UNIVERSITÀ
DI PARMA

Università degli studi di Parma

Dottorato di Ricerca in Fisica

Ciclo XXXVIII

Towards Optimal Extraction of Cosmological Information from the Large Scale Structure

Coordinatore:
Chiar.ma Prof.ssa Raffaella Burioni

Dottorando:
Matteo Peron

Tutore:
Chiar.mo Prof. Massimo Pietroni

Anni Accademici 2022/2023 – 2024/2025

To look up to the high mountains. . .

Acknowledgements

I would like to express my sincere gratitude and respect to my supervisor, Massimo Pietroni, for the guidance, support, and all the opportunities for growth he provided to me through this journey. He is a great mentor who is understanding, approachable and graceful, yet harsh when truly needed. As my PhD nears its epilogue, I hope I can continue to count on him as a friend and as a collaborator whenever new opportunities will arise.

I want to express my gratitude to all my collaborators as well. Michele Liguori in particular has been the catalyst for my ever-growing interest in the field of Bayesian statistics, and (willingly or not) he steered my life in directions which were in equal amounts unexpected and yet meaningful. Takahiro Nishimichi and Atsushi Taruya both have offered great insight and deepened my understanding of physics and statistics beyond what would've been possible without them, on top of making my time in Japan memorable (if the reader has the chance to go to karaoke with them, don't refuse).

I also want to thank all the people I met in Parma for giving me so many good memories. I shared many good moments and useful discussions with my peer, flatmate, and friend, Dennis Linde, despite him threatening my liver multiple times. Marco Marinucci, Guido d'Amico, all my fellow office coworkers, and many more provided a bounty of good times and opportunities to learn. I found many of these with all the people I had the pleasure to meet at conferences all around the world; even if I can't name everyone here, the memories will live on with me.

A sincere thank you is reserved to Marko Simonovic and Fabian Schmidt, who accepted to read and correct this manuscript, their validation has been a great reassurance.

Then, a most important thank you goes to my closest friends, for their moral support, and to my mamma and papà, and siblings, for their patience and genuine love. Their presence has been a pillar that guided me through difficult times and sustained me through much needed love. And as for love, a final thank you goes to my dear partner, Li Xing; I would've never met her if I didn't embark on this adventure, and her presence in my life makes it all the more special.

Publications and Permissions

List of published articles

1. Matteo Peron, Andrea Ravenni, Sarah and Libanore, Michele Liguori and Maria Celeste Artale (March 2024). "Clustering of binary black hole mergers: a detailed analysis of the eagle + mobse simulation." *Monthly Notices of the Royal Astronomical Society* **530**(1): 1129-1143. doi:10.1093/mnras/stae893.

2. Matteo Peron, Gabriel Jung, Michele Liguori and Massimo Pietroni (July 2024). "Constraining primordial non-Gaussianity from large scale structure with the wavelet scattering transform." *Journal of Cosmology and Astroparticle Physics* **2024**(07): 021. doi:10.1088/1475-7516/2024/07/021.

3. Matteo Peron, Takahiro Nishimichi, Massimo Pietroni and Atsushi Taruya (October 2025). "Renormalized perturbation theory at field-level: the LSS bootstrap in GridSPT." *Journal of Cosmology and Astroparticle Physics* **2025**(10): 098. doi:10.1088/1475-7516/2025/10/098.

Copyright permission statement

I confirm that permissions to reuse the published material listed above have been obtained and that reuse complies with the publishers' policies.

Abstract

Forthcoming galaxy redshift surveys such as DESI and Euclid demand analysis pipelines that translate statistical precision on non-linear scales into robust cosmological inference. This thesis advances two complementary directions toward near-optimal information extraction from large-scale structure. First, it develops and evaluates wavelet scattering transforms (WST) as compact, interpretable summaries that capture non-Gaussian information beyond the power spectrum. Using the public Quijote and QuijotePNG simulations, Fisher forecasts are produced for standard cosmological parameters and the amplitudes of primordial non-Gaussianity (local, equilateral, orthogonal), and are benchmarked against power-spectrum-only and joint power–bispectrum analyses. With a controlled $k_{\max}$ filtering and transparent hyperparameter choices, WST is shown to outperform the power spectrum and to be competitive with state-of-the-art non-Gaussian summaries while being interpretable. Second, the thesis formulates a renormalized perturbative forward model for field-level inference based on the bootstrap approach to large-scale structure dynamics. By making the ultraviolet cutoff explicit and using a Wilsonian approach to take into account the inevitable discretization effects induced by a finite grid, the framework yields "bootstrap" coefficients which are independent on grid size effects, and identifies the higher-derivative operators required for unbiased recovery of these coefficients at fifth (and third) perturbative order. The results clarify how theoretical control at finite resolution enables principled field-level likelihoods and prepares the ground for realistic extensions. The final chapter outlines ongoing steps toward survey observables: incorporation of redshift-space distortions within the renormalized bootstrap framework and full marginalization over initial conditions with efficient high-dimensional samplers. Together, task-specific compression and controlled forward modeling provide a coherent path to stringent tests of fundamental physics with next-generation data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A leap forward in the precision and scope of large-scale structure (LSS) observational campaigns is expected to happen in the near future. Forthcoming galaxy redshift surveys, notably DESI and Euclid [1–3], will map cosmic structure over volumes that render statistical errors small and put systematic errors and information-efficient analyses as primary concerns. Within the standard cosmological model, ΛCDM, the late-time distribution of matter arises from initially nearly Gaussian perturbations seeded during inflation and subsequently amplified by gravitational collapse. LSS observables such as the baryon acoustic oscillation feature in two-point statistics have already been used to deliver stringent constraints on expansion history and the growth of structure. Yet many open questions remain: the nature of dark matter and dark energy, potential deviations from general relativity on cosmic scales, the mass of neutrinos, and the possibility of primordial non-Gaussianity (PNG). The unifying theme of this thesis is to address some of these open questions and come close to optimal extraction of cosmological information from the LSS by exploring novel statistical compression methods and field-level approaches, pushing past the limitations of traditional two-point analyses while retaining interpretability and theoretical control.

A central insight motivating this program is that the power spectrum, though foundational, cannot capture the phase information that encodes the geometry of the cosmic web: its filaments, sheets, and nodes. The bispectrum is added to include non-Gaussian information and has been incorporated into joint power spectrum and bispectrum (P+B) analyses [4–8], aided by perturbative models valid on mildly non-linear scales. However, the bispectrum does not exhaust the information content; extending to higher n-point correlators is hampered by rapidly growing dimensionality, challenging covariance estimation, and complex survey window effects. This tension between information content and complexity has driven the exploration of alternative non-Gaussian summaries, like

marked statistics [9–13], skew spectra [14–17], Minkowski functionals [18, 19], halo mass function probes [20–25], and void abundances [26, 27]. Against this landscape, the *wavelet scattering transform* (WST, [28, 29]) has emerged as a compelling candidate: it is a composition of wavelet convolutions and modulus non-linearities that forms a compact multi-scale summary with a connection to Fourier-space correlators. Recent simulation studies have shown that WST can substantially shrink constraints relative to power spectrum-only analyses, including in scenarios with massive neutrinos [30–34]. Another approach is to abandon entirely the idea of compressing the information present in matter clustering observables, and instead perform statistical inference directly on the field of density perturbations [35–50]. This *field-level inference* is expected to deliver maximum cosmological information, but requires very fine control over the theoretical and statistical modeling process, and high computational performance. These two approaches, summary statistics that compress the information in the field and field-level, uncompressed analyses, are not necessarily opposed to each other. Summary statistics remain attractive because of their computational efficiency and resilience to systematics; they highlight scales and features where most of the information resides, and prove to be great diagnostic tools in more elaborate approaches. Field-level analyses connect to theory at a more fundamental level, enabling more interpretable results and more powerful exploration of alternative models. Indeed, works such as Schmidt [49] explore directly the connection between the two.

From this broad context, this thesis is aligned in the direction of exploring these advanced statistical tools to extract information with high precision on new physics. In particular, of the two works presented here, the first will address the problems of detecting PNG with the WST, since by construction the WST includes non-linear operations that can capture non-Gaussian information from a random field. The performance of the WST is quantified numerically to overcome difficulties with interpretability, however an analysis on its connection to N-point correlators is also explored. The second work applies the bootstrap approach of LSS [51], which derives perturbative kernels consistent with extended Galilean invariance and other symmetries, and introduces time-dependent coefficients that can be constrained to probe departures from the ΛCDM model of cosmology. The resulting generalized perturbation theory model is then applied at field-level in a simple setting to study the subtleties in renormalizing the theory when applied to discrete fields. This latter direction is then further explored to introduce survey realism in the form of redshift-space distortions (RSD) into the model, and finally complete the implementation of the full field-level analysis by enabling sampling and marginalization over the modes of the initial field. By advancing both of these approaches, this work aims to deliver practical advances, such as algorithms, models and diagnostics, that can move LSS analysis toward optimality while maintaining the theoretical discipline necessary for robust inference in the non-linear Universe.

The chapter are delineated as follows. Chapter 2 introduces the physical models at the foundation of LSS cosmology, from the ΛCDM model that characterizes the evolution of the background metric, to cosmic inflation as the engine of density perturbations, to finally the evolution of these perturbation via perturbation theory. Chapter 3 follows with a tour of the statistics background that justifies the methodology employed by later chapters, including parametric inference, uncertainty quantification and advanced sampling methodologies. In chapter 4 the first work on WST is layed out: in it, the constraining power of WST is compared to conventional P and P+B analyses via the Fisher information matrix; the $f_{\rm NL}$ parameter for standard PNG shapes, local, equilateral and orthogonal, are constrained together with cosmological parameters for dark matter and halo density fields in real and redshift space from the public QUIJOTEPNG suites of N-body simulations, and a connection to standard correlators is explored for a specific set of WST. Chapter 5 begins from the bootstrap approach to generalize the standard perturbative model of matter clustering, and the recognition that an ulterior expansion of the model on discrete fields to include an effective-theoretical treatment of small-scale effects would induce systematics in the inferred parameters; by adopting a Wilsonian perspective, the systematic *running* can be put under perturbative control, and an explicit renormalization procedure is laid out for models at third and particularly fifth order, where higher-derivative operators play a crucial role; field-level inference is finally carried out at fixed initial conditions in real space, demonstrating the effectiveness of the renormalization scheme. Then, chapter 6 introduces RSD systematics within the field-level forward model and presents the work done to prepare the ground for a full field-level analysis that samples and marginalizes the high-dimensional posterior distribution over the amplitudes and phases of the initial field; this requires high performance computing applied to the forward model, and advanced sampling techniques like Hamiltonian Monte Carlo, coupled with novel computational methods such as automatic differentiation and, possibly, machine learning methods. The work concludes with chapter 7 by summarizing the presented work, its limitatiotion and future developments.

Chapter 2

Cosmology from the large-scale structure

Cosmology is concerned with describing the dynamical evolution of the Universe at the largest of scales. There are several observables that can be measured in order to infer cosmological theory, and the *large-scale structure* (LSS), the distribution of matter observable (directly or indirectly) at scales larger than virialized systems, such as galaxy clusters, is the focus of this dissertation. While the hydrodynamics of gas and radiative processes govern the dynamics of galaxies and star systems, these become far less influential on cosmic scales; instead gravity dominates, driving the evolution of the LSS. Gravity is described by general relativity, so the theory provides a natural foundation for the description of the LSS. The chapter starts by giving a brief overview of the theoretical framework that defines modern cosmology in section 2.1; beginning from its core proposition, the cosmological principle, the notion of an expanding Universe and how general relativity can be used to describe its evolution is introduced. Section 2.2 briefly introduces the hypothesized phenomenon of cosmic inflation, an extension of standard cosmological theory that addresses many inconsistencies (the most notable being the "horizon" and the "flatness" problems), and provides a natural mechanism for the origin of the primordial perturbations that seeded the structures observed today; the dynamical evolution of these perturbations under gravity is then expanded in section 2.3. For more detailed treatments of the cosmological background and perturbations in the matter density field, the interested reader may consult Dodelson and Schmidt [52], Hartle [53], Bernardeau et al. [54], and Desjacques, Jeong, and Schmidt [55].

2.1 The Λ cold dark matter model

The assumption that underpins modern cosmology is the *cosmological principle*. The cosmological principle states that the Universe, when viewed on sufficiently large scales, appears to have the same properties for all observers. This implies that there is no privileged position or direction, the Universe is invariant under translations (homogenous) and rotations (isotropic). Isotropy is based on strong observational evidence, namely the mapping of the cosmic microwave background and of the distribution of galaxies across the sky coming from many surveys, like WMAP [56–59], Planck [60, 61], and the SDSS [62, 63]. Homogeneity can only be stated if the Universe appears isotropic for all comoving observers, however this cannot be experimentally verified, so enforcing the full cosmological principle implies the assumption that our frame of reference is not special.[1] Another important piece of evidence that a model of the Universe must take into account is the observation that the Universe is expanding, a fact first noticed by Edwin Hubble while observing the redshift of distant galaxies. Starting from Einstein's theory of general relativity, the spacetime geometry of a homogenous and isotropic Universe is given by the Friedmann-Lemaître-Robertson-Walker metric $g_{\mu\nu} = \mathrm{diag}(-1, a^2(t), a^2(t), a^2(t))$, where $a(t)$ is the *scale factor* that encodes the expansion of space. Note that a flat spatial geometry was assumed for simplicity. This is the first ingredient needed to solve Einstein's field equations and find the exact behavior of $a(t)$. The other ingredient is the stress-energy tensor $T_{\mu\nu}$, which specifies the matter-energy content of the Universe. Einstein's equations read:

$$G_{\mu\nu} = \frac{1}{M_p^2} T_{\mu\nu} + \Lambda g_{\mu\nu} \,, \tag{2.1}$$

where $M_p \equiv (8\pi G)^{-1/2}$ is the reduced Planck mass, with G being the gravitational constant, $G_{\mu\nu}$ is the Einstein tensor, which is completely described by the metric $g_{\mu\nu}$, and Λ is the cosmological constant. Assuming that the matter-energy content of the Universe behaves like a perfect fluid, $T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p g^{\mu\nu}$, where ρ, p and u^μ are density, pressure and four-velocity of the fluid respectively. From equation (2.1) and the conservation of energy, $\nabla_\mu T^{\mu 0} = 0$ (∇_μ is the covariant derivative), two independent equations can be recovered,

[1] This assumption takes the name of *Copernican principle*. Strictly speaking, spatial homogeneity cannot be established without some cosmological assumptions, since mapping redshifts to distances is model dependent; however, it can be tested under mild assumptions, for example by verifying large-scale clustering statistics across a wide range of fiducial cosmologies. For an example of how such tests are carried out, see Gonçalves et al. [64].

$$H^2 = \frac{1}{3M_p^2}\rho + \frac{\Lambda}{3}, \tag{2.2}$$

$$\dot{\rho} = -3H\left(\rho + p\right) . \tag{2.3}$$

These are known as the *Friedmann equations*, and fully describe the dynamical evolution of space-time and matter-energy content of the Universe. To be more specific, the first equation details the balance between the expansion of the Universe, given by the *Hubble rate* $H(t) \equiv \dot{a}(t)/a(t)$ (a dot denotes the time derivative), the contraction due to the density of matter $\rho(t)$, and the effect of Λ, which is currently interpreted as the energy density term of *dark energy* [65], and has the effect to accelerate the expansion. Before continuing it is useful to rewrite equations (2.2) and (2.3) by defining some quantities. Let $\rho_\Lambda \equiv M_p^2\Lambda$ be a density term associated to dark energy, then the term can be absorbed into $\rho_c = \rho + \rho_\Lambda$, and equation (2.2) can be rewritten as $H^2 = \rho_c/(3M_p^2)$; this gives the definition of the *critical density*, $\rho_c \equiv 3H^2M_p^2$. Just like ρ_Λ can be defined for dark energy, ρ_i can be defined for the i-th species that contributes to the total $\rho = \sum_i \rho_i$. The density parameters today are then defined as $\Omega_i \equiv \rho_{i,0}/\rho_{c,0}$. The sum of Ω_i (which now includes $i = \Lambda$) is equal to 1, so the density parameters determine the fractional contribution to the total density budget from the different constituents of the Universe. Furthermore, it is often customary to define the dimensionless Hubble constant $h \equiv H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$. The main constituents at present day are matter Ω_m and dark energy Ω_Λ. Other contributions are given by photons and neutrinos, Ω_γ and Ω_ν, which today are not as relevant as they were at earlier times. Equation (2.3) can be written in terms of an equation of state $w = p/\rho$, which assumes a unique value for each of the above constituents:

$$\dot{\rho} = -3H\left(1 + w\right)\rho\,. \tag{2.4}$$

This differential equation is solved by separation of variables and taking some initial density $\rho_{\mathrm{in}} \equiv \rho(t_{\mathrm{in}})$. Using the definition for the density parameters:

$$\Omega_i(t) = \frac{\rho_{in}}{\rho_c}\left(\frac{a(t)}{a_{in}}\right)^{-3(1+w_i)} , \tag{2.5}$$

where w varies based on the constituents $i \in \{m,\, \gamma,\, \nu,\, \Lambda\}$. For non-relativistic matter $w_m = 0$, for radiation and neutrinos $w_\gamma = w_\nu = 1/3$ and for dark energy $w_\Lambda = -1$. Equations (2.2) and (2.3) can be solved to find the complete expression for $a(t)$. In particular, the Hubble rate $H(z)$ can be defined from $a \equiv (1+z)^{-1}$ in the following form:

$$H^2(z) = H_0^2 \left[\Omega_m(1+z)^3 + \Omega_\gamma(1+z)^4 + \Omega_\nu(1+z)^{3(w_\nu(z)+1)} + \Omega_\Lambda\right] , \tag{2.6}$$

where, for neutrinos, the time-dependent effective equation of state interpolates between 1/3 and 0. It is instructive to see how the scale factor evolves in time when either matter, radiation or dark energy dominate the expansion of the Universe:

$$\text{matter-dominated: } a(t) \propto t^{1/2} \qquad \rho \propto a^{-3}, \tag{2.7}$$

$$\text{radiation-dominated: } a(t) \propto t^{2/3} \qquad \rho \propto a^{-4}, \tag{2.8}$$

$$\text{dark energy-dominated: } a(t) \propto \exp\left(t\sqrt{\Lambda/3}\right) \qquad \rho = \text{const.} \tag{2.9}$$

Current measurements [61, 66] show that the Universe is expanding in an accelerated way, with $\Omega_{m,0} \approx 0.32$, $\Omega_{\gamma,0} \approx 10^{-4}$ and $\Omega_{\Lambda,0} \approx 0.68$ at t_0, and is therefore largely dominated by dark energy. An estimation of the redshift z at which one component started to dominate over the other can be found by equating the various components: for matter, the ratio of $\Omega_m(t)/\Omega_{m,0} = [a(t_0)/a(t)]^3 = a(t)^{-3}$, since the scale factor can be rescaled so that $a(t_0) \equiv 1$. For dark energy instead $\Omega_\Lambda(t)/\Omega_{\Lambda,0} = 1$, so at t for which $\Omega_m(t) = \Omega_\Lambda(t)$:

$$a(t) = \left(\frac{\Omega_{\Lambda,0}}{\Omega_{m,0}}\right)^{-1/3} \approx 0.78\,, \tag{2.10}$$

dark energy began dominating the energy balance of the Universe at around $z \approx 0.29$ (corresponding to 3.4 Gyr of lookback time). Repeating the same procedure for matter and radiation, the former began dominating over the latter at $z \approx 3\,199$ (13.7 Gyr lookback time). So the Universe, according to the Big Bang model, was initially dominated by radiation, then matter, and finally dark energy.

