A. E. Evrard et al., *Cosmological Constraints from the Sloan Digital Sky Survey maxBCG Cluster Catalog*, *ApJ* **708** (Jan., 2010) 645–660, [0902.3702].

[111] T. de Haan, B. A. Benson, L. E. Bleem, S. W. Allen, D. E. Applegate, M. L. N. Ashby et al., *Cosmological Constraints from Galaxy Clusters in the 2500 Square-degree SPT-SZ Survey*, *ApJ* **832** (Nov., 2016) 95, [1603.06522].

[112] DES Collaboration, T. Abbott, M. Aguena, A. Alarcon, S. Allam, S. Allen et al., *Dark Energy Survey Year 1 Results: Cosmological Constraints from Cluster Abundances and Weak Lensing*, *arXiv e-prints* (Feb., 2020) arXiv:2002.11124, [2002.11124].

[113] G. O. Abell, *The Distribution of Rich Clusters of Galaxies.*, *ApJS* **3** (May, 1958) 211.

[114] T. Szabo, E. Pierpaoli, F. Dong, A. Pipino and J. Gunn, *An Optical Catalog of Galaxy Clusters Obtained from an Adaptive Matched Filter Finder Applied to Sloan Digital Sky Survey Data Release 6*, *ApJ* **736** (July, 2011) 21, [1011.0249].

[115] B. P. Koester, T. A. McKay, J. Annis, R. H. Wechsler, A. E. Evrard, E. Rozo et al., *MaxBCG: A Red-Sequence Galaxy Cluster Finder*, *ApJ* **660** (May, 2007) 221–238, [astro-ph/0701268].

[116] J. Hao, T. A. McKay, B. P. Koester, E. S. Rykoff, E. Rozo, J. Annis et al., *A GMBCG Galaxy Cluster Catalog of 55,424 Rich Clusters from SDSS DR7*, *ApJS* **191** (Dec., 2010) 254–274, [1010.5503].

[117] E. S. Rykoff, E. Rozo, M. T. Busha, C. E. Cunha, A. Finoguenov, A. Evrard et al., *redMaPPer. I. Algorithm and SDSS DR8 Catalog*, *ApJ* **785** (Apr., 2014) 104, [1303.3562].

[118] E. S. Rykoff, E. Rozo, D. Hollowood, A. Bermeo-Hernand ez, T. Jeltema, J. Mayers et al., *The RedMaPPer Galaxy Cluster Catalog From DES Science Verification Data*, *ApJS* **224** (May, 2016) 1, [1601.00621].

[119] E. S. Sheldon, D. E. Johnston, R. Scranton, B. P. Koester, T. A. McKay, H. Oyaizu et al., *Cross-correlation Weak Lensing of SDSS Galaxy Clusters. I. Measurements*, *ApJ* **703** (Oct., 2009) 2217–2231, [0709.1153].

[120] D. E. Johnston, E. S. Sheldon, R. H. Wechsler, E. Rozo, B. P. Koester, J. A. Frieman et al., *Cross-correlation Weak Lensing of SDSS galaxy Clusters II: Cluster Density Profiles and the Mass–Richness Relation*, *arXiv e-prints* (Sept., 2007) arXiv:0709.1159, [0709.1159].

[121] E. S. Rykoff, T. A. McKay, M. R. Becker, A. Evrard, D. E. Johnston, B. P. Koester et al., *Measuring the Mean and Scatter of the X-Ray Luminosity-Optical Richness Relation for maxBCG Galaxy Clusters*, *ApJ* **675** (Mar., 2008) 1106–1124, [0709.1158].

[122] E. Rozo et al., *Constraining the Scatter in the Mass-richness Relation of maxBCG Clusters with Weak Lensing and X-ray Data*, *ApJ* **699** (jul, 2009) 768–781, [0809.2794].

[123] F. Schmidt, F. Elsner, J. Jasche, N. M. Nguyen and G. Lavaux, *A rigorous EFT-based forward model for large-scale structure*, *JCAP* **2019** (jan, 2019) 042, [1808.02002].

[124] L. Senatore, *Bias in the effective field theory of large scale structures*, *JCAP* **11** (Nov., 2015) 007, [1406.7843].

[125] M. Mirbabayi, F. Schmidt and M. Zaldarriaga, *Biased tracers and time evolution*, *JCAP* **7** (July, 2015) 30, [1412.5169].