The results detailed here are known as the standard model of cosmology, or the Λ *cold dark matter model*. The "Λ" is the cosmological constant appearing in Einstein's field equations, however the "cold dark matter" part is not explained by any of the above equations. This is because it arises from observational evidence, in particular from comparison between Big Bang nucleosynthesis (BBN) predictions and observed deuterium abundances [67], and the cosmic microwave background (CMB) [61]; it is found that $\Omega_{m,0} \approx 0.05$, however this amount of matter cannot explain several empirical facts like

- the velocity curves of spiral galaxies, which they were found to require significantly more gravitational mass than the stars and gas in the system;

- the amount of observed gravitational lensing around galaxy clusters requires much more matter than the observed luminous one;
- both the growth of structures of LSS and the anisotropies of the CMB require $\Omega_{m,0} \approx 0.3$.

So there must be an extra $\approx 25\%$ of the matter-energy content represented by some type of matter that is "dark", as it interacts mainly through gravity. In ΛCDM, dark matter is "cold" because it is considered to be made up of non-relativistic particles; the reason for this is that gravitationally bound structures below a certain scale, such as observed galaxy clusters, could not form if dark matter was relativistic, dark matter particles would just escape the gravitational potential before the density can grow large enough. For a recent review on the evidence and state of the research on dark matter, the interested reader can consult Balazs et al. [68] and references therein.

2.2 Cosmic inflation

The notion of a Universe that expands over time leads to a trivial but important observation. By running time backwards, one can expect that the Universe shrinks, with all the matter observed today coming progressively closer together; eventually all, points in space, and so all matter, would collapse into a physical singularity at $t = 0$. While current knowledge of physics breaks down as it approaches this singularity, it is clear that the Universe, at some point in its evolution, was very dense and hot. This is the idea behind the so called "Big Bang model", which was first proposed by G. Gamow [69], and accepted as part of current cosmological models with the confirmation of the predictions of the associated theory for primordial nucleosynthesis, the BBN, and, later, the discovery of the CMB [70], the relic radiation left over from the primordial plasma that filled the early Universe. During the epochs in which radiation and matter dominated, the Universe's expansion was decelerated, as can be seen by taking the second derivative of the scale parameter for a matter-dominated (2.7) and radiation-dominated (2.8) Universe. This fact however is in contrast with observations: when we look at the CMB in different directions, we see that the deviation in the observed temperature with respect to the mean never exceeds one part in 10^5, which is consistent with radiation generated from a plasma in thermal equilibrium. In order for this to be the case, the entire Universe had to be causally connected, that is photons had to be able to travel the entire span of the Universe up until the CMB formed. However the CMB is observed at $z \approx 1\,100$, and under the ΛCDM model, the distance photons could have covered at that epoch would only extend around one degree in the sky we observe today, so points separated

by angular distances greater than that could not have been causally connected. This is known as the "horizon problem". To solve this and other issues, such as the fine-tuning that would be required to otherwise explain why the Universe is observed to have flat curvature, one can postulate an initial phase of rapid accelerated expansion, also know as *cosmic inflation* [71]. What follows describes only a subset of simple models that still fit current observational evidence, those that describe *single field inflation*.

Similarly to matter, radiation and dark energy, it is possible to postulate that the Universe was dominated by yet another component during its early stage. Using the formalism of field theory, this component can be identified in a scalar field ϕ, called the *inflaton* field, which evolves according to some potential $V(\phi)$. The simplest covariant action that gives rise to Einstein's equations is the Einstein-Hilbert action, and when this is combined with the simplest possible action for a scalar field like the inflaton, then the total action in curved spacetime takes the form of

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{c^4}{16\pi G}R - \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right) . \tag{2.11}$$

From the principle of least action, varying S with respect to $g_{\mu\nu}$ yields Einstein's field equations, for which now the stress-energy tensor $T_{\mu\nu}$ is sourced by the inflaton:

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi + V(\phi)\right) . \tag{2.12}$$

Setting the metric to the FLRW metric, and assuming that the inflaton only depends on time, $\phi \equiv \phi(t)$, the comparison of the stress energy tensor in equation (2.12) with a perfect fluid yields a pressure and density term for the inflaton,[2]

$$p_\phi = \frac{\dot\phi^2}{2} - V(\phi) \qquad \text{and} \qquad \rho_\phi = \frac{\dot\phi^2}{2} + V(\phi)\,, \tag{2.13}$$

where $\dot\phi \equiv \partial\phi/\partial t$. The Friedmann equations are

$$H^2 = \frac{1}{3M_p^2}\rho_\phi, \tag{2.14}$$

$$\dot H = -\frac{1}{2M_p^2}\dot\phi^2\,, \tag{2.15}$$

and, when combined, they give an expression for the acceleration of the scale factor:

[2]Note that pressure and density here are given in units of energy density. See main text.

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = \frac{1}{3M_p^2}\left(V(\phi) - \dot{\phi}\right) ; \tag{2.16}$$

in order for accelerated expansion to occur, $\ddot{a} > 0$, so $V(\phi) > \dot{\phi}$: the inflaton potential must dominate the dynamics of the scalar field. Going back to equation (2.11), varying S with respect to the inflaton itself ϕ yields the *Klein-Gordon equation*,

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 , \tag{2.17}$$

where $\phi \equiv \phi(t)$ is implied, and $V'(\phi) \equiv \mathrm{d}V/\mathrm{d}\phi$. Equation (2.17) is analogous to the dynamics of a particle subjected to a force supplied by the potential $V(\phi)$ and a friction term $3H\dot{\phi}$ supplied by the expansion of the Universe through the Hubble rate H. If the expansion is fast enough, the friction term dominates over the acceleration and damps the dynamical evolution of $\phi(t)$, leading to the so-called *slow-roll regime*. During slow-roll, as the name implies, the inflaton slowly rolls down the potential until it settles at its minimum. This means that $\ddot{\phi} \approx 0$, and

$$V'(\phi) \approx -3H\dot{\phi} . \tag{2.18}$$

Since the potential is nearly flat, the above expression implies that for large values of H, $\dot{\phi}$ must also be very small. In inflationary theories, it is customary to use the *slow-roll parameters* to describe inflation quantitatively:

$$\epsilon_H = -\frac{\dot{H}}{H^2} \qquad \text{and} \qquad \eta_H = -\frac{\ddot{\phi}}{H\dot{\phi}} , \tag{2.19}$$

both of which are much smaller than unity during slow-roll, $|\epsilon_H| \ll 1$ and $|\eta_H| \ll 1$. Inflation ends when the inflaton approaches the minimum of the potential; there $|\epsilon_H| \approx 1$, and the inflaton acquires a non-negligible amount of kinetic energy while it starts to oscillate around the minimum, following the motion of a damped harmonic oscillator. The transition to a radiation-dominated Universe soon after this phase requires that the inflaton must be coupled to other quantum fields in the standard model, so that the energy dissipated from it can go into particle generation: this process takes the name of *reheating* [72–74]

2.2.1 The origin of the large scale structure

The discussion so far assumed that the inflaton ϕ is uniform in space and only evolves in time. However ϕ is a quantum field subject to quantum dynamics, which predict the presence of vacuum fluctuations $\delta\phi(\vec{x},\, t)$ around some background,

$$\phi(\vec{x},\, t) = \bar{\phi}(t) + \delta\phi(\vec{x},\, t)\,. \tag{2.20}$$

Decomposing the fluctuations with a Fourier transform, one sees that each phase $\vec{k}\cdot\vec{x}$ of the plane waves in comoving coordinates can be transformed to physical coordinates $\vec{k}_{\rm phy}\cdot\vec{r}$ using the fact that $\vec{r} = a(t)\vec{x}$ and $\vec{k}_{\rm phy} = \vec{k}/a(t)$. This highlights how the physical wavenumbers of the fluctuations shrink as the Universe expands following the inflationary dynamics sketched previously. As long as the modes have physical wavelength within the *Hubble horizon* H^{-1}, they continue to oscillate like an harmonic oscillator, but the picture changes once the wavelength of each Fourier mode crosses the horizon. There the mode "freezes" at constant amplitude, because the Hubble horizon represents the region where events are causally connected. Since the presence of these fluctuations causes $a(t)$ to grow at slightly different rates at different points in space $\vec{x}$, then the fluctuation translate directly into curvature perturbations $\zeta(\vec{x})$ [75]. Once inflation ceases and the Universe goes through reheating, the frozen modes re-enter the horizon as it slowly grows, and since their amplitude remained constant, they provide the initial conditions for density perturbations to grow from. In the matter-dominated era, the curvature perturbations are related to the Newtonian potential $\Phi(\vec{k})$ by

$$\Phi(\vec{k}) = \frac{3}{5}T(k)\zeta(\vec{k})\,, \tag{2.21}$$

where $T(k) = 1$ for $k \ll k_{\rm eq}$, $k_{\rm eq} \approx 0.02h\,{\rm Mpc}^{-1}$ being the frequency that entered the horizon at matter-radiation equality. The density of matter $\rho(\vec{k})$ can be expressed in terms of its fluctuations $\delta(\vec{k})$ around the average over a volume,

$$\delta(\vec{k}) \equiv \frac{\rho(\vec{k})}{\bar{\rho}} - 1\,, \tag{2.22}$$

so with this, one obtains an expression relating the density fluctuations, or *density contrast*, $\delta(\vec{x})$ with the curvature perturbations by plugging in equation (2.21) into the Poisson equation:

$$\delta(\vec{k},\, a) = \frac{2}{5}\frac{k^2}{H_0^2\Omega_{m,0}}T(k)D(a)\zeta(\vec{k}) \equiv \mathcal{M}(\vec{k},\, a)\zeta(\vec{k})\,. \tag{2.23}$$

$D(a)$ is called the linear growth factor, which will be introduced in section 2.3.1. The inflationary dynamics described in section 2.2 predict that the curvature perturbations are nearly Gaussian,[3] so the information of $\delta(\vec{k},\, a)$ at early times is entirely summarized in its power spectrum.

A significant degree of *primordial non-Gaussianity* can be present in case of more complex inflationary scenarios (e.g., multi-field inflation, "ultra-slow roll" regimes, fluctuations from non-adiabatic vacuum states, for a review see [71]) and can be detected by observing different shapes of the bispectrum. The power spectrum and bispectrum will be discussed in section 2.3.4. Current results from observations of the CMB have detected no significant degree of primordial non-Gaussianity, greatly restricting the space of viable models of inflation [77, 78].

2.3 Perturbation theory of density fluctuations

The previous section 2.2 introduced cosmic inflation and how it can produce fluctuations in the density field of matter. These fluctuations are tiny, $\delta(\vec{x}) \approx 10^{-5}$, but non-zero, and so they can grow at later times thanks to gravitational collapse: the overdense regions begin to accumulate more and more matter in bound structures which end up forming the familiar complex of galaxy clusters, walls and filaments observed presently in the LSS. At the end of section 2.1 cold dark matter was established as the major contributor to the matter content of the Universe, and so it is the main driver when it comes to gravitational collapse [79]. Furthermore, baryonic matter is coupled to dark matter through gravity, so studying the evolution of the dark matter density field is crucial to understand the observed galaxy distribution. For this reason, the following discussion refers only to dark matter.

If perturbations are small, $\delta(\vec{x}) \ll 1$, perturbation theory can be applied to obtain an accurate description of the density field. In the study of the LSS, this approach is often referred to as Standard Perturbation Theory (SPT). Cosmological models predict that the variance in the perturbations is scale-dependent, being small at large scales and large at small scales [80, 81]. This suggests the distinction between two regimes:

[3] This is true when the action for the curvature perturbations is truncated to second order. Non-Gaussian contributions are added by higher order terms, but these are suppressed during slow-roll [76].

- the quasi-linear regime, where perturbation theory can, up to sufficiently high order, describe the clustering of matter;
- the non-linear regime, where perturbation theory fails to accurately characterize the density field.

Therefore, using a perturbative framework means restricting the theory to large scale clustering, where the quasi-linear regime is valid. Defining k_{NL} as the scale that separates these two regimes, it is found that k_{NL} shrinks as the distribution of matter evolves in time under the effect of gravity, and this reduces the span of scales that can be probed with SPT.

2.3.1 Standard perturbation theory

Assume that matter is distributed in the Universe according to a certain distribution function $f(\vec{r},\, \vec{v},\, t)$ that depends on time t, space $\vec{r}$ and velocity $\vec{v}$. This can be thought of as the continuous limit of an ensemble of collisionless particles interacting via gravity, so a conservative system that obeys Hamiltonian dynamics. As such, by the Liouville theorem, $f(\vec{r},\, \vec{v},\, t)$ is conserved:

$$\begin{aligned} \frac{\mathrm{d}f}{\mathrm{d}t} &= \frac{\partial f}{\partial t} + \frac{\partial \vec{r}}{\partial t} \cdot \vec{\nabla}_r f + \frac{\partial \vec{v}}{\partial t} \cdot \vec{\nabla}_v f \\ &= \frac{\partial f}{\partial t} + \vec{v} \cdot \vec{\nabla}_r f - \vec{\nabla}_r \Phi \cdot \vec{\nabla}_v f = 0\,. \end{aligned} \tag{2.24}$$

This partial differential equation is called the Vlasov (or collisionless Boltzmann) equation, and together with the Poisson equation,

$$\nabla_r^2 \Phi(t,\, \vec{r}) = 4\pi G \rho(t,\, \vec{r}) = \frac{1}{2M_p^2} \int \mathrm{d}^3 v\, f(\vec{r},\, \vec{v},\, t)\,, \tag{2.25}$$

it describes the time evolution of the distribution of matter in the Universe. At the scales of interest, much smaller than the Hubble horizon and much larger than the typical size of individual galaxy clusters, relativistic effects are negligible so the Newtonian description accurately describes the force of gravity. Since the Universe is expanding, equations (2.24) and (2.25) are better expressed in terms of comoving quantities. The change from physical coordinates $(\vec{r},\, \vec{v},\, t)$ to comoving coordinates $(\vec{x},\, \vec{w},\, t)$ occurs by the following change of variables:

$$\vec{x} = \frac{\vec{r}}{a(t)}\,, \tag{2.26}$$

$$\vec{w} = \vec{v} - \dot{a}\vec{x} = a\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}\,, \tag{2.27}$$

where the determinant of the Jacobian is $a^3(t)$, so $\mathrm{d}t\,\mathrm{d}^3r\,\mathrm{d}^3v = a^3(t)\mathrm{d}t\,\mathrm{d}^3x\,\mathrm{d}^3w$. Given the above transformation, it is convenient to redefine the phase-space distribution in comoving coordinates as $F(\vec{x},\,\vec{w},\,t) \equiv a^3(t)f(\vec{x},\,\vec{w},\,t)$, so that no factors of $a(t)$ appear during integration in the new variables. Splitting the Newtonian potential into a background bulk term $\Phi_{\rm bg}$, and a term given by the density fluctuations ϕ, the Vlasov and Poisson equations acquire the forms

$$\frac{\partial F}{\partial t} + \frac{\vec{w}}{a}\cdot\vec{\nabla}_x F - \left(\frac{1}{a}\vec{\nabla}_x\phi + H\vec{w}\right)\cdot\vec{\nabla}_w F = 0\,, \tag{2.28}$$

$$\nabla_x^2\phi(\vec{x},\,t) = \frac{1}{2M_p^2}a^2(t)\bar{\rho}(t)\delta(\vec{x},\,t)\,. \tag{2.29}$$

The Vlasov equation can be turned into an infinite system of differential equations expressed in terms of velocity momenta [82],

$$M^{(n)}_{i_1\cdots i_n}(\vec{x},\,t) = \int \mathrm{d}^3w\, w_{i_1}\cdots w_{i_n} F(\vec{x},\,\vec{w},\,t)\,. \tag{2.30}$$

The system must be then truncated to a given order n. From the first three momenta one finds the density $\rho(\vec{x},\,t)$, the bulk peculiar velocity $\vec{u}(\vec{x},\,t)$ and the velocity dispersion tensor $\sigma_{ij}(\vec{x},\,t)$:

$$\rho(\vec{x},\,t) = M^{(0)}\,, \tag{2.31}$$

$$u_i(\vec{x},\,t) = \frac{M^{(1)}_i}{M^{(0)}}\,, \tag{2.32}$$

$$\sigma_{ij}(\vec{x},\,t) = \frac{1}{M^{(0)}}\int \mathrm{d}^3w\,(w_i - u_i)(w_j - u_j)F(\vec{x},\,\vec{w},\,t)\,. \tag{2.33}$$

In practice, the computation of the hierarchy stops here, because on scales where perturbative treatments of δ hold, the velocity dispersion tensor σ_{ij} is negligible. This happens up to *shell-crossing*, when the trajectories of fluid elements of dark matter intersect. Carrying out the first two moment calculations on the Vlasov equation, one obtains

the continuity and Euler equations:

$$\frac{\partial \delta}{\partial \eta} + \theta = -\vec{u} \cdot \vec{\nabla}_x \delta - \delta\theta, \tag{2.34}$$

$$\frac{\partial \theta}{\partial \eta} + \mathcal{H}\theta + \nabla_x^2 \phi = -\vec{\nabla}_x \cdot \theta\vec{u}, \tag{2.35}$$

where the time dependence is expressed in terms of the conformal time $\mathrm{d}\eta \equiv \mathrm{d}t/a(t)$, the density is expressed in terms of the contrast $\delta(\vec{x}, \eta)$, and the velocity $\vec{u}(\vec{x}, \eta)$ is decomposed in terms of its divergence $\theta \equiv \vec{\nabla} \cdot \vec{u}$ and vorticity $\vec{\omega} \equiv \vec{\nabla} \times \vec{u}$, with the latter assumed to be negligible [83]. $\mathcal{H} \equiv (1/a)\mathrm{d}a/\mathrm{d}\eta = aH$ if the conformal Hubble rate. In order to express the continuity equation in terms of $\delta(\vec{x}, \eta)$, the average density $\bar{\rho}(\eta)$ can be factored out by using the Friedmann equations (2.2) and (2.3), appropriately expressed in conformal time. In particular, the matter term Ω_m can be used in equation (2.29),

$$\nabla_x^2 \phi(\vec{x}, \eta) = \frac{3}{2}\mathcal{H}^2(\eta)\Omega_m(\eta)\delta(\vec{x}, \eta)\,, \tag{2.36}$$

to further simplify Euler equation (2.35).

In order to solve equations (2.34) and (2.35) it is convenient to move to Fourier space:

$$g(\vec{k}) = \mathcal{F}[g(\vec{x})] \equiv \int \mathrm{d}^3x\, g(\vec{x}) e^{-i\vec{k}\cdot\vec{x}};$$

$$g(\vec{x}) = \mathcal{F}^{-1}[g(\vec{k})] \equiv \int \frac{\mathrm{d}^3k}{(2\pi)^3} g(\vec{k}) e^{i\vec{k}\cdot\vec{x}} \equiv \int_{\vec{k}} g(\vec{k}) e^{i\vec{k}\cdot\vec{x}},$$

and introduce the following definition:

$$\mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n} \equiv \int \frac{d^3q_1 \cdots d^3q_n}{(2\pi)^{3n}} (2\pi)^3 \delta_D(\vec{k} - \vec{q}_{1\cdots n})\,, \tag{2.37}$$

where δ_D indicates the Dirac delta function and $\vec{q}_{1\cdots n} \equiv \vec{q}_1 + \cdots + \vec{q}_n$. Equation (2.34) becomes:

$$\frac{\partial \delta}{\partial \eta} + \theta = -\mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2} \alpha(\vec{q}_1, \vec{q}_2)\theta(\vec{q}_1)\delta(\vec{q}_2), \tag{2.38}$$

where $\alpha(\vec{q}_1, \vec{q}_2) = \vec{q}_{12} \cdot \vec{q}_1/q_1^2$. Following the same procedure and using the Poisson and Friedmann equations, the Euler equation (2.35) takes the form

$$\frac{\partial\theta}{\partial\eta} + \mathcal{H}\theta + \frac{3}{2}\mathcal{H}^2\Omega_m\delta = -\mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2}\beta(\vec{q}_1,\,\vec{q}_2)\theta(\vec{q}_1)\theta(\vec{q}_2), \tag{2.39}$$

where $\beta(\vec{q}_1,\,\vec{q}_2) = q_{12}^2(\vec{q}_1\cdot\vec{q}_2)/(2q_1^2q_2^2)$.

A perturbative expansion for $\delta(\vec{k},\,\eta)$ and $\theta(\vec{k},\,\eta)$ can then be assumed,

$$\delta(\vec{k},\,\eta) = \sum_{n=1}^{\infty}\delta^{(n)}(\vec{k},\,\eta), \quad \theta(\vec{k},\,\eta) = \sum_{n=1}^{\infty}\theta^{(n)}(\vec{k},\,\eta), \tag{2.40}$$

where the n-th terms are convolutions involving n linear fields $\delta^{(1)}(\vec{k},\,\eta)$. At linear order, the two equations (2.38) and (2.39) can be combined in a second order differential equation for $\delta^{(1)}(\vec{k},\,\eta)$:

$$\frac{\partial^2\delta^{(1)}}{\partial\eta^2} + \mathcal{H}\frac{\partial\delta^{(1)}}{\partial\eta} + \frac{3}{2}\mathcal{H}^2\Omega_m\delta^{(1)} = 0. \tag{2.41}$$

Notice that each Fourier mode evolves in time independently from the others, and all the coefficients in the equation are scale-independent. For this reason $\delta^{(1)}(\vec{k},\,\eta)$ can be factored into a time-dependent component and a scale-dependent one,

$$\delta^{(1)}(\vec{k},\,\eta) = D(\eta)\varphi_0(\vec{k})\,, \tag{2.42}$$

with $D(\eta)$ the *linear growth factor* for the growing mode, that for the ΛCDM model is given by [84]

$$D(a) = a\;{}_2F_1\left(\frac{1}{3}, 1; \frac{11}{6}; -\frac{\Omega_\Lambda}{\Omega_m}a^3\right), \tag{2.43}$$

where ${}_2F_1$ is the Gauss hypergeometric function. The n-th order solutions can be found by solving equations (2.38) and (2.39) iteratively, starting from the linear ones. They are convolutions of the form

$$\delta^{(n)}(\vec{k},\,\eta) = \mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n}F_n(\vec{q}_1,\,\cdots,\,\vec{q}_n,\,\eta)\varphi_0(\vec{q}_1,\,\eta)\cdots\varphi_0(\vec{q}_n,\,\eta)\,, \tag{2.44}$$

$$\theta^{(n)}(\vec{k},\,\eta) = \mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n}G_n(\vec{q}_1,\,\cdots,\,\vec{q}_n,\,\eta)\varphi_0(\vec{q}_1,\,\eta)\cdots\varphi_0(\vec{q}_n,\,\eta)\,. \tag{2.45}$$

The matter kernels F_n, together with the respective velocity kernels G_n, can be computed iteratively from equations (2.38) and (2.39), starting from the linear solution found

in equation (2.42) and considering that $F_1 = G_1 = 1$. Their explicit expressions are given for example in Bernardeau et al. [85]. For an *Einstein-de Sitter* (EdS) Universe $\Omega_m = 1$ and $\Omega_\Lambda = 0$, so equation (2.43) yields $D(a) = a$, or in terms of conformal time, $D(\eta) = e^\eta$; beyond that, continuity and Euler equations can be parametrized in terms of the linear growth rate $f \equiv \mathrm{d}\log D/\mathrm{d}\log a$, which is equal to 1 for EdS, so no extra time dependence enters the kernels. For ΛCDM both Ω_m and f are time-dependent but vary slowly, and the time dependence introduced in F_n and G_n is mild [85], so the EdS kernels are taken as an efficient approximation; this is no longer the case at low redshifts [86].

2.3.2 Redshift space distortions

Redshift space distortions arise from the fact that we cannot directly measure the distance of cosmological objects such as distant galaxies. Distance measurement can either come from the distance ladder, where the distance of close objects is used to infer the distance of those further away, or by measuring the redshift z of an object and using the Friedmann equation (2.6), which is cosmology-dependent [87, 88].[4] Distances measured from z depend not only on the cosmological redshift supplied by the expansion of the Universe, but also on the kinematic redshift induced by any component of the peculiar velocity along the line of sight $\hat{r}$ from an observer, $u_r \equiv \vec{u}\cdot\hat{r}$. This results in an anisotropic displacement of the inferred positions from the real ones along the radial line of sight. Rescaling the peculiar velocity used previously as $\vec{u}' \equiv \vec{u}/(\mathcal{H}f)$, the transformation between redshift space $\vec{s}$ and real space $\vec{x}$ coordinates is given by

$$\vec{s} = \vec{x} + f u_r' \hat{r}\,. \tag{2.46}$$

When considering the field of density fluctuations observed in redshift space, $\delta_S(\vec{s})$, number conservation implies that

$$(1 + \delta_S(\vec{s}))\,\mathrm{d}^3 s = (1 + \delta(\vec{x}))\,\mathrm{d}^3 x\,, \tag{2.47}$$

leading to the equation:

[4] Standard sirens, like binary black hole or binary neutron star mergers, provide an independent and direct measurement of the distance of the merger from the observer, potentially overcoming the problem of redshift space distortions. There are difficulties involved in the distance determination however, for example the inclination angle of the orbital plane of the binary system is degenerate with the distance, leading to large uncertainties [89].

$$\delta_S(\vec{s}) = \frac{1+\delta(\vec{x})}{\det J_s(\vec{x})} - 1\,. \tag{2.48}$$

The determinant of the Jacobian of the transformation (2.46), $J_{s,ij} = \partial s_i/\partial x_j = \delta^K_{ij} + f(\partial u'_r/\partial x_j)\hat{r}_i$, can be found by noting that given a tensor product $\vec{a}\otimes\vec{b}$, a matrix of the form $\mathbb{I}+\vec{a}\otimes\vec{b}$, as is the case for $J_s(\vec{x})$, has determinant $\det(\mathbb{I}+\vec{a}\otimes\vec{b}) = 1 + b\cdot a$,

$$\det(J_s) = \det(\mathbb{I} + \hat{r}\otimes f\vec{\nabla}_x u'_r) = 1 + f\vec{\nabla}_x u'_r\cdot\hat{r} = 1 + f\partial_r u'_r\,, \tag{2.49}$$

where ∂_r denotes the directional derivative along the line of sight. Continuing from equation (2.48):

$$\delta_S(\vec{s}) = \frac{\delta - f\partial_r u'_r}{1 + f\partial_r u'_r}\,, \tag{2.50}$$

where the dependence on $\vec{x}$ of the right-hand side is implied. This result is exact as long as the line of sight direction $\hat{r}$ does not change with the position $\vec{x}$, this is the so-called *distant observer* approximation; the distant observer approximation can be discarded to include wide-angle corrections, see, e.g., Raccanelli et al. [90] for a treatment. From linear perturbation theory $\delta(\vec{x},\,\eta) = \delta^{(1)}(\vec{x},\,\eta) = D(\eta)\varphi_0(\vec{x})$, so the linearized continuity equation, equation (2.38) with null right-hand side, gives $\theta'^{(1)} = \vec{\nabla}_x\cdot\vec{u}'^{(1)} = -\delta^{(1)}$, or equivalently in Fourier space,

$$i\vec{k}\cdot\vec{u}'^{(1)}(\vec{k}) = -\delta^{(1)}(\vec{k})\,, \tag{2.51}$$

At linear order the peculiar velocity field has vanishing vorticity, so the velocity is entirely sourced by the gradient of its potential $\vec{u}'(\vec{k}) = i\vec{k}\psi(\vec{k})$, and so $\vec{u}'\,||\,\vec{k}$. This means that equation (2.51) can be manipulated to find the linear velocity field,

$$\begin{aligned}
\vec{k}\cdot u'^{(1)}(\vec{k})\hat{k} &= i\delta^{(1)}(\vec{k}) \\
k\,u'^{(1)}(\vec{k}) &= i\delta^{(1)}(\vec{k}) \\
u'^{(1)}(\vec{k})\hat{k} &= i\frac{\hat{k}}{k}\delta^{(1)}(\vec{k}) \\
\vec{u}'^{(1)}(\vec{k}) &= i\frac{\vec{k}}{k^2}\delta^{(1)}(\vec{k})\,.
\end{aligned} \tag{2.52}$$

Finally, projecting along $\hat{r}$ and taking the directional derivative in Fourier space along

the line of sight, $\partial_r \to i\vec{k}\cdot\hat{r} = ik\mu$, with $\mu \equiv \hat{k}\cdot\hat{r}$ the cosine of the angle between the wavevector and the line of sight, yields,

$$ik\mu\, u_r'(\vec{k}) = -\mu^2\delta(\vec{k}), \tag{2.53}$$

where the linear order notation has been omitted. This expression can be plugged into equation (2.50) expanded at linear order around $\partial_r u_r' = 0$ (note that in this approximation, equation (2.46) implies $\vec{s} \approx \vec{x}$),

$$\begin{aligned}\delta_S(\vec{k}) &\approx \delta(\vec{k}) - if k\mu\, u_r'(\vec{k}) \\ &= (1 + f\mu^2)\delta(\vec{k})\,.\end{aligned} \tag{2.54}$$

Equation (2.54) is the *Kaiser effect*, and characterizes the impact of redshift space distortions on large scales [91].

What happens if one drops all assumptions and tries to derive the exact relation between δ_S and δ? Starting from equation (2.47) and taking the Fourier transform with respect to $\vec{s}$ leads to the expression

$$(2\pi)^3\delta_D(\vec{k}) + \delta_S(\vec{k}) = \int \mathrm{d}^3x\, e^{-i\vec{k}\cdot\vec{x}} e^{-if k\mu u_r'}\,(1 + \delta(\vec{x}))\ . \tag{2.55}$$

This relation is exact and valid even at scales where multi-streaming applies. In the single-stream regime, where a single position $\vec{x}$ maps to a single $\vec{s}$, the relation in equation (2.50) applies, and its Fourier transform can be derived with the Jacobian,

$$\delta_S(\vec{k}) = \int \mathrm{d}^3x\, e^{-i\vec{k}\cdot\vec{x}} e^{-if k\mu u_r'}\,(\delta(\vec{x}) - f\partial_r u_r')\ . \tag{2.56}$$

Note how the exponential $e^{-if k\mu u_r'}$ introduces a non-linear term of the velocity; expanding the term at first order in the argument, one finds again equation (2.54), and more generally, expanding at order n and using equations (2.44) and (2.45) gives the redshift space kernel Z_n [85, 92]:

$$\delta_S(\vec{k}) = \sum_{n=1}^{\infty} \frac{k_r^{n-1}}{(n-1)!} \mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n} \left(\delta(\vec{q}_1) - \frac{q_{r,1}^2}{q_1^2}\theta'(\vec{q}_1)\right) \prod_{i=2}^{n-1} \frac{q_{r,i}}{q_i^2}\theta'(\vec{q}_i) \tag{2.57}$$
$$= \sum_{n=1}^{\infty} \mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n} Z_n(\vec{q}_1, \cdots, \vec{q}_n)\delta^{(1)}(\vec{q}_1)\cdots\delta^{(1)}(\vec{q}_n)\,,$$

where $\{k_r,\, q_r\} \equiv \{k\mu,\, q\mu\}$. Going back to the exponential in equation (2.55), one can see the velocity term $u'_r(\vec{x})$ as made up of a large-scale field $u'^L_r(\vec{x})$ describing bulk flows (which is well described in first approximation by the Kaiser effect), and a small-scale component $u'^S_r(\vec{x})$ encoding random motion inside virialized structures such as galaxy clusters. Since the latter can produce large radial velocities, the displacement in the measured redshift space positions can affect the value of the density at large scales. In particular, this can be seen as a damping factor in front of $\delta_S(\vec{k})$ given by averaging the exponential over the small-scale velocity field, $D_{\rm FoG}(k,\,\mu) \equiv \left\langle e^{ik\mu u'^S_r} \right\rangle_x$; FoG stands for "*Fingers of God*", a name given to the peculiar anisotropic pattern visible in galaxy surveys, where galaxies in virialized clusters appear smeared along the line of sight. The fingers of God are modeled by treating $u'^S_r(\vec{x})$ as a random variable following some distribution, then $D_{\rm FoG}(k,\,\mu)$ ends up being the characteristic function of the distribution. Accounting for this small-scale contribution, the redshift space field is given by

$$\delta_S(\vec{k}) \approx D_{\rm FoG}(k,\,\mu)\,(1 + f\mu^2)\delta(\vec{k})\,. \tag{2.58}$$

2.3.3 The large-scale structure bootstrap

The results presented in section 2.3.1 have been derived starting from the Vlasov equation, leading to the equations of motion for dark matter as a pressureless, perfect fluid in the form of the continuity and Euler equations. D'Amico et al. [51] introduces an alternative approach, where the analytical form of the kernels is constrained by symmetry arguments instead of deriving them as a perturbative solution of the equations of motion in a given cosmological model. The exact form is derived by enforcing symmetries, such as translational and rotational invariance, and "Bose symmetry", which allows the kernels to be symmetrized in the external momenta. A special role is played by *extended Galilean invariance* (EGI), or invariance under time-dependent translations, as discussed in Peloso and Pietroni [93] and Kehagias and Riotto [94]. Following a translation $\vec{x} \to \vec{x} + \vec{d}(\eta)$, EGI induces a nonlinear transformation on the field,

$$\delta(\vec{k},\,\eta) \to e^{i\vec{k}\cdot\vec{d}(\eta)}\delta(\vec{k},\,\eta)\,. \tag{2.59}$$

Expanding the exponential and treating $\vec{d}(\eta)$ as the $\vec{k} = 0$ limit of a linear field $\vec{d}^{(1)}$, then one can write

$$\delta^{(1)}(\vec{k},\,\eta) \to \delta^{(1)}(\vec{k},\,\eta) + (2\pi)^3\delta_D(\vec{k},\,\eta)i\vec{k}\cdot\vec{d}^{(1)}(\eta) \tag{2.60}$$

$$\delta^{(n)}(\vec{k},\eta) \to \sum_{m=0}^{\infty}\frac{\left(i\vec{k}\cdot\vec{d}^{(1)}(\eta)\right)^m}{m!}\delta^{(n-m)}(\vec{k},\eta)\,, \tag{2.61}$$

where δ was expanded and matched order by order. The transformation in equations (2.60) and (2.61) can then be plugged into equation (2.44) to relate the kernels F_n to F_{n-m}. This turns out to induce a soft limit for the n-th kernel:

$$\lim_{\vec{q}_1,\cdots,\vec{q}_m\to 0} F_n(\vec{q}_1,\,\cdots,\,\vec{q}_n) = \prod_{i=1}^{m}\frac{\vec{q}_i\cdot\vec{q}_{m+1\cdots n}}{q_i^2}F_{n-m}(\vec{q}_{m+1},\,\cdots,\,\vec{q}_n)\,. \tag{2.62}$$

EGI can be used to derive further constraints by considering higher orders in $\vec{d}^{(n)}(\eta)$ as the bulk displacement induced by the peculiar velocity field $u^{(n)}$ integrated in time up to η, which induce similar constraints that relate F_{n+m} with G_n and F_m. Finally, imposing mass and momentum conservation, as is appropriate for dark matter density and velocity fields, provides constraints for the kernels in the limit of the vanishing sum of the external momenta $\vec{q}_{1\cdots n} \to 0$. With all these constraints derived from symmetries, the form of the kernels is set by taking the set of rotationally invariant objects, constructing a basis of dimensionless functions with them and finally applying the constraints to set the general coefficients of their linear combination. For $n = 2$, the set of possible rotationally invariant objects that can be constructed with $\vec{q}_1$ and $\vec{q}_2$ is $\{\vec{q}_1\cdot\vec{q}_2,\, q_1^2,\, q_2^2\}$. Then, the most generic dimensionless function is of the form

$$\left(q_1^2\right)^a\left(q_2^2\right)^b(\vec{q}_1\cdot\vec{q}_2)^c \quad \text{with} \quad a+b+c=0, \tag{2.63}$$

with a, b, $c \in \mathbb{Z}$. EGI and Bose symmetry can be used to derive the constraints $2a + c \geq -1$ and $2b + c \geq -1$ from the soft limit behavior of F_2 (G_2),[5] and the simplest possible functions that satisfy these requirements are:

[5]Note that $F_1 = G_1 = 1$.

$$B_1(\vec{q}_1,\,\vec{q}_2)\equiv 1,\quad B_2(\vec{q}_1,\,\vec{q}_2)\equiv\frac{\vec{q}_1\cdot\vec{q}_2}{q_1^2},\tag{2.64}$$
$$B_3(\vec{q}_1,\,\vec{q}_2)\equiv\frac{\vec{q}_1\cdot\vec{q}_2}{q_2^2}\quad\text{and}\quad B_4(\vec{q}_1,\,\vec{q}_2)\equiv\frac{(\vec{q}_1\cdot\vec{q}_2)^2}{q_1^2q_2^2}\,,$$

however, it is convenient to work with a different basis:

$$B_1(\vec{q}_1,\,\vec{q}_2),\quad \alpha(\vec{q}_1,\,\vec{q}_2)\equiv B_2-B_3,\quad \beta(\vec{q}_1,\,\vec{q}_2)\equiv\frac{B_2+B_3}{2}+B_4\quad\text{and}\quad \gamma(\vec{q}_1,\,\vec{q}_2)\equiv B_1-B_4\,.\tag{2.65}$$

The most generic kernel is a time-dependent linear combination of the four, however since $\alpha(\vec{q}_1,\,\vec{q}_2)$ is antisymmetric, it cannot appear in the second order kernels as it would violate Bose symmetry, so in general

$$F_2(\vec{q}_1,\,\vec{q}_2,\,\eta)=a_0^{(2)}(\eta)+a_1^{(2)}(\eta)\,\beta(\vec{q}_1,\,\vec{q}_2)+a_2^{(2)}(\eta)\,\gamma(\vec{q}_1,\,\vec{q}_2)\,.\tag{2.66}$$

Equation (2.62) sets the soft limit for the kernel,[6]

$$\lim_{\vec{q}_1\to 0}F_2(\vec{q}_1,\,\vec{q}_2)=a_1^{(2)}\frac{\vec{q}_1\cdot\vec{q}_2}{2q_1^2}\overset{\text{EGI}}{=}\frac{\vec{q}_1\cdot\vec{q}_2}{q_1^2}\quad\longrightarrow\quad a_1^{(2)}=2\,,\tag{2.67}$$

while mass conservation dictates

$$\lim_{\vec{q}_{12}\to 0}F_2(\vec{q}_1,\,\vec{q}_2)=F_2(\vec{q}_1,\,-\vec{q}_1)=a_0^{(2)}\overset{\text{m.c.}}{=}0\,.\tag{2.68}$$

The same arguments apply for G_2, leading to the general second order kernels:

$$F_2(\vec{q}_1,\,\vec{q}_2,\,\eta)=\beta(\vec{q}_1,\,\vec{q}_2)+\frac{a_\gamma^{(2)}(\eta)}{2}\,\gamma(\vec{q}_1,\,\vec{q}_2)\,,\tag{2.69}$$
$$G_2(\vec{q}_1,\,\vec{q}_2,\,\eta)=\beta(\vec{q}_1,\,\vec{q}_2)+\frac{d_\gamma^{(2)}(\eta)}{2}\,\gamma(\vec{q}_1,\,\vec{q}_2)\,,\tag{2.70}$$

where a factor 2! at the denominator coming from the second order term of a functional

[6]Note that $a_0^{(2)}(\eta)$ cannot appear in the soft limit, as it would generate an unphysical second order contribution when one of the momenta vanishes.