[126] C. Modi, Y. Feng and U. Seljak, *Cosmological Reconstruction From Galaxy Light: Neural Network Based Light-Matter Connection*, *ArXiv e-prints* (May, 2018) , [1805.02247].

[127] D. K. Ramanah, G. Lavaux, J. Jasche and B. D. Wand elt, *Cosmological inference from Bayesian forward modelling of deep galaxy redshift surveys*, *A&A* **621** (Jan., 2019) A69, [1808.07496].

[128] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Physics Letters B* **195** (Sep, 1987) 216–222.

[129] R. M. Neal, *Slice Sampling*, *arXiv e-prints* (Sept., 2000) physics/0009028, [physics/0009028].

[130] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, *Biometrika* **57** (04, 1970) 97–109, [https://academic.oup.com/biomet/ARTICLE-pdf/57/1/97/8545904/57-1-97.pdf].

[131] S. Geman and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6** (1984) 721–741.

[132] A. E. Gelfand and S. A. F. M., *Sampling-based approaches to calculating marginal densities*, *Journal of the American Statistical Association* **85** (1990) 398–409.

[133] W. R. Gilks, N. G. Best and K. K. C. Tan, *Adaptive rejection metropolis sampling within gibbs sampling*, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **44** (1995) 455–472.

[134] G. B. Rybicki and W. H. Press, *Interpolation, Realization, and Reconstruction of Noisy, Irregularly Sampled Data*, *ApJ* **398** (Oct., 1992) 169.

[135] G. Lavaux and J. Jasche, *Unmasking the masked Universe: the 2M++ catalogue through Bayesian eyes*, *MNRAS* **455** (Jan., 2016) 3169–3179, [1509.05040].

[136] F. Elsner, F. Schmidt, J. Jasche, G. Lavaux and N.-M. Nguyen, *Cosmology inference from a biased density field using the EFT-based likelihood*, *JCAP* **2020** (Jan., 2020) 029, [1906.07143].

[137] M. C. Neyrinck, M. A. Aragón-Calvo, D. Jeong and X. Wang, *A halo bias function measured deeply into voids without stochasticity*, *Monthly Notices of the Royal Astronomical Society* **441** (05, 2014) 646–655, [https://academic.oup.com/mnras/ARTICLE-pdf/441/1/646/3039316/stu589.pdf].

[138] V. Assassi, D. Baumann, D. Green and M. Zaldarriaga, *Renormalized halo bias*, *JCAP* **2014** (Aug., 2014) 056, [1402.5916].

[139] R. Casas-Miranda, H. J. Mo, R. K. Sheth and G. Boerner, *On the distribution of haloes, galaxies and mass*, *MNRAS* **333** (July, 2002) 730–738, [astro-ph/0105008].

[140] M. Biagetti, T. Lazeyras, T. Baldauf, V. Desjacques and F. Schmidt, *Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity*, *Mon. Not. Roy. Astron. Soc.* **468** (2017) 3277–3288, [1611.04901].

[141] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *arXiv e-prints* (Apr., 2011) arXiv:1104.2932, [1104.2932].

[142] R. Scoccimarro, L. Hui, M. Manera and K. C. Chan, *Large-scale bias and efficient generation of initial conditions for nonlocal primordial non-Gaussianity*, *PhRvD* **85** (Apr., 2012) 083002, [1108.5512].

[143] J. E. Gunn and I. Gott, J. Richard, *On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution*, *ApJ* **176** (Aug., 1972) 1.

[144] W. H. Press and P. Schechter, *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, *ApJ* **187** (Feb., 1974) 425–438.

[145] A. V. Kravtsov and S. Borgani, *Formation of Galaxy Clusters*, *ARA&A* **50** (Sept., 2012) 353–409, [1205.5556].

[146] S. P. Gill, A. Knebe and B. K. Gibson, *The Evolution substructure 1: A New identification method*, *Mon.Not.Roy.Astron.Soc.* **351** (2004) 399, [astro-ph/0404258].

[147] S. R. Knollmann and A. Knebe, *AHF: Amiga's Halo Finder*, *ApJS* **182** (June, 2009) 608–624, [0904.3662].

[148] N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith and T. Baldauf, *Minimizing the stochasticity of halos in large-scale structure surveys*, *PhRvD* **82** (Aug., 2010) 043515–+, [1004.5377].

[149] D. Ginzburg, V. Desjacques and K. C. Chan, *Shot noise and biased tracers: A new look at the halo model*, *PhRvD* **96** (Oct., 2017) 083528, [1706.08738].