Taylor expansion was absorbed inside the kernels.

The same procedure can be extended to biased tracers, where mass and momentum conservation do not apply, resulting in a higher number of independent terms. In particular, the first order kernel becomes $K_1(\vec{q}_1, \eta) = c_0^{(1)}(\eta)$, and EGI gives $c_1^{(2)} = 2c_0^{(1)}$, yielding the kernel:

$$2!\, K_2(\vec{q}_1, \vec{q}_2, \eta) = c_0^{(2)} + 2c_0^{(1)}\beta(\vec{q}_1, \vec{q}_2) + c_\gamma^{(2)}(\eta)\,\gamma(\vec{q}_1, \vec{q}_2)\,. \tag{2.71}$$

The new tracer dependent coefficients can be related to the bias parameters derived from other bases (see Desjacques, Jeong, and Schmidt [95] for a review on galaxy bias).

These results show that the bootstrap of LSS can be used to constrain the non-linear evolution of δ by assuming the symmetries listed above but no specific theoretical model. Such a tool can be applied to constrain deviations from known cosmologies and thus detect new physics [96–98].

2.3.4 Power spectrum and bispectrum

Section 2.2.1 established that basic models of inflation enforce approximately Gaussian statistics for the density field at scales where gravity has little effect. This means that most of the cosmological information that can be obtained from the large scale distribution of matter is contained the 2-point correlation function of $\delta(\vec{x}, \eta)$:

$$\xi(\vec{x}_1, \vec{x}_2) = \langle \delta(\vec{x}_1)\delta(\vec{x}_2)\rangle\,. \tag{2.72}$$

On the assumption that the cosmological principle holds, this relation must be invariant under global translations (homogeneity) and rotations (isotropy). This means that the 2-point correlation function only depends on the distance between $\vec{x}_1$ and $\vec{x}_2$, regardless of their respective orientation,

$$\xi(\vec{x}_1, \vec{x}_2) = \langle \delta(\vec{x}_1)\delta(\vec{x}_2)\rangle = \xi(|\vec{x}_1 - \vec{x}_2|) \equiv \xi(r)\,. \tag{2.73}$$

The Fourier transform of the 2-point correlation function is the power spectrum $P(k)$:

$$\langle \delta(\vec{k})\overline{\delta}(\vec{k}')\rangle = \langle \delta(\vec{k})\delta(-\vec{k}')\rangle = (2\pi)^3\delta_D(\vec{k} + \vec{k}')P(\vec{k})\,, \tag{2.74}$$

where $\delta_D(\vec{k}+\vec{k}')$ arises from homogeneity. The standard perturbation theory defined in section 2.3.1 can be used inside the ensemble average to find a perturbative expansion for the power spectrum:

$$\begin{aligned}\langle\delta(\vec{k})\delta(-\vec{k}')\rangle &= \left\langle \sum_{n=1}^{\infty}\delta^{(n)}(\vec{k})\sum_{m=1}^{\infty}\delta^{(m)}(-\vec{k}')\right\rangle \\ &= \langle\delta^{(1)}(\vec{k})\delta^{(1)}(-\vec{k}')\rangle + 2\langle\delta^{(1)}(\vec{k})\delta^{(2)}(-\vec{k}')\rangle + \\ &\quad + 2\langle\delta^{(1)}(\vec{k})\delta^{(3)}(-\vec{k}')\rangle + \langle\delta^{(2)}(\vec{k})\delta^{(2)}(-\vec{k}')\rangle + \cdots ,\end{aligned} \tag{2.75}$$

where the expansion has been carried out explicitly up to second order in the fields. Wick's theorem implies that averages including an odd number of linear fields vanish in the limit in which these are exactly gaussian. Therefore, at leading order (LO), the linear matter power spectrum is

$$P_{\rm LO}(k) \equiv P_{11}(k) \equiv \langle\delta^{(1)}(\vec{k})\delta^{(1)}(-\vec{k}')\rangle' , \tag{2.76}$$

while the next-to-leading order (NLO) correction is given by:

$$P_{\rm NLO}(k) \equiv P_{13}(k) + P_{22}(k) \equiv 2\langle\delta^{(1)}(\vec{k})\delta^{(3)}(-\vec{k}')\rangle' + \langle\delta^{(2)}(\vec{k})\delta^{(2)}(-\vec{k}')\rangle' , \tag{2.77}$$

where primes indicate that the $(2\pi)^3\delta_D(\vec{k}+\vec{k}')$ term in equation (2.74) has been factored out.

This description is based on the assumption that the single-stream perfect fluid approximation introduced in section 2.3.1 can be extended at all scales, however it is found to diverge at $k \gtrsim 0.2h\,\mathrm{Mpc}^{-1}$. This invites an effective description of the standard perturbation theory, where the small scale dynamics that would otherwise be added to the theory explicitly are instead accounted for by introducing *counterterms*. A way to approach this problem is detailed in Baumann et al. [99], Pietroni et al. [100], and Carrasco, Hertzberg, and Senatore [101], here instead follows a simple illustration on how the leading order counterterm can be obtained from the power spectrum. Consider $P_{13}(k)$, following equation (2.44) and applying Wick's theorem, it is given by the integral

$$P_{13}(k) = 6P_{11}(k)\int\frac{\mathrm{d}^3q}{(2\pi)^3}\,F_3(\vec{k},\,\vec{q},\,-\vec{q})P_{11}(q)\,. \tag{2.78}$$

To regulate small-scale behavior, a cutoff scale Λ can be imposed on the above integral, and the behavior of the F_3 kernel can be studied in the limit where $q \geq \Lambda \gg k$,

$$P_{13}^{\rm UV}(k;\,\Lambda) \propto P_{11}(k)\,k^2 \int^{\Lambda} \frac{{\rm d}^3 q}{(2\pi)^3}\,\frac{P_{11}(q)}{q^2}\,, \tag{2.79}$$

this gives the leading order behavior of the small-scale "UV" contribution to the large scales as a consequence of momentum conservation, up to a constant term. The integral above can be interpreted as a dissipative term expressed by an effective sound speed $c_s^2(\Lambda)$, and a matching counterterm $-c_s^2(\Lambda)k^2P_{11}(k)$ can then be added to $P_{\rm LO}(k) + P_{\rm NLO}(k)$ to correct for it. Since this must hold at any Λ, in practice the *bare* $c_s^2(\Lambda)$ is expressed as a *renormalized* constant term at some fixed scale $\Lambda = \Lambda_{\rm uv}$ plus a Λ-dependent term that matches the UV contributions, $c_s^2(\Lambda) = c_s^2(\Lambda_{\rm uv}) + \delta c_s^2(\Lambda;\,\Lambda_{\rm uv})$. The constant is then fitted to simulations, while the Λ-dependent term always cancels out the undesired contributions to the perturbation theory, computed now only with momenta $q < \Lambda_{\rm uv}$. Other counterterms can be derived similarly. One thing to note is that once $\Lambda_{\rm uv}$ is set, the relative gap in counterterms defined at some new $\Lambda' < \Lambda_{\rm uv}$ is computable in perturbation theory in the limit $k \ll \Lambda'$; the linear fields contributing to the "IR" scales $q < \Lambda_{\rm uv}$, $\delta^{(1)}_{\Lambda_{\rm uv}}$, can be seen as processed by a low-pass filter that removes all the modes beyond $\Lambda_{\rm uv}$, and so can be decomposed as a sum of a similar field filtered up to Λ' and the remainder between the two UV cutoffs, $\delta^{(1)}_{\Lambda_{\rm uv}} = \delta^{(1)}_{\Lambda'} + \Delta\delta^{(1)}_{\Lambda',\Lambda_{\rm uv}}$. With this in mind, the contributions given by mixing $\Delta\delta^{(1)}_{\Lambda',\Lambda_{\rm uv}}$ and $\delta^{(1)}_{\Lambda_{\rm uv}}$ can be computed perturbatively, and depending on their scale dependence they can be matched to the counterterm operators with the same behavior in the theory at Λ'. This will prove to be key in understanding the work presented in chapter 5.

To study possible non-Gaussian information present at large scales, a straightforward approach is to use the three-point function, also called the *bispectrum*:

$$\left\langle \delta(\vec{k}_1)\delta(\vec{k}_2)\delta(\vec{k}_3)\right\rangle = (2\pi)^3\delta_D(\vec{k}_1+\vec{k}_2+\vec{k}_3)B(\vec{k}_1,\,\vec{k}_2,\,\vec{k}_3)\,. \tag{2.80}$$

The possibility of non-Gaussianity injected into the initial curvature perturbations $\zeta(\vec{k})$ was hinted the end of section 2.2.1. The simplest way to introduce non-Gaussian features is to perform some non-linear transformation $f\left(\zeta_G(\vec{x})\right)$ to a Gaussian field $\zeta_G(\vec{x})$. For small curvature perturbations, an expansion of f would converge quickly, so the quadratic contribution is sufficient,

$$\zeta(\vec{x}) = \zeta_G(\vec{x}) + f_{\rm NL}\left(\zeta_G^2(\vec{x}) - \left\langle\zeta_G^2(\vec{x})\right\rangle\right)\,. \tag{2.81}$$

The bispectrum of the curvature perturbations would then be given by the four-point function (or *trispectrum*), which can be split in products of the power spectrum via the

Wick theorem:

$$B_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3) = 2f_{\rm NL}\left[P_\zeta(k_1)P_\zeta(k_2) + P_\zeta(k_1)P_\zeta(k_3) + P_\zeta(k_2)P_\zeta(k_3)\right] . \tag{2.82}$$

From equation (2.23), this is related to the matter bispectrum by:

$$B(\vec{k}_1, \vec{k}_2, \vec{k}_3) = \mathcal{M}(\vec{k}_1)\mathcal{M}(\vec{k}_2)\mathcal{M}(\vec{k}_3)B_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3) . \tag{2.83}$$

Equation (2.81) covers primordial non-Gaussianity of the local type, or more specifically *quadratic local primordial non-Gaussianity*. The expansion can also be carried out to higher order, introducing higher order terms like $g_{\rm NL}\zeta^3$. Other configurations of primordial non-Gaussianity are possible [102].

Chapter 3

Statistical Inference

The notion of density perturbations, introduced in chapter 2, is central to the study of the large-scale structure (LSS) of the Universe. Their evolution through gravitational collapse and small scale physics defines the theoretical models used to describe the structures observed today. These models allow for degrees of freedom that need to be fit to the observables, however their value must account for uncertainty and the fundamentally random nature of the initial fluctuations that evolved to present time. The problem of inferring parameters of a random process from observations is addressed by the field of statistics. Statistical inference is at the basis of all data analysis techniques applied to physics. This is the starting point of this chapter; section 3.1 dives into the specifics of statistical inference applied to parametric models, enunciating important results that will justify the methodologies followed in later chapters. In section 3.2 the focus will shift on one of the most popular and powerful methodologies for sampling randomly distributed quantities, presenting different methods and providing some practical guidance. For the concepts touched in this chapter and a more general outlook on statistics, the interested reader can take a look at Wasserman [103] and Gelman et al. [104] and the review of Hamiltonian Monte Carlo methods by Betancourt [105].

3.1 Parametric inference

In statistics, inference is the process of determining the distribution that generated some measured data from the data themselves. It is based on the assumption that the observed sample is drawn from some random process, and so one can estimate its statistical properties by working backwards using the tools of probability theory. Inference need not be the exact determination of the entire distribution, instead it could be limited

to determining some feature of it like the mean and the variance. Particular attention should be given to distinguish between frequentist and Bayesian approaches. While this dissertation focuses on the Bayesian formulation of statistics, it is good practice to give a brief comparison of inferential fundamentals when seen from a frequentist perspective as well. What follows briefly presents the specific problems of point estimation 3.1.1 and quantification of uncertainty 3.1.2 from both sides, since the research presented in this thesis is exclusively concerned with them.

3.1.1 Point estimation

Point estimation is the class of problems that revolve around giving a single best estimate for a quantity of interest. For example, given a sample $\{x\}_{i=1}^{N}$ of observed random variables, all drawn from the same parametric *probability density function* (PDF), $x_i \sim f(x_i;\, \theta)$, an estimator for parameter θ can then be constructed from the data, $\hat{\theta}(x_i)$. If the expected value is equal to the true value, $\mathbb{E}[\hat{\theta}] = \theta$, then the estimator is said to be unbiased. Note that since $\hat{\theta}$ is a function of the sample, estimators are random variables themselves. In the frequentist setting, the distribution over which the estimator is sampled is called the *sampling distribution*. As $N \to \infty$, the sampling distribution of an unbiased estimator converges to the real value θ, which is fixed. In contrast, the Bayesian approach is to generalize θ as a random variable, to which is assigned a *prior* distribution $\pi(\theta)$ that encodes knowledge about θ before observing the sample. Once the sample is observed, a *posterior* probability distribution $P(\theta \,|\, \{x\}_{i=1}^{N})$ can be constructed from Bayes theorem by picking an appropriate *likelihood* $\mathcal{L}(\{x\}_{i=1}^{N} \,|\, \theta)$, on which all kinds of estimators can be constructed, for example the mean $\hat{\theta} = \int \mathrm{d}\theta\; \theta P(\theta \,|\, \{x\}_{i=1}^{N})$. One of the most important results when it comes to statistical estimators is the *maximum likelihood estimator* (MLE): if the sample is *independent and identically distributed* (i.i.d.) from a distribution $f(x;\, \theta)$, then the likelihood of the sample is given by

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} f(x_i;\, \theta)\,. \tag{3.1}$$

Note that while this is proportional to the joint probability distribution of $\{x\}_{i=1}^{N}$, it is not a probability density function, since its integral is not guaranteed to be normalized to 1. The MLE is then the maximum of the likelihood:

$$\left.\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta}\right|_{\hat{\theta}_{\mathrm{MLE}}} = 0\,, \tag{3.2}$$

although often it is more convenient to maximize the logarithm of $\mathcal{L}$. The MLE is very appealing because it benefits from many useful properties:

- it is guaranteed to converge to the real value θ;
- given a parameter $\gamma = g(\theta)$, then $\hat{\gamma}_{\mathrm{MLE}} = g(\hat{\theta}_{\mathrm{MLE}})$;
- it is asymptotically normal: as $N \to \infty$, $\hat{\theta}_{\mathrm{MLE}} \sim \mathcal{N}\left(\theta, \sqrt{F_N^{-1}(\theta)}\right)$, where $\mathcal{N}(\mu, \sigma)$ denotes a normal distribution of mean μ and variance σ^2, and $F_N(\theta)$ is the *Fisher information*, which will be introduced in section 3.1.2;
- among all the well-behaved estimators, the MLE is optimal, or in other words it has the lowest variance for large enough N.

Despite this, it must be stressed that the MLE is only good in the limit of sufficiently large N, with a non-pathological likelihood and a well-specified probabilistic model $f(x;\, \theta)$ for the data-generating process.

In the Bayesian framework, the MLE estimator can be recovered with an appropriate choice of prior on θ. In particular, when the prior is flat[1] over the interval where the high probability region of the likelihood resides, then the maximum $\hat{\theta}_{\mathrm{MAP}}$ of the posterior distribution of θ,

$$\begin{aligned} 0 &= \left.\frac{\mathrm{d}P(\theta \,|\, \{x\}_{i=1}^N)}{\mathrm{d}\theta}\right|_{\hat{\theta}_{\mathrm{MAP}}} \\ &= \left.\frac{\mathrm{d}\mathcal{L}(\{x\}_{i=1}^N \,|\, \theta)}{\mathrm{d}\theta}\pi(\theta)\right|_{\hat{\theta}_{\mathrm{MAP}}} + \mathcal{L}(\{x\}_{i=1}^N \,|\, \theta)\left.\frac{\mathrm{d}\pi(\theta)}{\mathrm{d}\theta}\right|_{\hat{\theta}_{\mathrm{MAP}}} \\ &= \left.\frac{\mathrm{d}\mathcal{L}(\{x\}_{i=1}^N \,|\, \theta)}{\mathrm{d}\theta}\right|_{\hat{\theta}_{\mathrm{MAP}}}, \end{aligned} \tag{3.3}$$

will coincide with $\hat{\theta}_{\mathrm{MLE}}$, since the derivative of the prior vanishes. This is also the case in the limit of $N \to \infty$ with a more informative prior, since the likelihood dominates over it. This is formalized in an important result, the Bernstein Von-Mises theorem: as long as the prior is smooth and does not vanish near the maximum of the likelihood, the posterior distribution will always converge to the sampling distribution of $\hat{\theta}_{\mathrm{MLE}}$ as $N \to \infty$. This means that Bayesian and frequentist statistics are effectively equivalent in the limit of large samples. But the strength of the Bayesian approach lies in its flexibility: while the MLE fails when certain regularity conditions are not satisfied, a posterior

[1] Such a prior is often said to be "uninformative".

distribution can always be constructed to be optimal within the given constraints. To illustrate this point, consider the case where a small sample $\{x\}_{i=1}^N$, with N small, has been collected, and there is a good guess for the parametric distribution $f(x;\theta)$ that generated the data. If the goal is to compute the probability that $x_{N+1} > a \in \mathbb{R}$, then the first step is to estimate θ. Naively, one can compute $\hat{\theta}_{\text{MLE}}$ from the small sample, and obtain $P(x_{N+1} > a \,|\, \{x\}_{i=1}^N;\, \theta = \hat{\theta}_{\text{MLE}})$ from the cumulative distribution function (CDF) of $f(x;\, \theta = \hat{\theta}_{\text{MLE}})$. While this works for large N, for small N the value of $\hat{\theta}_{\text{MLE}}$ is strongly uncertain, which leads to a suboptimal conclusion; nothing forbids that the data has been sampled by other values of θ, infinitely many so, so how can the problem be solved optimally? An intuitive way is to perform a "weighted average" of $f(x_{N+1};\, \theta)$ as θ changes. From the point of view of Bayesian statistics, the Bayes theorem provides an immediate solution:

$$\begin{aligned} P(x_{N+1} \,|\, \{x\}_{i=1}^N) &= \int \mathrm{d}\theta \, f(x_{N+1} \,|\, \theta) P(\theta \,|\, \{x\}_{i=1}^N) \\ &= \int \mathrm{d}\theta \, f(x_{N+1} \,|\, \theta) \frac{\mathcal{L}(\{x\}_{i=1}^N \,|\, \theta)\pi(\theta)}{\int \mathrm{d}\theta \, \mathcal{L}(\{x\}_{i=1}^N \,|\, \theta)\pi(\theta)} \,. \end{aligned} \tag{3.4}$$

By construction, the Bayesian framework naturally accounts for the uncertainty in the estimation of θ. While this integral is hard to solve in most cases, it is possible to obtain a closed form expression with an appropriate choice of prior, namely a *conjugate prior*. Given an expression for the likelihood, a conjugate prior π yields a posterior of the same family of distributions as π. Aside from algebraic convenience, conjugate priors lead to posteriors with useful properties in inference problems. In particular conjugate priors can be seen as contributing a number ν of *pseudo-samples*, which stabilize the likelihood in the low N regime. The posterior mean of θ obtained from a conjugate posterior distribution is

$$\mathbb{E}[\theta \,|\, \{x\}_{i=1}^N] = \frac{N}{N+\nu}\hat{\theta}_{\text{MLE}} + \frac{\nu}{N+\nu}\mu_\pi, \tag{3.5}$$

where μ_π is the prior mean.[2] This introduces a bias-variance trade-off, where the posterior mean is more concentrated around μ_π when N is small, while converging to the expected MLE when N is large. A simple example can show this more clearly.

[2]Here was used an abuse of notation. It is correct to use the expected value $\mathbb{E}$ to indicate the mean, however the expected value is not a random variable, while (3.5) is (indeed, the right-hand side contains the MLE, a random variable). Whenever the posterior mean is used as an estimator in the remainder of the text, the expected value will be dropped in favor of the " $\hat{\cdot}$ " notation.

Consider a sample drawn from a normal distribution with zero mean and unknown variance σ^2, $\{x\}_{i=1}^N \sim \mathcal{N}(0,\, \sigma)$. The log-likelihood for the sample is simply:

$$\log \mathcal{L}(\sigma) = C - N \log \sigma - \frac{1}{2} \sum_{i=1}^{N} \frac{x_i^2}{\sigma^2}\,, \tag{3.6}$$

with C the normalization constant. The MLE is:

$$\begin{aligned} 0 &= \left.\frac{\partial \log \mathcal{L}}{\partial \sigma}\right|_{\hat{\sigma}_{\mathrm{MLE}}} \\ &= -N \frac{1}{\hat{\sigma}_{\mathrm{MLE}}} + \sum_{i=1}^{N} \frac{x_i^2}{\hat{\sigma}_{\mathrm{MLE}}^3} \\ &= \frac{1}{\hat{\sigma}_{\mathrm{MLE}}} \left(-N + \sum_{i=1}^{N} \frac{x_i^2}{\hat{\sigma}_{\mathrm{MLE}}^2} \right)\,. \end{aligned} \tag{3.7}$$

The derivative vanishes only when the factor within parentheses does, therefore:

$$\hat{\sigma}_{\mathrm{MLE}}^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2\,; \tag{3.8}$$

$\hat{\sigma}_{\mathrm{MLE}}^2$ is a random variable given by a sum of squares of normally distributed random variables, so it is distributed according to a chi-square distribution with N degrees of freedom. Mean and variance of the MLE can be found using the fact that the expected values of a chi-squared variable χ given by the sum of squares of N variables $\tilde{x}_i \sim \mathcal{N}(0,\, 1)$ are $\mathbb{E}[\chi] = N$ and $\mathrm{Var}[\chi] = 2N$:

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_{\mathrm{MLE}}^2] &= \frac{1}{N} \mathbb{E}\left[\sum_{i=1}^{N} x_i^2 \right] \\ &= \frac{1}{N} \mathbb{E}\left[\sigma^2 \sum_{i=1}^{N} \tilde{x}_i^2 \right] \\ &= \frac{\sigma^2}{N} \mathbb{E}[\chi] \\ &= \sigma^2\,, \end{aligned} \tag{3.9}$$

$$\begin{aligned}
\mathrm{Var}[\hat{\sigma}^2_{\mathrm{MLE}}] &= \mathbb{E}[\hat{\sigma}^4_{\mathrm{MLE}}] - \left(\mathbb{E}[\hat{\sigma}^2_{\mathrm{MLE}}]\right)^2 && (3.10)\\
&= \frac{\sigma^4}{N^2}\mathbb{E}\left[\chi^2\right] - \sigma^4 && (3.11)\\
&= \frac{\sigma^4}{N^2}(\mathrm{Var}\left[\chi\right] + N^2) - \sigma^4 && (3.12)\\
&= \frac{2\sigma^4}{N}\,. && (3.13)
\end{aligned}$$

This is the typical frequentist approach: $\hat{\sigma}^2_{\mathrm{MLE}}$ is unbiased, but crucially its variance is inversely proportional to N. The Bayesian approach introduces the freedom of choosing a prior for σ. With a Gaussian likelihood, a good choice for the conjugate prior of a scaling parameter like σ^2 is the inverse gamma distribution:

$$\pi(\sigma^2;\,\alpha,\,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\sigma^{-2(1+\alpha)}\exp\left[-\frac{\beta}{\sigma^2}\right]\,. \tag{3.14}$$

The resulting posterior is simply:

$$p(\sigma^2 \mid \{x\}_{i=1}^N) \propto (\sigma^2)^{-(1+\alpha+N/2)}\exp\left[-\frac{\beta + \frac{1}{2}\sum_{i=1}^N x_i^2}{\sigma^2}\right]\,, \tag{3.15}$$

itself an inverse gamma distribution with parameters $\alpha' = \alpha + N/2$ and $\beta' = \beta + \frac{1}{2}\sum_{i=1}^N x_i^2$. The posterior mean, which is guaranteed to exist for $\alpha > 1$ when $N \geq 1$, is

$$\begin{aligned}
\hat{\sigma}^2_{\mathrm{mean}} &= \frac{\beta'}{\alpha' - 1} && (3.16)\\
&= \frac{\beta + \frac{1}{2}\sum_{i=1}^N x_i^2}{\alpha + \frac{N}{2} - 1}\\
&= \frac{N}{2(\alpha-1)+N}\frac{\sum_{i=1}^N x_i^2}{N} + \frac{2(\alpha-1)}{2(\alpha-1)+N}\frac{\beta}{\alpha-1}\,,
\end{aligned}$$

exactly the relation introduced in (3.5): the number of pseudo-samples introduced by the inverse gamma prior is $\nu = 2(\alpha - 1)$. The variance of the posterior mean is

$$\begin{aligned}\mathrm{Var}\left[\hat{\sigma}^2_{\mathrm{mean}}\right] &= \frac{1}{\left(\alpha+\frac{N}{2}-1\right)^2}\left(\beta^2+N\sigma^2+\frac{1}{4}\sigma^4(2N+N^2)-\beta^2-N\sigma^2-\frac{1}{4}N^2\sigma^4\right) \\ &= \frac{N\sigma^4}{2\left(\alpha+\frac{N}{2}-1\right)^2} \\ &= \left(\frac{N}{2(\alpha-1)+N}\right)^2 \mathrm{Var}\left[\hat{\sigma}^2_{\mathrm{MLE}}\right],\end{aligned} \tag{3.17}$$

which is strictly smaller than the variance of the MLE. Due to it being a biased estimator, to better understand the advantage of the Bayesian posterior mean compared to MLE, it is good measure to study its mean squared error:

$$\begin{aligned}\mathrm{MSE}[\hat{\sigma}^2_{\mathrm{mean}}] &= \mathrm{Var}\left[\hat{\sigma}^2_{\mathrm{mean}}\right]+\left(\mathrm{Bias}\left[\hat{\sigma}^2_{\mathrm{mean}}\right]\right)^2 \\ &= \mathrm{Var}\left[\hat{\sigma}^2_{\mathrm{mean}}\right]+\left(\mathbb{E}\left[\hat{\sigma}^2_{\mathrm{mean}}\right]-\sigma^2\right)^2 \\ &= \left(\frac{N}{2(\alpha-1)+N}\right)^2 \mathrm{MSE}\left[\hat{\sigma}^2_{\mathrm{MLE}}\right]+\left(\frac{2(\alpha-1)}{2(\alpha-1)+N}\right)^2\left(\frac{\beta}{\alpha-1}-\sigma^2\right)^2.\end{aligned} \tag{3.18}$$

It can be observed that, for large N, the MSE of the posterior mean converges to the MSE of the MLE. For small N, as long as the second term is not larger than the MSE of the MLE, then the posterior mean will estimate σ^2 more effectively. This happens when the prior mean is not too far from the true value.

All of this can be verified in figures 3.1 and 3.2, which were obtained for a sample with $\sigma^2 = 1$. In the first figure, it can be seen that for an inverse gamma prior with $\alpha = 1.2$ and $\beta = 1$, the MSE is worse than the MLE. This is to be expected, because the second term in equation (3.18) dominates, since the mean of $\frac{\beta}{\alpha-1} = 5$ is significantly larger than σ^2, and the number of pseudo-samples introduced by the prior is small. However, the situation changes quickly when α is increased, in particular for $\alpha = 3.2$ the MSE of the posterior mean is clearly lower than the MLE. The situation continues to improve as α increases, but then it reverses, with $\alpha = 5.2$ the MSE increases again. This is because the bias introduced by the prior weighs more than the variance reduction it provides, however note how choosing significantly different values of α for the prior distribution still yields an estimator $\hat{\sigma}^2_{\mathrm{mean}}$ that reliably outperforms the MLE in terms of mean squared error. The prior improving the performance of the posterior to estimate parameters is a property that has general validity, even beyond conjugate priors. While the effect of conjugate

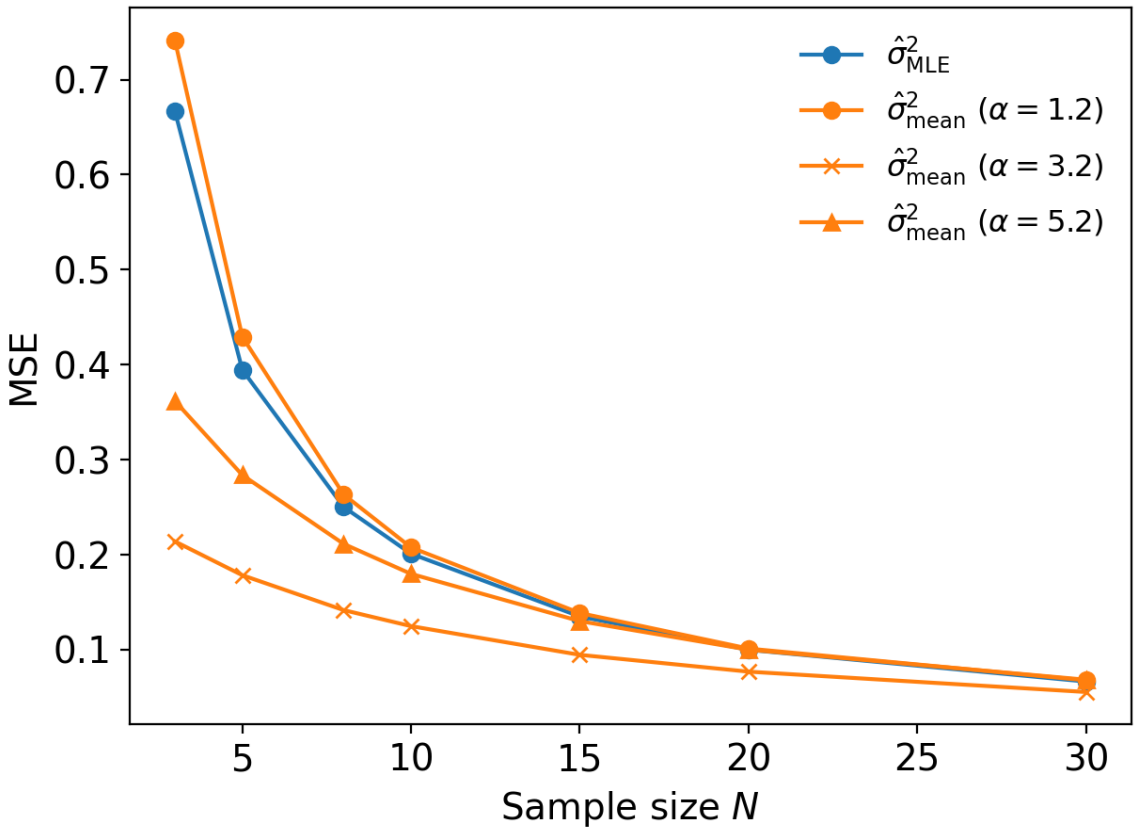

Figure 3.1: Comparison between mean squared error (MSE) of the maximum likelihood estimator (MLE, in blue) and MSE of the posterior mean (in orange) obtained with inverse gamma priors chosen for different values of α (keeping $\beta = 1$).

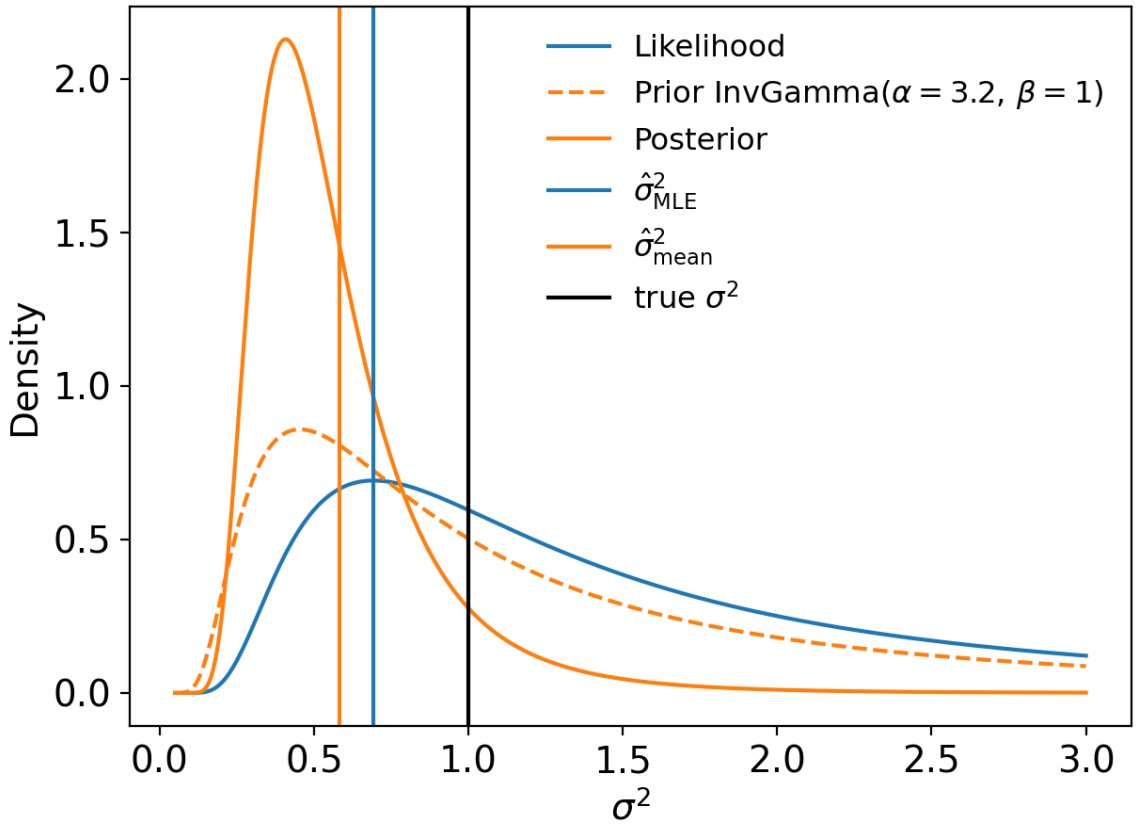

Figure 3.2: Likelihood and MLE (in blue) against the posterior and its mean (in orange) for $N = 5$ samples. The prior for $\alpha = 3.2$ and $\beta = 1$ is shown with a dashed line.

priors is made clear by the introduction of pseudo-samples, the effect of a generic prior is more complicated to characterize. Considering the logarithm of the posterior

$$\log P(\theta \,|\, \{x\}_{i=1}^{N}) = C + \log \mathcal{L}(\{x\}_{i=1}^{N} \,|\, \theta) + \log \pi(\theta)\,, \tag{3.19}$$

the prior looks like a regularization term, that adds to the likelihood in order to suppress or enhance certain values of θ. When N is small, the likelihood can have high variance, small fluctuations in the data lead to large changes in θ, and if the number of parameters to estimate is larger than N, it can lead to a likelihood with many maxima, so the MLE is not unique. The regularization introduced by the prior helps estimators by focusing the posterior on certain intervals of θ. This demonstration on how powerful priors can be should be sufficient to show how Bayesian statistics provides an ideal tool to physicists, however it should also be taken as a warning. While the value of inferred parameters are reasonably protected by the Bernstein Von-Mises theorem in the regime of large data, an improper choice of priors can significantly skew the results, and lead to the wrong conclusions.

Coming back to the example of the conjugate prior, the choice of posterior mean over other possible estimators, like the MAP, is not random. If the MSE is chosen as a *loss function*, a metric to measure the performance of the estimator, then the posterior mean is the estimator that minimizes its expected value given some data $\{x\}_{i=1}^{N}$. An estimator that minimizes the expectation of a loss function in this way is called a *Bayes estimator*. From equation (3.18) it is immediate to see that $\mathrm{MSE}[\hat{\theta}] = \mathbb{E}[L(\hat{\theta}, \theta)] = \mathbb{E}[(\hat{\theta} - \theta)^2]$, however here the expected value is over the data given θ. Instead the Bayes estimator is given by the expected value over θ given the data, in other words the expectation computed with respect of the posterior $P(\theta \,|\, \{x\}_{i=1}^{N})$,

$$\mathbb{E}[L(\hat{\theta}, \theta) \,|\, \{x\}_{i=1}^{N}] = \hat{\theta}^2 - 2\hat{\theta}\mathbb{E}[\theta \,|\, \{x\}_{i=1}^{N}] + \mathbb{E}[\theta^2 \,|\, \{x\}_{i=1}^{N}]\,. \tag{3.20}$$

The Bayes estimator is the maximum,

$$\begin{aligned} 0 &= \left.\frac{\mathrm{d}\mathbb{E}[L(\hat{\theta}, \theta) \,|\, \{x\}_{i=1}^{N}]}{\mathrm{d}\hat{\theta}}\right|_{\hat{\theta}_{\mathrm{Bayes}}} \\ &= 2\hat{\theta}_{\mathrm{Bayes}} - 2\mathbb{E}[\theta \,|\, \{x\}_{i=1}^{N}] \\ \hat{\theta}_{\mathrm{Bayes}} &= \int \mathrm{d}\theta\, \theta P(\theta \,|\, \{x\}_{i=1}^{N})\,, \end{aligned} \tag{3.21}$$

which corresponds exactly to the posterior mean. Any choice of loss function is valid, and different $L(\hat{\theta}, \theta)$ will result in different estimators. This is result is particularly useful, because it provides a way to construct estimators that are optimal for a given problem where some performance metric is better suited than others.