[150] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes et al., *Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality*, *ApJ* **688** (Dec., 2008) 709–728, [0803.2706].

[151] T. Lazeyras, C. Wagner, T. Baldauf and F. Schmidt, *Precision measurement of the local bias of dark matter halos*, *JCAP* **2016** (Feb., 2016) 018, [1511.01096].

[152] M. Schmittfull, T. Baldauf and M. Zaldarriaga, *Iterative initial condition reconstruction*, *PhRvD* **96** (July, 2017) 023505, [1704.06634].

[153] J. Birkin, B. Li, M. Cautun and Y. Shi, *Reconstructing the baryon acoustic oscillations using biased tracers*, *MNRAS* **483** (Mar., 2019) 5267–5280, [1809.08135].

[154] N. Porqueres, D. K. Ramanah, J. Jasche and G. Lavaux, *Explicit Bayesian Treatment of Unknown Foreground Contaminations in Galaxy Surveys*, *A&A* **624** (apr, 2019) A115, [1812.05113].

[155] R. A. Sunyaev and Y. B. Zeldovich, *The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies*, *Comments on Astrophysics and Space Physics* **4** (1972) 173.

[156] J. P. Ostriker and E. T. Vishniac, *Generation of Microwave Background Fluctuations from Nonlinear Perturbations at the ERA of Galaxy Formation*, *ApJL* **306** (July, 1986) L51.

[157] P. R. Phillips, *Calculation of the Kinetic Sunyaev-Zeldovich Effect from the Boltzmann Equation*, *ApJ* **455** (Dec., 1995) 419.

[158] M. Birkinshaw, *The Sunyaev-Zel'dovich effect*, *PhR* **310** (Mar., 1999) 97–195, [astro-ph/9808050].

[159] Planck Collaboration, *Planck intermediate results - XXXVII. Evidence of unbound gas from the kinetic Sunyaev-Zeldovich effect*, *A&A* **586** (feb, 2016) , [1502.015976].

[160] M. S. Madhavacheril, N. Battaglia, K. M. Smith and J. L. Sievers, *Cosmology with kSZ: breaking the optical depth degeneracy with Fast Radio Bursts*, *arXiv e-prints* (jan, 2019) arXiv:1901.02418, [1901.02418].

[161] R. Keisler and F. Schmidt, *Prospects for measuring the relative velocities of galaxy clusters in photometric surveys using the kinetic Sunyaev-Zel'dovich Effect*, *ApJL* **765** (2013) , [1211.0668].

[162] Y.-Z. Ma and G.-B. Zhao, *Dark energy imprints on the kinematic Sunyaev-Zel'dovich signal*, *Physics Letters B* **735** (jul, 2014) 402–411, [1309.1163].

[163] E.-M. Mueller, F. de Bernardis, R. Bean and M. D. Niemack, *Constraints on gravity and dark energy from the pairwise kinematic Sunyaev-Zeldovich effect*, *ApJ* **808** (jul, 2015) , [1408.6248].

[164] E.-M. Mueller, F. de Bernardis, R. Bean and M. D. Niemack, *Constraints on massive neutrinos from the pairwise kinematic Sunyaev-Zel'dovich effect*, *PhRvD* **92** (sep, 2015) 063501, [1412.0592].

[165] C. Hernández-Monteagudo, Y.-Z. Ma, F. S. Kitaura, W. Wang, R. Génova-Santos, J. Macías-Pérez et al., *Evidence of the Missing Baryons from the Kinematic Sunyaev-Zeldovich Effect in Planck Data*, *PhRvL* **115** (Nov, 2015) 191301, [1504.04011v3].

[166] S. Lim, H. Mo, H. Wang and X. Yang, *The detection of missing baryons in galaxy halos with kinetic Sunyaev-Zel'dovich effect*, *arXiv e-prints* (dec, 2017) arXiv:1712.08619, [1712.08619].

[167] M. Fukugita, C. J. Hogan and P. J. E. Peebles, *The Cosmic Baryon Budget*, *ApJ* **503** (aug, 1998) 518–530, [astro-ph/9712020].

[168] J. M. Shull, B. D. Smith and C. W. Danforth, *The Baryon Census in a Multiphase Intergalactic Medium: 30% of the Baryons May Still be Missing*, *ApJ* **759** (Nov., 2012) 23, [1112.2706].