3.1.2 Quantification of uncertainty

While estimating a single value for some quantity of interest is fundamental for the process of inference, statistics is all about probabilities and uncertainty, so it is desirable to have some measure of how uncertain a specific estimate is. In the frequentist framework, uncertainty is measured by confidence intervals: given a sample $\{x\}_{i=1}^{N}$, a $1-\alpha$-confidence interval is an interval $\left(a(\{x\}_{i=1}^{N}),\, b(\{x\}_{i=1}^{N})\right)$ that includes the real value of a parameter θ a fraction $1-\alpha$ of the times. This means that, for many experiments where the sample is collected, $(a,\, b)$ will box θ a fraction $1-\alpha$ of the times. Section 3.1.1 touched on how estimators $\hat{\theta}$ are random variables distributed according to some sampling distribution; consider the example of the previous section, of a normally distributed sample $x_i \sim \mathcal{N}(0,\, \sigma)$ with unknown variance. The goal is to find $(a,\, b)$ so that $P(a < \sigma < b) = 1-\alpha$. The MLE $\hat{\sigma}^2_{\text{MLE}}$ is distributed according to a chi square distribution with N degrees of freedom when rescaled,

$$\hat{\varsigma} = \frac{n\hat{\sigma}^2_{\text{MLE}}}{\sigma^2} = \frac{1}{\sigma}\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} \tilde{x}_i \sim \chi^2(N)\,. \tag{3.22}$$

From the CDF of $\chi^2(N)$, X_N, we can find $\chi^2_{N,\alpha/2} = X_N^{-1}(\alpha/2)$ and $\chi^2_{N,1-\alpha/2} = X_N^{-1}(1-\alpha/2)$, which lead to the probability

$$\begin{aligned} 1-\alpha &= P\left(\chi^2_{N,\alpha/2} < \hat{\varsigma} < \chi^2_{N,1-\alpha/2}\right) \\ &= P\left(\chi^2_{N,\alpha/2} < \frac{n\hat{\sigma}^2_{\text{MLE}}}{\sigma^2} < \chi^2_{N,1-\alpha/2}\right) \\ &= P\left(\frac{n\hat{\sigma}^2_{\text{MLE}}}{\chi^2_{N,\alpha/2}} < \sigma^2 < \frac{n\hat{\sigma}^2_{\text{MLE}}}{\chi^2_{N,1-\alpha/2}}\right)\,. \end{aligned} \tag{3.23}$$

This method requires knowledge of the sampling distribution of the estimator and a transformation in terms of the true value like the one presented in equation (3.22). Due to the fact that, for many estimators, the sampling distribution converges to a normal distribution, the above method is often carried out using its CDF. When the

sampling distribution is unknown or difficult to treat, or the model used for the sample is non-parametric, then the *bootstrap method* provides a straightforward way to estimate confidence intervals:

1. given a sample $\{x\}_{i=1}^N$, construct the empirical distribution function $E_N(x) = \frac{1}{N}\sum_{i=1}^N \mathbb{I}(x_i \leq x)$, where the indicator function $\mathbb{I}(x_i \leq x) = 1$ when the argument is satisfied, and 0 otherwise;

2. draw a new sample $\{x^*\}_{i=1}^N$ from E_N, and compute the estimator $\hat{\theta}_1(\{x^*\}_{i=1}^N)$;

3. repeat the previous step M times, generating the sequence $\{\hat{\theta}\}_{i=1}^M$;

4. finally, compute the variance of $\hat{\theta}$ from the sequence,

$$\widehat{\mathrm{Var}}[\hat{\theta}] = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \frac{1}{M}\sum_{j=1}^{M}\hat{\theta}_j\right) .$$

In the Bayesian setting, uncertainty over a parameter is automatically accounted for by construction, since the posterior distribution is by definition the probability density function of the parameters of interest. One can just compute the variance of the posterior, or alternatively some quantiles using the posterior's CDF in order to obtain confidence intervals. This requires some way to obtain a sample of the posterior distribution, a problem addressed in section 3.2.

An important quantity when it comes to the uncertainty of estimators is the *Fisher information*. Consider a likelihood of some sample, $\mathcal{L}(\{x\}_{i=1}^N \,|\, \theta)$. The amount of information encoded by the sample on θ can be intuitively seen as the response of $\mathcal{L}$ as θ changes. This statement can be naturally expressed by the derivative of the (logarithm of the) likelihood with respect to the parameter,

$$s(\{x\}_{i=1}^N \,|\, \theta) = \frac{\mathrm{d}\log\mathcal{L}(\{x\}_{i=1}^N \,|\, \theta)}{\mathrm{d}\theta} = \sum_{i=1}^{N}\frac{\mathrm{d}\log f(x_i \,|\, \theta)}{\mathrm{d}\theta} = \sum_{i=1}^{N} s(x_i \,|\, \theta) \,. \tag{3.24}$$

$s(x \,|\, \theta)$ is called the score function, and its expected value is 0:[3]

[3]Note that there can be cases where the expectation of the score does not vanish. An example where this can happen is when the domain of the integral is dependent on θ, in which case the derivative cannot be brought out of the integral.

$$\begin{aligned}\mathbb{E}[s] &= \int \mathrm{d}x\, s(x\,|\,\theta) f(x\,|\,\theta) \\ &= \int \mathrm{d}x\, \frac{\mathrm{d}\log f(x\,|\,\theta)}{\mathrm{d}\theta} f(x\,|\,\theta) \\ &= \int \mathrm{d}x\, \frac{\mathrm{d}f(x\,|\,\theta)}{\mathrm{d}\theta} \\ &= \frac{\mathrm{d}}{\mathrm{d}\theta} \int \mathrm{d}x\, f(x\,|\,\theta) = 0\,.\end{aligned} \tag{3.25}$$

The variance of the score function however is not 0,

$$\begin{aligned}F(\theta) &= \mathrm{Var}[s] \\ &= \mathbb{E}[s^2] \\ &= \int \mathrm{d}x\, \frac{1}{f^2(x\,|\,\theta)} \left(\frac{\mathrm{d}f(x\,|\,\theta)}{\mathrm{d}\theta}\right)^2 f(x\,|\,\theta) \\ &= -\int \mathrm{d}x\, \frac{1}{f(x\,|\,\theta)} \frac{\mathrm{d}^2 f(x\,|\,\theta)}{\mathrm{d}\theta^2} f(x\,|\,\theta) + \int \mathrm{d}x\, \frac{1}{f^2(x\,|\,\theta)} \left(\frac{\mathrm{d}f(x\,|\,\theta)}{\mathrm{d}\theta}\right)^2 f(x\,|\,\theta) \\ &= -\int \mathrm{d}x\, \frac{\mathrm{d}^2 \log f(x\,|\,\theta)}{\mathrm{d}\theta^2} f(x\,|\,\theta) \\ &= -\mathbb{E}\left[\frac{\mathrm{d}^2 \log f(x\,|\,\theta)}{\mathrm{d}\theta^2}\right].\end{aligned} \tag{3.26}$$

$F_N(\theta) = -\sum_{i=1}^N \mathbb{E}\left[\frac{\mathrm{d}^2 \log f(x_i\,|\,\theta)}{\mathrm{d}\theta^2}\right] = -\mathbb{E}\left[\frac{\mathrm{d}^2 \log \mathcal{L}(\{x\}_{i=1}^N\,|\,\theta)}{\mathrm{d}\theta^2}\right]$ is the Fisher information of the entire sample. The Fisher information can be defined also in the multivariate case for a vector of parameters $\vec{\theta}$, leading to the matrix

$$F_{ij} \equiv F_N(\theta_i,\, \theta_j) = -\mathbb{E}\left[\frac{\partial^2 \log \mathcal{L}(\{\vec{x}\}_{i=1}^N\,|\,\vec{\theta})}{\partial\theta_i \partial\theta_j}\right]. \tag{3.27}$$

To see why the Fisher matrix is important, consider an estimator $\hat{\theta}$; from the definition of the score function in equation (3.24), the identity $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}[\hat{\theta}] = 1 + \frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{Bias}[\hat{\theta}] = \mathbb{E}[s\hat{\theta}] = \mathrm{Cov}[s,\, \hat{\theta}]$ is trivial. The Cauchy-Schwarz inequality for probability theory states

$$\mathrm{Cov}^2[x,\, y] \leq \mathrm{Var}[x]\mathrm{Var}[y]\,, \tag{3.28}$$

then it follows that

$$1 + \frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{Bias}[\hat{\theta}] \leq \mathrm{Var}[s]\mathrm{Var}[\hat{\theta}] \\ = F(\theta)\mathrm{Var}[\hat{\theta}]\,. \tag{3.29}$$

This result takes the name of *Cramer-Rao bound*, and it shows that the variance of any estimator $\hat{\theta}$ is lower bounded by the inverse of the Fisher information. Crucially, it also shows that biased estimators can reach arbitrarily low variance as long as the bias is a decreasing function of the parameter to estimate. It can be shown that $\hat{\theta}_{\mathrm{MLE}}$ saturates the bounds asymptotically, as $N \to \infty$.

The Cramer-Rao bound will be used extensively in chapter 4; it and has important practical applications, because by computing the Fisher information we can estimate a lower bound on the variance of a test summary statistic of choice, and predict how well it will constrain some parameters from the data. In cosmology, it is usual to generate synthetic data according to the specifications of some future experiment, and pick the best estimators for some task based on the Fisher information.

3.2 Markov chain Monte Carlo

Most of the results discussed in the previous section of this chapter deal with samples of i.i.d. random variables, in which case the joint probability of the sample $\{x\}_{i=1}^{N}$ is $P(\{x\}_{i=1}^{N}) = \prod_{i=1}^{N} P(x_i)$. When a sample contains dependent random variables, then in general $P(\{x\}_{i=1}^{N}) = P(x_1)P(x_2\,|\,x_1)\cdots P(x_N\,|\,x_1, \cdots, x_{N-1}) = \prod_{i=1}^{N} P(x_i\,|\,\{x\}_{j=1}^{i-1})$. Any stochastic process generates samples with a joint probability density factorizable in this way, and the exact structure of conditional dependence will depend from process to process. The stochastic process discussed in this section is the one with the simplest conditional dependence,

$$P(\{x\}_{i=1}^{N}) = \prod_{i=1}^{N} P(x_i\,|\,x_{i-1})\,. \tag{3.30}$$

Such a stochastic process is called a *Markov chain* (MC), and it proves to be very useful for sampling even the most complex posterior probability densities. The probability $P(X_{t+1} = x_j\,|\,X_t = x_i)$ takes the name of transition probability, and defines how likely it is that the process, the chain, transitions from some state with value x_j at time $t+1$ from state x_i at time t. If the same probability does not change with time (that is, the chain

retains the same transition probability from x_i to x_j regardless of when the transition happen), then the MC is said to be *homogeneous*. Consider a probability distribution $f(x)$, if the following condition holds,

$$p(x \,|\, y) f(y) = p(y \,|\, x) f(x), \tag{3.31}$$

then $f(x)$ satisfies *detailed balance*. An homogenous MC with transition probability $p(x \,|\, y)$ can converge to a stationary distribution $f(x)$ that satisfies equation (3.31) if it enjoys a few properties. For starters, a stationary distribution is such that $f(x) = \int \mathrm{d}y \, p(x \,|\, y) f(y)$, so $f(x)$ is the probability of the chain to transition from all states y to x weighted by the probability of being in y, or in other words, it just gives the probability of the chain of being in x at any given step; this implies that the chain must be able to reach any state x from any state y, if so the MC is said to be *ergodic*. If the MC does this without periodically returning to the same states, then detailed balance is satisfied, and $f(x)$ is unique. As the number N of steps goes to infinity, the chain will converge asymptotically to $f(x)$.

Detailed balance guarantees that a MC can be defined with a transition probability such that the chain will sample any PDF $f(x)$. This powerful result fully realizes the practical application of the Bayesian methods discussed in this chapters. Here follow three *Markov chain Monte Carlo* (MCMC) algorithms that implement this sampling principle.

3.2.1 Metropolis-Hastings sampling

Given a distribution $f(\vec{\theta})$ and an initial guess $\vec{\theta}_0$, a MC can be constructed by:

1. sampling a vector $\vec{\theta}'$ according to some known proposal distribution $p(\vec{\theta}' \,|\, \vec{\theta}_0)$ (a popular choice is a normal distribution);

2. accept the proposal with probability[4]

$$\alpha(\vec{\theta}' \,|\, \vec{\theta}_0) = \min \left[1, \, \frac{f(\vec{\theta}')}{f(\vec{\theta}_0)} \frac{p(\vec{\theta}_0 \,|\, \vec{\theta}')}{p(\vec{\theta}' \,|\, \vec{\theta}_0)} \right] ; \tag{3.32}$$

3. if the proposal is accepted, which can be done by verifying that $x \leq \alpha(\vec{\theta}' \,|\, \vec{\theta}_0)$, where $x \sim \mathrm{Uniform}(0, 1)$ is sampled uniformly in the unit interval, then set $\vec{\theta}_1 = \vec{\theta}'$,

[4]Note that this doesn't require $f(\vec{\theta})$ to be properly normalized, so it is especially suited for Bayesian approaches, where the normalization of the posterior can be expensive to compute.

otherwise reject the proposal and repeat.

The transition probability from $\vec{x}$ to $\vec{y}$ for a chain sampled with Metropolis-Hastings (MH) is $\alpha(\vec{y}\,|\,\vec{x})p(\vec{y}\,|\,\vec{x})$, since the proposal distribution provides a proposed step with some probability p, which is then accepted (independently) with probability given by α. Then, for detailed balance to hold, the product of the transition probability and the stationary distribution must be commutative in $\vec{x}$ and $\vec{y}$:

$$\begin{aligned}\alpha(\vec{y}\,|\,\vec{x})p(\vec{y}\,|\,\vec{x})f(\vec{x}) &= \min\left[1,\ \frac{f(\vec{y})}{f(\vec{x})}\frac{p(\vec{x}\,|\,\vec{y})}{p(\vec{y}\,|\,\vec{x})}\right]p(\vec{y}\,|\,\vec{x})f(\vec{x}) \\ &= \min\left[p(\vec{y}\,|\,\vec{x})f(\vec{x}),\ f(\vec{y})p(\vec{x}\,|\,\vec{y})\right] \\ &= \min\left[\frac{p(\vec{y}\,|\,\vec{x})}{p(\vec{x}\,|\,\vec{y})}\frac{f(\vec{x})}{f(\vec{y})},\ 1\right]p(\vec{x}\,|\,\vec{y})f(\vec{y}) \\ &= \alpha(\vec{x}\,|\,\vec{y})p(\vec{x}\,|\,\vec{y})f(\vec{y})\,,\end{aligned} \tag{3.33}$$

MH is guaranteed to sample a MC that converges to the stationary distribution of choice. There's freedom in the way the proposal distribution is chosen. A common choice is a normal distribution centered in the previous step and with variance/covariance tuned to increase sampling efficiency.

3.2.2 Gibbs sampling

Given a distribution $f(\vec{\theta})$ for which we can sample the conditional distributions $f(\theta_i\,|\,\{\theta\}_{j\neq i})$ and an initial guess $\vec{\theta}_0$, a MC can be constructed by:

1. sampling each θ_i' from $f(\theta_i'\,|\,\{\theta_0\}_{j\neq i})$;
2. accept θ_i with probability 1, as it is sampled directly from the distribution of choice.

Since steps are always accepted, the transition probability of each single update from $\vec{x}$ to y_i for a chain sampled with the Gibbs algorithm is $1\cdot f(y_i\,|\,\{x\}_{j\neq i})$. For detailed balance,

$$\begin{aligned}f(y_i\,|\,\{x\}_{j\neq i})f(\vec{x}) &= f(y_i\,|\,\{x\}_{j\neq i})f(\{x\}_{j\neq i})f(x_i\,|\,\{x\}_{j\neq i}) \\ &= f(y_i,\,\{x\}_{j\neq i})f(x_i\,|\,\{x\}_{j\neq i})\,,\end{aligned} \tag{3.34}$$

which means that the chain can move back and forth between x_i and y_i while all the other parameters are fixed, detailed balance is satisfied. The Gibbs MCMC algorithm is very efficient and works for distributions where all conditional distributions are known; it can even be applied in combination with other MCMC algorithms when only a subset of conditional distributions is known.

3.2.3 Hamiltonian Monte Carlo sampling

Consider $\Phi(\vec{\theta}) \equiv -\log f(\vec{\theta})$. This logarithm can be treated as the potential term of a Hamiltonian system:

$$H(\vec{\theta},\, \vec{p}) = \frac{1}{2}\vec{p}^{T} M^{-1}\vec{p} + \Phi(\vec{\theta})\,, \tag{3.35}$$

with $\vec{\theta}$ the parameters we are interested in and $\vec{p}$ their associated momenta in phase space. The Hamiltonian describes a dynamical system governed by Hamilton's equations:

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i}\,, \quad \frac{\mathrm{d}p_i}{\mathrm{d}t} = \frac{\partial H}{\partial \theta_i} = -\frac{\partial \log f}{\partial \theta_i}\,. \tag{3.36}$$

Solutions for this system of coupled differential equations can be readily found using a numerical integrator. Symplectic integrators are used because they preserve energy, making them especially suitable for Hamiltonian systems. How does equation (3.35) help in sampling $f(\vec{\theta})$? Consider

$$e^{-H(\vec{\theta},\vec{p})} = f(\vec{\theta}) \exp\left[-\frac{1}{2}\vec{p}^{T} M^{-1}\vec{p}\right], \tag{3.37}$$

this is our target distribution times a multivariate normal distribution over the momenta. $e^{-H(\vec{\theta},\vec{p})}$ takes the name of *canonical distribution*, and its marginal over $\vec{p}$ is the target distribution to be sampled. Starting from the canonical distribution and an initial guess $\vec{\theta}_0$, a MC can be constructed by:

1. sampling an initial momentum $\vec{p}_i$ from the kinetic term, in this case $\vec{p}_i \sim \mathcal{N}_D(0,\, M)$, where D is the number of parameters;

2. solving the system in equations (3.36) by integrating numerically for some time $\tau = N_{\text{step}}\Delta\tau$ from the initial state $(\vec{\theta}_0,\, \vec{p}_i)$ up to some final state $(\vec{\theta}',\, -\vec{p}_f)$, where the minus sign is needed to enforce detailed balance;

3. accepting the proposal with probability

$$\alpha(\vec{\theta}', \vec{p}_f \,|\, \vec{\theta}_0, \vec{p}_i) = \min\left[1, \exp\left[H(\vec{\theta}', -\vec{p}_f) - H(\vec{\theta}_0, \vec{p}_i)\right]\right] = 1\,, \tag{3.38}$$

since the proposed step is part of a trajectory that preserves the Hamiltonian.

The random sampling of the initial momentum effectively acts as a proposal distribution like in the Metropolis-Hastings algorithm, which explains the acceptance probability and also trivially satisfies detailed balance. Hamiltonian Monte Carlo (HMC) is especially suited for sampling in high-dimensional parameter spaces.

3.2.4 Which algorithm to use and some caveats

The algorithms introduced in the previous sections are all guaranteed to generate chains that converge to any target distribution by satisfying detailed balance. However specific algorithms are best suited for sampling specific distributions. When conditional distributions are known, Gibbs sampling is always a good choice due to its simplicity and guaranteed acceptance rate; this is especially true for distributions with highly correlated or degenerate parameters, since their conditional distributions are well behaved. MH works well in a low and moderate number of dimensions and with distributions that are not differentiable in $\vec{\theta}$, however it can struggle with highly correlated parameters, requiring reparametrizations and tuning: take a proposal $\vec{\theta}' \sim \mathcal{N}(\vec{\theta}_i, \Sigma)$, then the covariance Σ need not be fixed a priori; Σ can be chosen to be simple, like the identity matrix, and tuned later by estimating the covariance of some initial "warm-up" sample. The random walk behavior of MH sampling fails when the number of dimensions grows to be very high. This is because of the unintuitive phenomenon of "concentration of measure": when the number of dimensions grows, volumes tend to concentrate around a region of space that takes the name of *typical set*. For a sphere in D dimensions, its volume will concentrate within a thin annulus, not around the center, as one would expect. For probability density functions, this means that the probability mass located at the typical set is large compared to the region of high probability; so to sample it effectively, the sampler needs to stay within the typical set. Due to the expression of the acceptance probability, MH tends to converge to regions where the probability *density* is high, which is not where the typical set resides in general. HMC solves this issue for differentiable target distributions by exploiting the geometry of the high-dimensional space, which is built into the Hamiltonian. Instead of converging towards the mode, where the probability mass is low, HMC explores equal energy surfaces that, by construction, foliate the typical set. Like MH, the mass matrix M in equation (3.35) can be tuned with a warm-up phase.

HMC is very powerful, but requires extra tuning (for example, for the integration time τ), and the choice of kinetic term can also be dependent on the problem at hand [105]. Another subtlety to consider is that the symplectic numerical integrators adopted to solve equations (3.36) are prone to numerical oscillations around constant energy trajectories, so the acceptance rate is never truly 1, but more around the range of 0.7 to 0.9.

Once the algorithm is chosen, it can be run until a certain convergence criterion is reached, obtaining a D-dimensional sample $\{\vec{\theta}\}_{i=1}^{N}$. From this, mean and variance of each parameter can be computed by using the law of large numbers,

$$\bar{\theta}_i = \frac{1}{N}\sum_{j=1}^{N}\theta_{i,j}\,, \quad \bar{\sigma}^2_{\theta_i} = \frac{1}{N}\sum_{j=1}^{N}\theta^2_{i,j} - \bar{\theta}^2_i\,. \tag{3.39}$$

However, since a MC is guaranteed to converge only asymptotically with its size, the algorithm will only converge to the stationary distribution after a certain initial number of steps, called the "burn-in" phase, which must be identified and discarded. There is no way to prove that a chain has converged, however there is a simple test that can be used to quantify how well mixed a chain is, that is if the chain has extensively sampled some region of space; a chain that has converged must be well mixed, but the opposite is not necessarily true. The test is called a *Gelman-Rubin* test:

1. start with M initial guesses, and run the MCMC algorithm for each, constructing M Markov chains;
2. assuming that each chain has N samples, compute the average of each chain, and the average across all chains.

If the chains are well mixed, and all sample the same region of parameter space, then the two averages must be close to each other, otherwise the chains must be sampling each different regions, therefore convergence has not been reached. In more quantitative terms, the Gelman-Rubin test computes the quantity

$$\hat{R} = \frac{\hat{V}}{W} = \frac{\frac{N-1}{N}W + \frac{1}{N}B}{W}\,, \tag{3.40}$$

where B is the variance of the means of each chain compared to the global mean of all the chains, and W is the average variance between chains. If all chains are well mixed, then $B \to 0$, and $\hat{R} \to 1$. The Gelman-Rubin statistic $\hat{R}$ is the golden standard when diagnosing convergence, but there are other metrics that should be used in conjunction with it:

- very low acceptance rates are usually a symptom of poor convergence, because near convergence the chains are likely to accept new steps often, save for pathological distributions;

- trace plots that show the chain samples oscillating around the same value for some parameter is likely to indicate that the chain as converged, however this fails if the distribution has more than one mode, a value of maximum density, in that parameter;

- the effective sample size $\widehat{ESS}$ provides the effective number of independent samples in the chain. Consecutive samples in a MC of length N are correlated because one step is statistically dependent on the previous one, so the size of a truly i.i.d. sample of the target distribution is given by the ratio $\widehat{ESS} = N/\tau_{\text{ac}}$, where τ_{ac} is the integrated autocorrelation of the samples.[5]

[5]The integrated autocorrelation is given by the the ratio between the covariance of the chain with itself, shifted by k steps, and the variance, summed over all possible shifts $k = 0, \cdots, N$.

Chapter 4

The Wavelet Scattering Transform for PNG Detection

The contents of this chapter have been published in the paper

Matteo Peron, Gabriel Jung, Michele Liguori and Massimo Pietroni (July 2024). "Constraining primordial non-Gaussianity from large-scale structure with the wavelet scattering transform." *Journal of Cosmology and Astroparticle Physics* **2024**(07): 021. doi:10.1088/1475-7516/2024/07/021.[1]

which has been included, integrally or in part, according to the journal's copyright policy.

4.1 Introduction

A central aim of observational cosmology in the coming years is the development of methodologies that extract maximal information from non-linear scales in galaxy surveys. This necessarily requires going beyond power spectrum analysis, since the power spectrum alone cannot capture phase information and is therefore insensitive to structures such as filaments in the cosmic web. A natural next step is the bispectrum (the three-point function of the Fourier modes of the density field), the lowest-order non-Gaussian (NG) correlator. Joint power spectrum and bispectrum (P+B) Large Scale Structure (LSS) forecasts and analyses are already widespread in the cosmological literature [4–8]; this approach offers notable advantages, including the availability of validated P+B anal-

[1]The contents of the article are accessible at `https://www.doi.org/10.1088/1475-7516/2024/07/021`.

ysis pipelines and the interpretability of bispectrum measurements (at mildly non-linear scales) within cosmological perturbation theory. On the other hand, the bispectrum does not exhaust the available information, and a direct extension to higher-order correlators becomes increasingly cumbersome to implement, especially given the need for accurate covariance matrices in the relevant non-linear regime and for realistic observational effects (e.g., window functions). These considerations motivate the search for alternative NG summary statistics. Desiderata include ease of implementation, efficiency in capturing NG information with a limited number of modes, and clear physical interpretability. Several NG summaries have been explored with this in mind, including marked spectra [9–13], power spectra from cosmic web environments [106, 107], skew spectra [14–17], Minkowski functionals [18, 19], k-nearest neighbour cumulative distribution functions [108, 109], the probability distribution function (PDF) of late-time cosmic density fluctuations [110, 111], halo mass function probes [20–25], and the void abundance [26, 27].

This work focuses on another statistic that has received significant recent attention in observational cosmology, namely the Wavelet Scattering Transform (WST) [28, 29]. The WST has the structure of a three-layer convolutional network in which the operations at each stage are, in sequence, non-linear activation via the modulus function, convolution by a wavelet filter, and low-pass spatial averaging of the data. Recent applications to simulations [30–34] indicate that the WST efficiently captures NG information and can significantly improve constraints relative to power spectrum-only analyses, particularly for neutrino mass. The goal here is to investigate the WST as a tool for the study of primordial non-Gaussianity (PNG), to compare in detail its performance with that of a P+B analysis, and to provide a physical interpretation for this comparison.

Non-Gaussian features in the primordial cosmological density field are a distinctive prediction of inflationary models beyond standard single-field slow-roll dynamics. Primordial NG signatures are encoded in the primordial bispectrum of cosmological perturbations; different departures from the single-field slow-roll scenario yield specific functional dependencies of the bispectrum on triangular configurations of Fourier modes. This model dependence makes PNG a powerful probe of inflation. At the same time, the expected PNG signal is small and, on nonlinear scales, strongly degenerate with the bispectrum produced at late times by gravitational evolution and galaxy bias. Beyond-bispectrum statistics, such as the WST, can help break this degeneracy, which is one of the main motivations for the present analysis. An alternative, more ambitious route is forward-modelled field-level inference, which is increasingly employed in cosmological data analysis and promises to extract the full information content of the data (see, e.g., [44, 47, 112–123], but also chapter 5). The more traditional approach, based on an initial

compression of the data into a set of summary statistics, remains a valid alternative: although lossy, it often facilitates physical interpretation and comparison with analytical models, and it helps identify where (in domain and scale) the bulk of information resides and how it can be efficiently recovered.

The analysis considers the three standard primordial bispectrum shapes, local, equilateral and orthogonal, and produces Fisher forecasts for the corresponding $f_{\rm NL}$ amplitude parameters, jointly with standard cosmological parameters (σ_8, Ω_m, n_s, h). To build the Fisher matrix up to a scale $k_{\rm max} = 0.5\,h\,{\rm Mpc}^{-1}$, covariance matrices and summary responses to parameter changes are computed by Monte Carlo averaging and numerical differentiation, using the publicly available Quijote and QuijotePNG N-body simulation suites. This is a set of simulations with 512^3 dark matter particles in boxes of size $1\ h^{-1}$Gpc.

The remainder is organized as follows. section 4.2 introduces the summary statistics considered in the analysis: power spectrum, bispectrum and WST. section 4.3 describes the methodology to compute Fisher matrices. section 4.4 presents the input Quijote simulation dataset. section 4.5 details the analysis and the main results. section 4.6 discusses the interpretation of the results and compares the performance of P+B and WST. section 4.7 summarizes the conclusions.

4.2 Summary statistics

4.2.1 Power spectrum and bispectrum

To extract cosmological information from the density contrast field $\delta(\vec{x})$, its two- and three-point correlators in Fourier space are considered: the power spectrum and the bispectrum.

From a Fourier-transformed density field $\delta(\vec{k})$ defined on a three-dimensional grid, the power spectrum $P(k)$ is obtained using a binned estimator (see, e.g., [124]):

$$\hat{P}(k_i) = \frac{1}{VN_i} \sum_{k\in\Delta_i} \delta(\vec{k})\delta^*(\vec{k})\,, \tag{4.1}$$

where V is the surveyed volume, Δ_i defines the binning of the relevant k-range, N_i is the count of modes k in the i-th bin and k_i the average modulus.

Rather than computing the binned bispectrum $B(k_1, k_2, k_3)$, a modal approach based

on its expansion on a basis of separable functions (i.e., symmetrized products of one-dimensional functions in k) is adopted [125–128]. With this choice, evaluating the information content of the bispectrum reduces to estimating modal modes of the form

$$\hat{\beta}_n = \frac{1}{V}\int \mathrm{d}^3x\, M_p(\vec{x})M_q(\vec{x})M_r(\vec{x}), \quad \text{with} \quad M_p(\vec{x}) \equiv \int \frac{\mathrm{d}^3k}{(2\pi)^3}\frac{q_p(k)\delta(\vec{k})}{\sqrt{kP(k)}}e^{i\vec{k}\cdot\vec{x}}\,, \tag{4.2}$$

where the $q_p(k)$ are the chosen one-dimensional functions for the bispectrum expansion. As shown in [129–131], only a small number of such functions, constructed from simple polynomials (e.g., Legendre polynomials) or from the tree-level matter bispectrum (typically of order 100 modal modes β_n) is sufficient to study the matter and halo bispectrum from linear to non-linear scales ($k_{\max} = 0.5\,h\,\mathrm{Mpc}^{-1}$).

4.2.2 Wavelet scattering transform

The observables in the WST are the *scattering coefficients*, built from nonlinear operations on $\delta(\vec{x})$ and organized in different orders, or *layers*. This work adopts the implementation encoded in the KYMATIO[2] package [29]. The zeroth-order coefficient is the spatial average of the modulus of $\delta(\vec{x})$ raised to a power q,

$$S_0^q \equiv \langle|\delta(\vec{x})|^q\rangle\,. \tag{4.3}$$

To obtain higher-order coefficients a *mother wavelet* function is defined, which, for KYMATIO, is

$$\psi_{0\,l}^m(\vec{x},\,\sigma) \equiv \frac{C_l}{(2\pi\sigma^2)^{3/2}}\exp\left[-\frac{x^2}{2\sigma^2}\right]R_l^m\left(\frac{\vec{x}}{\sigma}\right)\,, \tag{4.4}$$

where $x \equiv |\vec{x}|$, $\hat{x} \equiv \vec{x}/x$, and the $R_l^m(\vec{r})$ are the *regular solid harmonics*,

$$R_l^m(\vec{r}) \equiv \sqrt{\frac{4\pi}{2l+1}}\,r^l\,Y_l^m(\hat{r})\,, \tag{4.5}$$

with $Y_l^m(\hat{r})$ the Laplacian Spherical Harmonics. The coefficients C_l are given in appendix A.2. From the mother wavelet, a family of rescaled ones, $\psi_{j\,l}^m$, with $j = 1, 2, \cdots$, is generated by successive doublings of the length scale σ,

[2] `https://www.kymat.io`

$$\psi^m_{jl}(\vec{x},\,\sigma) \equiv \psi^m_{0l}(\vec{x},\,2^j\sigma) = 2^{-3j}\,\psi^m_{0l}(2^{-j}\vec{x},\,\sigma)\,. \tag{4.6}$$

Following [30], in this analysis based on simulation data the choice

$$\sigma = 0.8\,\frac{L}{N^{1/3}}\,, \tag{4.7}$$

is adopted, where L is the side of the cubic box and N the number of grid points.

In Fourier space, the wavelets read

$$\tilde{\psi}^m_{jl}(\vec{p},\,\sigma) \equiv \int d^3r\, e^{i\vec{p}\cdot\vec{r}}\psi^m_{jl}(\vec{x},\,\sigma) = i^l C_l\,\exp\left[-\frac{(2^j\sigma\,p)^2}{2}\right]\,R^m_l\left(2^j\sigma\vec{p}\right)\,, \tag{4.8}$$

with $p \equiv |\vec{p}|$, $\hat{p} \equiv \vec{p}/p$.

First-order coefficients are then defined as

$$S^q_1(j,l) \equiv \left\langle \left(\sum_{m=-l}^{l} \left| \delta(\vec{x}) * \psi^m_{jl}(\vec{x},\,\sigma) \right|^2 \right)^{\frac{q}{2}} \right\rangle\,, \tag{4.9}$$

where the symbol "$*$" indicates convolution in configuration space. Due to the sum over m, the coefficients depend, for fixed q, on j and l only. In the following, results are also shown for the special case $q = 2$, where the interpretation of these coefficients simplifies considerably. In this case, the first-order coefficients are directly related to the nonlinear power spectrum $P(p)$,

$$\begin{aligned} S^2_1(j,l) &= \int \frac{d^3p}{(2\pi)^3}\frac{d^3p'}{(2\pi)^3} e^{-i\vec{x}\cdot(\vec{p}+\vec{p}')} \langle \tilde{\delta}(\vec{p})\tilde{\delta}(\vec{p}')\rangle\, \tilde{\psi}^m_{jl}(\vec{p})\tilde{\psi}^m_{jl}(\vec{p}')\,, \\ &= C_l^2 \int \frac{d^3p}{(2\pi)^3} W_l^2\left[2^j\sigma p\right]\; P(p) \\ &= C_l^2\,\langle \delta^2(\vec{x};\,j,\,l)\rangle\,, \end{aligned} \tag{4.10}$$

where the property of spherical harmonics has been used,

$$\frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y^m_l(\hat{p}) Y^m_l(-\hat{p}) = 1\,, \tag{4.11}$$

and the Fourier-space filter is defined as

$$W_l[z] \equiv z^l \, \exp(-z^2/2)\,, \tag{4.12}$$

showing that, for fixed variance of the Gaussian filter (set by j), higher values of l probe larger wavenumbers. The last line of equation 4.10 is the space-averaged variance of the inverse Fourier transform of the filtered density field $\tilde{\delta}(\vec{p};\, j,\, l) \equiv W_l\,[2^j \sigma p]\,\tilde{\delta}(\vec{p})$.

To define second-order coefficients, first introduce the nonlinear field

$$U(j_1,\, l)(\vec{x}) \equiv \left(\sum_{m=-l}^{l} \left| \delta(\vec{x}) * \psi^m_{j_1 l}(\vec{x},\, \sigma) \right|^2 \right)^{\frac{1}{2}}\,, \tag{4.13}$$

and then convolve it again with the wavelet at scale j_2, using the same operation as in 4.9,

$$S_2^q(j_2,\, j_1,\, l) \equiv \left\langle \left(\sum_{m=-l}^{l} \left| U(j_1,\, l)(\vec{x}) * \psi^m_{j_2 l}(\vec{x},\, \sigma) \right|^2 \right)^{\frac{q}{2}} \right\rangle\,, \tag{4.14}$$

where, following [30], the same value of l is used for both the internal (j_1) and external (j_2) convolution.

The $q = 2$ case yields

$$\begin{aligned} S_2^2(j_2,\, j_1,\, l) &= C_l^2 \int \frac{d^3p}{(2\pi)^3} W_l \left[2^{j_2}\sigma p\right]^2 \; P_U(p; j_1,\, l)\,, \\ &= C_l^2 \, \langle U^2(\vec{x};\, j_2,\, j_1,\, l) \rangle\,, \end{aligned} \tag{4.15}$$

where the power spectrum in the integral refers to the field $U(j_1,\, l)(\vec{x})$,

$$\langle \tilde{U}(j_1,\, l)(\vec{p})\, \tilde{U}(j_1,\, l)(\vec{p}') \rangle \equiv (2\pi)^3 \delta_D(\vec{p} + \vec{p}') P_U(p; j_1,\, l)\,, \tag{4.16}$$

and the filtered field $U(\vec{x};\, j_2,\, j_1,\, l)$ is the inverse Fourier transform of $W_l\,[2^{j_2}\sigma p]\,\tilde{U}(j_1,\, l)(\vec{p})$.

The $q = 2$ case is particularly useful for comparing the performance of the WST with that of $P(k)$, since all information beyond the nonlinear $P(k)$ is then encoded in scattering coefficients beyond first order. The second-order coefficients are simply the variances of the filtered fields $U(\vec{x} : j_2,\, j_1,\, l)$, and therefore contain information on the bispectrum and higher-order correlators.

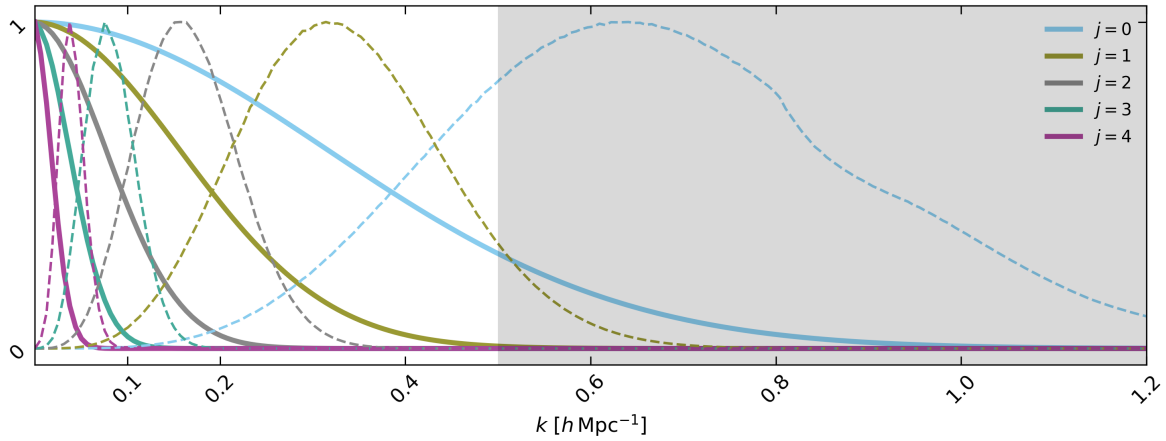

Figure 4.1: Radial profile of the wavelets at $l = 0$ (solid lines) and $l = 4$ (dashed lines) at $m = 0$ and for different values of j, where $j = 0$ is equivalent to the mother wavelet. As discussed in the text, a low-pass filter with $k_{\max} = 0.5\,h\,\mathrm{Mpc}^{-1}$ is applied to the field, this is highlighted by the shaded region.

Fixing σ as in equation (4.7), the number of coefficients is set as follows. For first order, take $j = 0, 1, \cdots, J$ with $J = 4$, corresponding to Gaussian smoothing filters in real space of widths from σ to $2^4\,\sigma$. In figure 4.1, the Fourier-space filters used to analyze the Quijote suite of simulations presented in section 4.4 are shown. For the angular dependence, take $l = 0, 1, \cdots, L$ with $L = 4$. For second order, the external convolution (controlled by j_2) is restricted to scales larger than the internal one, i.e., $j_1 < j_2 = 1 \cdots, J$ with $J = 4$ as above. This yields a total of 76 coefficients: one of zeroth order, 25 of first order, and 50 of second order. As higher orders are not considered, these coefficients form the data vector.

4.3 Simulation-based Fisher matrices

To benchmark the information on PNG amplitudes, and more generally on cosmological parameters, extracted from the non-linear density field using the summary statistics of section 4.2, a standard approach is the Fisher matrix formalism, first presented at the end of section 3.1.2.

Under the reasonable assumption of normally distributed summary statistics $\vec{s}$, depending on a set of parameters $\vec{\theta}$ through their mean $\bar{\vec{s}}$, the Fisher matrix is

$$F_{ij} = \left(\frac{\partial \bar{\vec{s}}}{\partial \theta_i}\right)^{\mathrm{T}} \Sigma^{-1} \left(\frac{\partial \bar{\vec{s}}}{\partial \theta_j}\right), \tag{4.17}$$

where Σ is the covariance matrix. For each parameter θ_i, $\sqrt{(F^{-1})_{ii}}$ is the Cramér–Rao bound, i.e., the expected error bar from an optimal estimator.