[169] M. Münchmeyer, M. S. Madhavacheril, S. Ferraro, M. C. Johnson and K. M. Smith, *Constraining local non-Gaussianities with kinetic Sunyaev-Zel'dovich tomography*, *PhRvD* **100** (oct, 2019) 083508, [1810.13424].

[170] Y.-Z. Ma, *Constraining the ionized gas evolution with CMB-spectroscopic survey cross-correlation*, *Nuclear Physics B* **920** (July, 2017) 402–418, [1707.03348].

[171] N. Hand et al., *Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel'dovich Effect*, *PhRvL* **109** (July, 2012) 041101, [1203.4219].

[172] B. Soergel et al., *Detection of the kinematic Sunyaev-Zel'dovich effect with DES Year 1 and SPT*, *MNRAS* **461** (Sept., 2016) 3172–3193, [1603.03904].

[173] F. De Bernardis et al., *Detection of the pairwise kinematic Sunyaev-Zel'dovich effect with BOSS DR11 and the Atacama Cosmology Telescope*, *JCAP* **2017** (mar, 2017) 008, [1607.02139].

[174] E. Schaan et al., *Evidence for the kinematic Sunyaev-Zel'dovich effect with the Atacama Cosmology Telescope and velocity reconstruction from the Baryon Oscillation Spectroscopic Survey*, *PhRvD* **93** (Apr., 2016) 082002, [1510.06442].

[175] B. Soergel, A. Saro, T. Giannantonio, G. Efstathiou and K. Dolag, *Cosmology with the pairwise kinematic SZ effect: calibration and validation using hydrodynamical simulations*, *MNRAS* **478** (aug, 2018) 5320–5335, [1712.05714].

[176] J.-F. Cardoso, M. Martin, J. Delabrouille, M. Betoule and G. Patanchon, *Component separation with flexible models. Application to the separation of astrophysical emissions*, *arXiv e-prints* (Mar., 2008) arXiv:0803.1814, [0803.1814].

[177] N. S. Sugiyama, T. Okumura and D. N. Spergel, *A direct measure of free electron gas via the kinematic Sunyaev-Zel'dovich effect in Fourier-space analysis*, *MNRAS* (2018) stx3362, [1705.07449].

[178] M. Li, R. E. Angulo, S. D. M. White and J. Jasche, *Matched filter optimization of kSZ measurements with a reconstructed cosmological flow field*, *MNRAS* **443** (2014) 2311–2326, [1404.0007v2].

[179] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, *ApJ* **622** (Apr., 2005) 759–771, [arXiv:astro-ph/0409513].

[180] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python*, *Journal of Open Source Software* **4** (Mar., 2019) 1298.

[181] P. S. Behroozi, R. H. Wechsler and H.-Y. Wu, *The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores*, *ApJ* **762** (Jan., 2013) 109, [1110.4372].

[182] A. Knebe, N. I. Libeskind, F. Pearce, P. Behroozi, J. Casado, K. Dolag et al., *Galaxies going MAD: the Galaxy-Finder Comparison Project*, *MNRAS* **428** (Jan., 2013) 2039–2052, [1210.2578].

[183] j. . a. k. . A. y. . . m. . j. e. . a. p. . a. a. . a. e. . . p. . a. a. . h. a. . P. Planck Collaboration, title = "Planck 2018 results. III. High Frequency Instrument data processing and frequency maps".

[184] Planck Collaboration, *Planck 2015 results. VIII. High Frequency Instrument data processing: Calibration and maps*, *A&A* **594** (Sept., 2016) A8, [1502.01587].

[185] R. Chown, Y. Omori, K. Aylor, B. A. Benson, L. E. Bleem, J. E. Carlstrom et al., *Maps of the Southern Millimeter-wave Sky from Combined 2500 deg^2 SPT-SZ and Planck Temperature Data*, *ApJS* **239** (Nov., 2018) 10, [1803.10682].

[186] J. Bobin, F. Sureau and J.-L. Starck, *Cosmic microwave background reconstruction from WMAP and Planck PR2 data*, *A&A* **591** (2016) 12, [1511.08690].

[187] G. Cabass and F. Schmidt, *The Likelihood for Large-Scale Structure*, *arXiv e-prints* (Sept., 2019) arXiv:1909.04022, [1909.04022].

[188] T. Lazeyras and F. Schmidt, *Beyond LIMD bias: a measurement of the complete set of third-order halo bias parameters*, *ArXiv e-prints* (Dec., 2017) , [1712.07531].

[189] M. M. Abidi and T. Baldauf, *Cubic Halo Bias in Eulerian and Lagrangian Space*, *ArXiv e-prints* (Feb., 2018) , [1802.07622].

[190] T. Baldauf, E. Schaan and M. Zaldarriaga, *On the reach of perturbative methods for dark matter density fields*, *JCAP* **3** (Mar., 2016) 007, [1507.02255].

[191] T. Baldauf, E. Schaan and M. Zaldarriaga, *On the reach of perturbative descriptions for dark matter displacement fields*, *JCAP* **3** (Mar., 2016) 017, [1505.07098].

[192] M. Schmittfull, M. Simonović, V. Assassi and M. Zaldarriaga, *Modeling Biased Tracers at the Field Level*, *arXiv e-prints* (Nov, 2018) arXiv:1811.10640, [1811.10640].

[193] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Ann. Math. Statist.* **9** (03, 1938) 60–62.

[194] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput. Phys. Commun.* **10** (1975) 343–367.

[195] T. Lazeyras and F. Schmidt, *A robust measurement of the first higher-derivative bias of dark matter halos*, *arXiv e-prints* (Apr, 2019) arXiv:1904.11294, [1904.11294].

[196] F. Schmidt, G. Cabass, J. Jasche and G. Lavaux, *Unbiased Cosmology Inference from Biased Tracers using the EFT Likelihood*, *arXiv e-prints* (Apr., 2020) arXiv:2004.06707, [2004.06707].

[197] G. Cabass and F. Schmidt, *The Likelihood for LSS: Stochasticity of Bias Coefficients at All Orders*, *arXiv e-prints* (Apr., 2020) arXiv:2004.00617, [2004.00617].

[198] A. Perko, L. Senatore, E. Jennings and R. H. Wechsler, *Biased Tracers in Redshift Space in the EFT of Large-Scale Structure*, *ArXiv e-prints* (Oct., 2016) , [1610.09321].

[199] L. Fonseca de la Bella, D. Regan, D. Seery and D. Parkinson, *Impact of bias and redshift-space modelling for the halo power spectrum: Testing the effective field theory of large-scale structure*, *ArXiv e-prints* (May, 2018) , [1805.12394].

[200] V. Desjacques, D. Jeong and F. Schmidt, *The Galaxy Power Spectrum and Bispectrum in Redshift Space*, *ArXiv e-prints* (June, 2018) , [1806.04015].

[201] T. Charnock, G. Lavaux, B. D. Wandelt, S. Sarma Boruah, J. Jasche and M. J. Hudson, *Neural physical engines for inferring the halo mass distribution function*, *MNRAS* **494** (Mar., 2020) 50–61, [1909.06379].

[202] R. M. Neal, *MCMC using Hamiltonian dynamics*, *arXiv e-prints* (Jun, 2012) arXiv:1206.1901, [1206.1901].

[203] R. M. Neal, *Probabilistic inference using markov chain monte carlo methods*, tech. rep., 1993.

[204] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *Equation of State Calculations by Fast Computing Machines*, *JChPh* **21** (Jun, 1953) 1087–1092.

[205] V. I. Arnold, *Mathematical methods of classical mechanics*. 1978.

[206] R. M. Neal, *Bayesian Learning for Neural Networks*. Springer-Verlag, 1996.

[207] J. F. Taylor, M. A. J. Ashdown and M. P. Hobson, *Fast optimal CMB power spectrum estimation with Hamiltonian sampling*, *MNRAS* **389** (Sep, 2008) 1284–1292, [0708.2989].

[208] M. Betancourt, *A Conceptual Introduction to Hamiltonian Monte Carlo*, *arXiv e-prints* (Jan, 2017) arXiv:1701.02434, [1701.02434].

[209] M. Creutz, *Global Monte Carlo algorithms for many-fermion systems*, *PhRvD* **38** (Aug., 1988) 1228–1238.

[210] D. J. Venzon and S. H. Moolgavkar, *A method for computing profile-likelihood-based confidence intervals*, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **37** (1988) 87–94.

[211] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *ApJ* **538** (2000) 473–476, [astro-ph/9911177].

[212] C. Howlett, A. Lewis, A. Hall and A. Challinor, *CMB power spectrum parameter degeneracies in the era of precision cosmology*, *JCAP* **1204** (2012) 027, [1201.3654].

[213] E. Lawrence, K. Heitmann, M. White, D. Higdon, C. Wagner, S. Habib et al., *CosmicEmu: Cosmic Emulator for the Dark Matter Power Spectrum*, Oct., 2010.

[214] D. J. Eisenstein and W. Hu, *Baryonic Features in the Matter Transfer Function*, *ApJ* **496** (Mar., 1998) 605–614, [astro-ph/9709112].

[215] D. J. Eisenstein, H.-J. Seo and M. White, *On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter*, *ApJ* **664** (Aug., 2007) 660–674, [astro-ph/0604361].

[216] D. J. Eisenstein, H.-J. Seo, E. Sirko and D. N. Spergel, *Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak*, *ApJ* **664** (Aug., 2007) 675–679, [astro-ph/0604362].

[217] Y. Noh, M. White and N. Padmanabhan, *Reconstructing baryon oscillations*, *PhRvD* **80** (Dec., 2009) 123501, [0909.1802].

[218] S. Tassev and M. Zaldarriaga, *Towards an optimal reconstruction of baryon oscillations*, *JCAP* **10** (Oct., 2012) 006, [1203.6066].

[219] A. Burden, W. J. Percival and C. Howlett, *Reconstruction in Fourier space*, *MNRAS* **453** (Oct., 2015) 456–468, [1504.02591].

[220] M. Schmittfull, Y. Feng, F. Beutler, B. Sherwin and M. Y. Chu, *Eulerian BAO reconstructions and N -point statistics*, *PhRvD* **92** (Dec., 2015) 123522, [1508.06972].

[221] X. Wang, H.-R. Yu, H.-M. Zhu, Y. Yu, Q. Pan and U.-L. Pen, *Isobaric Reconstruction of the Baryonic Acoustic Oscillation*, *ApJL* **841** (June, 2017) L29, [1703.09742].

[222] R. Hada and D. J. Eisenstein, *An iterative reconstruction of cosmological initial density fields*, *MNRAS* **478** (Aug., 2018) 1866–1874.

[223] Y. Shi, M. Cautun and B. Li, *New method for initial density reconstruction*, *PhRvD* **97** (Jan., 2018) 023505, [1709.06350].

[224] S. Tassev, *N-point statistics of large-scale structure in the Zel'dovich approximation*, *JCAP* **6** (June, 2014) 012, [1311.6316].

[225] B. D. Sherwin and M. Zaldarriaga, *Shift of the baryon acoustic oscillation scale: A simple physical picture*, *PhRvD* **85** (May, 2012) 103523, [1202.3998].

Acknowledgment

The work presented here represents my PhD journey over the last three years and a half. I now would like to thank the people who have collectively made this journey so much more memorable.

Firstly, I cannot express enough the gravity of my gratitude towards my advisor, Fabian Schmidt, for providing all the guidance, support, sympathy, and friendship. At the same time, he is always intimidatingly sharp which has made me hold myself to a higher standard. I really appreciate his infinite patience to answer even the most absurd questions of mine with the utmost seriousness, his exceptional ability to balance attentiveness without being overbearing and distance without seeming uninterested established a research environment which I have enjoyed tremendously.

Secondly, I must thank Eiichiro Komatsu, my formal advisor, not only for welcoming me into such a diverse and energetic research group at MPA, but also for his leadership and support. His passion for astrophysics and cosmology is second to none and many times a fire has been lit in me just by watching and listening to him talks about physics. I really enjoy the cosmology lecture series he has been organizing at MPA.

Thirdly, I want to thank my collaborators, Jens Jasche, Guilhem Lavaux and Franz Elsner for our fruitful collaboration. I would attribute half of my knowledge in programming and statistics to Guilhem and Jens, respectively. Franz was always kind and supportive.

I would also like to thank all the people of our LSS and cosmology group at MPA. In particular, I would like to thank Titouan Lazeyras, Aniket Agrawal, and Alex Barreira for the fruitful and cheerful discussions we have had. They and the rest of our group, including (not exhaustingly) Giovanni Cabass, Kaloian Lozanov, Linda Blot, Xun Shi, Chi-Ting Chiang, Chris Byrohl, have been excellent company.

I am also grateful to my family, especially my mother Nga, for starting life in great hardship and continued in that fashion so that my siblings and I can have something they never had, a choice with what to do with our lives. You have have been my foundation, my solace and my determination. Everything I am today is because of you.

Lastly but most importantly, I must thank the woman paramount to my happiness, Thanh, for her understanding, patience, sympathy and love. Thank you for putting up with me during this journey, for always being there to show support. Your smile is all I need to get through any speed bump or major setback along the way. Without you, this journey would not have been worth it.