To evaluate equation (4.17), a simulation-based approach is adopted (see section 4.4 for the simulations used). Unbiased estimates of inverse covariance matrices are obtained using the Hartlap/Anderson correction factor [132]:

$$\Sigma^{-1} = \frac{n_\mathrm{r} - n_\mathrm{s} - 2}{n_\mathrm{r} - 1}\,\hat{\Sigma}^{-1}, \qquad \text{with } \hat{\Sigma} = \frac{1}{n_r - 1}(\vec{s} - \bar{\vec{s}})(\vec{s} - \bar{\vec{s}})^\mathrm{T}\,, \tag{4.18}$$

where n_r is the number of realizations used and n_s the number of summary statistics. Derivatives at a fiducial point (denoted by the superscript $*$) are computed by central finite differences:

$$\left.\frac{\partial \bar{\vec{s}}}{\partial \theta_i}\right|_{\vec{\theta}^*} = \frac{\bar{\vec{s}}(\theta_i^* + \delta\theta_i) - \bar{\vec{s}}(\theta_i^* - \delta\theta_i)}{2\delta\theta_i}\,. \tag{4.19}$$

A direct application of this method requires a very large number of simulations to achieve numerical convergence; otherwise, spuriously optimistic Fisher error bars may result (see [133, 134] for a discussion in the PNG context). This issue can be mitigated using the method of [135], which computes the Fisher matrix from score-compressed statistics instead of the raw summaries. This compression is optimal, retaining full information on parameters, under the same conditions used to write equation (4.17) [136, 137]. Score-compressed statistics $\tilde{\vec{s}}$ are obtained via

$$\tilde{s}_i = \left.\frac{\partial \bar{\vec{s}}}{\partial \theta_i}\right|_{\vec{\theta}^*} \Sigma^{-1}(\vec{s} - \bar{\vec{s}}^*), \tag{4.20}$$

i.e., from the same ingredients as the Fisher matrix, enabling evaluation from the same simulation sets with faster numerical convergence.

4.4 Simulations

Accurately evaluating the statistical properties of the observables introduced in section 4.2 up to non-linear scales, as well as their dependence on cosmological parameters, requires a large number of simulations. This study uses the Quijote[3] and QuijotePNG[4] suites of N-body simulations [138, 139], which were designed for precisely this purpose.

[3] `https://quijote-simulations.readthedocs.io/en/latest/`
[4] `https://quijote-simulations.readthedocs.io/en/latest/png.html`

	$N_{\rm sims}$	σ_8	Ω_m	$\Omega_{\rm b}$	n_s	h	$f_{\rm NL}^{\rm local}$	$f_{\rm NL}^{\rm equil}$	$f_{\rm NL}^{\rm ortho}$	$M_{\rm min}(M_\odot/h)$
Fiducial	15000	0.834	0.3175	0.049	0.9624	0.6711	0	0	0	3.2×10^{13}
σ_8^+	500	0.849	0.3175	0.049	0.9624	0.6711	0	0	0	3.2×10^{13}
σ_8^-	500	0.819	0.3175	0.049	0.9624	0.6711	0	0	0	3.2×10^{13}
Ω_m^+	500	0.834	0.3275	0.049	0.9624	0.6711	0	0	0	3.2×10^{13}
Ω_m^-	500	0.834	0.3075	0.049	0.9624	0.6711	0	0	0	3.2×10^{13}
n_s^+	500	0.834	0.3175	0.049	0.9824	0.6711	0	0	0	3.2×10^{13}
n_s^-	500	0.834	0.3175	0.049	0.9424	0.6711	0	0	0	3.2×10^{13}
h^+	500	0.834	0.3175	0.049	0.9624	0.6911	0	0	0	3.2×10^{13}
h^-	500	0.834	0.3175	0.049	0.9624	0.6511	0	0	0	3.2×10^{13}
$f_{\rm NL}^{\rm local,+}$	500	0.834	0.3175	0.049	0.9624	0.6711	+100	0	0	3.2×10^{13}
$f_{\rm NL}^{\rm local,-}$	500	0.834	0.3175	0.049	0.9624	0.6711	−100	0	0	3.2×10^{13}
$f_{\rm NL}^{\rm equil,+}$	500	0.834	0.3175	0.049	0.9624	0.6711	0	+100	0	3.2×10^{13}
$f_{\rm NL}^{\rm equil,-}$	500	0.834	0.3175	0.049	0.9624	0.6711	0	−100	0	3.2×10^{13}
$f_{\rm NL}^{\rm ortho,+}$	500	0.834	0.3175	0.049	0.9624	0.6711	0	0	+100	3.2×10^{13}
$f_{\rm NL}^{\rm ortho,-}$	500	0.834	0.3175	0.049	0.9624	0.6711	0	0	−100	3.2×10^{13}
$M_{\rm min}^+$	500	0.834	0.3175	0.049	0.9624	0.6711	0	0	0	3.3×10^{13}
$M_{\rm min}^-$	500	0.834	0.3175	0.049	0.9624	0.6711	0	0	0	3.1×10^{13}

Table 4.1: The parameters of the Quijote and QuijotePNG N-body simulations and halo catalogues used in this work.

These are N-body simulations containing 512^3 dark matter particles in boxes of size 1 h^{-1}Gpc, in which the gravitational evolution of structures up to $z = 0$ is computed with the TreePM code Gadget-III [140] starting from initial conditions determined at $z = 127$ using 2LPTIC [141] or 2LPTPNG [139, 142].[5] The suite contains 15 000 simulations with Planck-like cosmological parameters [61], and subsets of 500 in which one parameter is slightly shifted above or below its fiducial value. The characteristics of these sets are summarized in Table 4.1.

In what follows, the parameter vector $\vec{\theta} = \{\sigma_8, \Omega_m, n_s, h, f_{\rm NL}^{\rm local}, f_{\rm NL}^{\rm equil}, f_{\rm NL}^{\rm ortho}\}$ is considered, and both the matter (at $z = 1$) and halo (at $z = 0$) density fields are analyzed. The definitions of the PNG parameters $f_{\rm NL}^{\rm local}$, $f_{\rm NL}^{\rm equil}$, and $f_{\rm NL}^{\rm ortho}$, characterizing different shapes of the primordial bispectrum, are given in [139]. Halos are identified using friends-of-friends (FOF) [143]. Only halos with mass above $M_{\rm min} = 3.2\times10^{13}M_\odot/h$ are included, and $M_{\rm min}$ is treated as a nuisance parameter by varying it in the halo analyses. As shown in previous analyses [134, 139], $M_{\rm min}$ can be regarded as a proxy for the linear bias parameter b_1.

[5] https://github.com/dsjamieson/2LPTPNG

The power spectra and bispectrum modal modes previously measured in [131] and [134] are used for the matter and halo fields, respectively. The matter density field is studied in real space, while the halo density field is transformed to redshift space by applying the velocity correction along the x direction.

WST coefficients are estimated from density fields obtained by interpolating dark matter particle or halo positions on regular three-dimensional grids of size $N_g = 256$ using the CIC scheme implemented in Pylians3.[6] A sharp, low-pass filter is applied to the input field $\delta(\vec{x})$ to exclude all modes above $k_{\rm max} = 0.5\,h\,{\rm Mpc}^{-1}$ (the smallest accessible scale with Quijote simulations), and all 76 scattering coefficients are then computed.

4.5 Results

4.5.1 First order scattering coefficients and $P(k)$

As shown in equation 4.10, first-order scattering coefficients with the specific choice $q = 2$, i.e., $S_1^2(j,\, l)$, contain no more information than the (nonlinear) power spectrum $P(k)$. To check this explicitly, a Fisher matrix analysis is performed on the non-linear dark matter field at redshift $z = 1$. The resulting constraints for $\vec{\theta} = \{\sigma_8,\, \Omega_m,\, n_s,\, h\}$ are displayed in figure 4.2. The near-perfect overlap between the 2D contours for $P(k)$ and $S_1^2(j,\, l)$ indicates that the two statistics carry nearly identical information. $P(k)$ yields slightly tighter constraints, suggesting that the equivalence between the 25 integrals of $P(k)$ in equation 4.10 and the 78 $P(k)$ bins is not exact. This could likely be improved by increasing the maximum j and/or l. However, to limit computational cost, this choice is retained since, as shown below, including second-order coefficients outperforms the power spectrum for all parameters, including PNG, both for matter and halos.

Constraints from first-order coefficients with $q = 0.8$ are also shown. In this case, the equivalence with $P(k)$ is lost, and some information from higher-order correlators is expected to leak into first-order scattering coefficients. As anticipated, the contours for $S_1^{0.8}(j,\, l)$ differ from those of $S_1^2(j,\, l)$ and of $P(k)$. Moreover, for all parameters except n_s, performance is worse, suggesting that moving from $q = 2$ to $q = 0.8$ redistributes information from first to higher orders.

[6] `https://github.com/franciscovillaescusa/Pylians3`

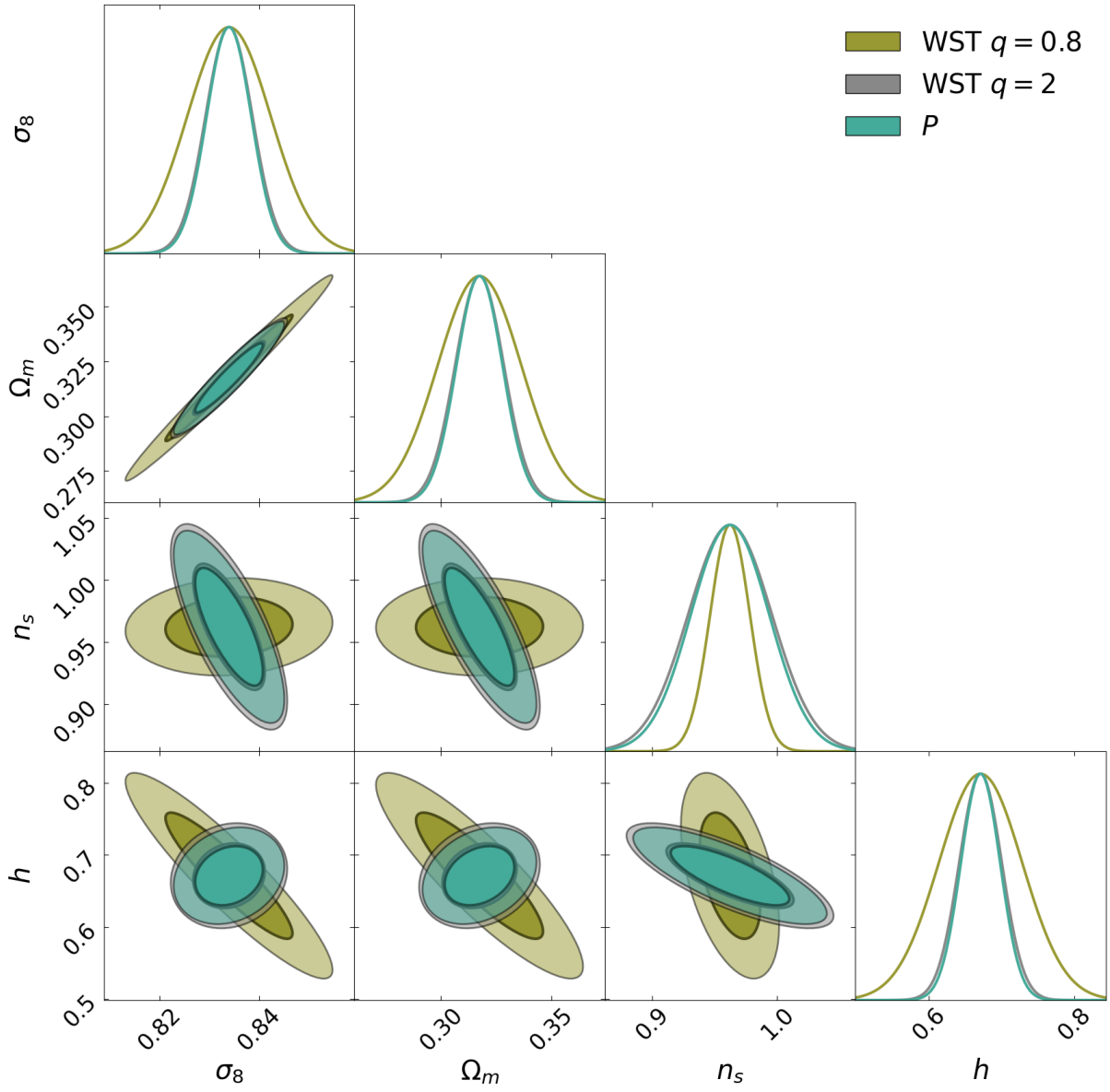

Figure 4.2: Expected distribution of the parameters around the fiducial QUIJOTE-PNG cosmology, with covariance given by the Fisher matrix computed on the non-linear matter field at $z = 1$. The statistics compared in the figure are 1st order WST coefficients $S_1^{0.8}(j, l)$ (olive), $S_1^2(j, l)$ (grey) and $P(k)$ (aquamarine).

4.5.2 Matter at $z = 1$

The analysis now turns to the matter field at $z = 1$ and extends the previous subsection by enlarging the parameter vector to include the PNG parameters, and by including second-order scattering coefficients and the bispectrum. Specifically, figures 4.3 and 4.4 use the parameter vector $\vec{\theta} = \{\sigma_8, \Omega_m, n_s, h, f_{\rm NL}^{\rm local}, f_{\rm NL}^{\rm equil}, f_{\rm NL}^{\rm ortho}\}$, while in the upper table of figure 4.7, each PNG shape is analyzed jointly with cosmological parameters and independently of the other two. For cosmological parameters, the largest bound of the three analyses is reported (differences are small).

In figures 4.3 and 4.4, constraints from the combination of all 76 WST coefficients are shown for $q = 0.8$ and $q = 2$, respectively, and compared with those from $P + B$ and from the full combination $P + B +$ WST. The corresponding 1σ Cramér–Rao bounds are listed in figure 4.7. Results for P alone are omitted, as they are outperformed by both WST variants on all parameters.

As discussed above, for $q = 2$ the extra information originates entirely from the second-order scattering coefficients. For $q = 0.8$, the inclusion of second-order coefficients improves performance, which becomes overall comparable to the $q = 2$ case: constraints are worse for all cosmological parameters except n_s, but slightly better for the PNG parameters. When combined with $P + B$, WST with $q = 0.8$ is slightly more effective than $q = 2$ at reducing parameter degeneracies.

4.5.3 Halos at $z = 0$

The Fisher matrix analysis is then applied to the non-linear dark matter halo field at $z = 0$. To test WST performance on a biased tracer in a more realistic scenario, the field is transformed to redshift space. The parameter set analyzed in figures 4.5 and 4.6 is $\vec{\theta} = \{\sigma_8, \Omega_m, n_s, h, M_{\rm min}, f_{\rm NL}^{\rm local}, f_{\rm NL}^{\rm equil}, f_{\rm NL}^{\rm ortho}\}$. Bounds from all 76 WST coefficients computed with $q = 0.8$ and $q = 2$ are shown and compared with those from the first two power spectrum multipoles (P_0+P_2), their combination with the bispectrum (P_0+P_2+B), and the combination of all statistics. The 1σ Cramér–Rao bounds for these cases are listed in the lower table of figure 4.7, where each PNG shape is analyzed independently of the others.

As for matter, WST outperforms the power spectrum alone for both values of q. This is particularly true for the PNG parameters, including $f_{\rm NL}^{\rm local}$ (constraints on the other two PNG shapes from P_0+P_2 are not reported in figure 4.7, as they are extremely weak). The performance of $P_0 + P_2 + B$ is comparable to WST for the cosmological parameters (and

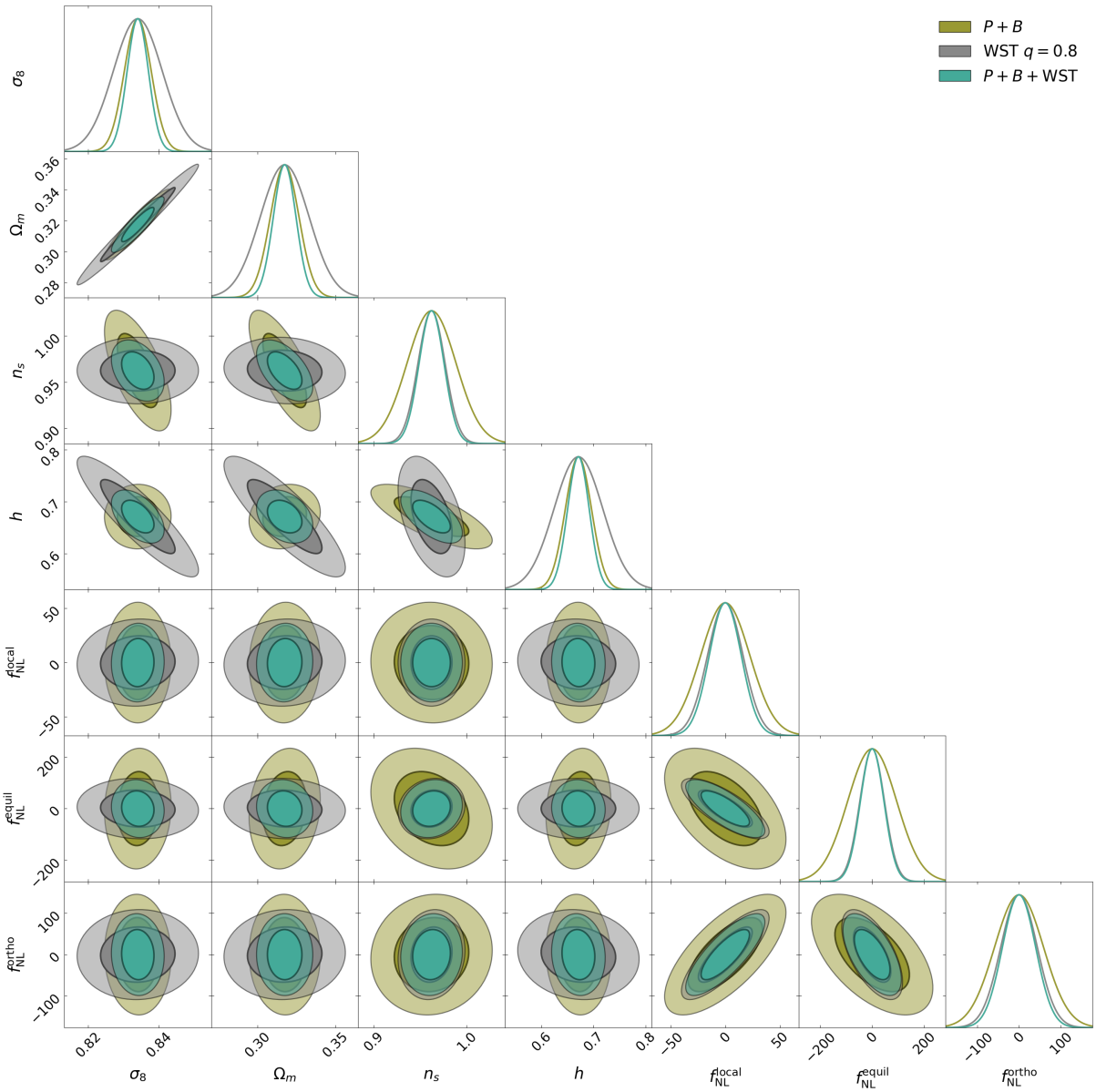

Figure 4.3: Expected distribution of the parameters around the fiducial QUIJOTE-PNG cosmology, with covariance given by the Fisher matrix computed on the non-linear matter field at $z = 1$. The statistics compared in the figure are $P + B$ (olive), WST coefficients $S_n^{0.8}(j, l)$ with $n = 0, 1, 2$ (grey), and the full combination $P + B +$ WST (aquamarine).

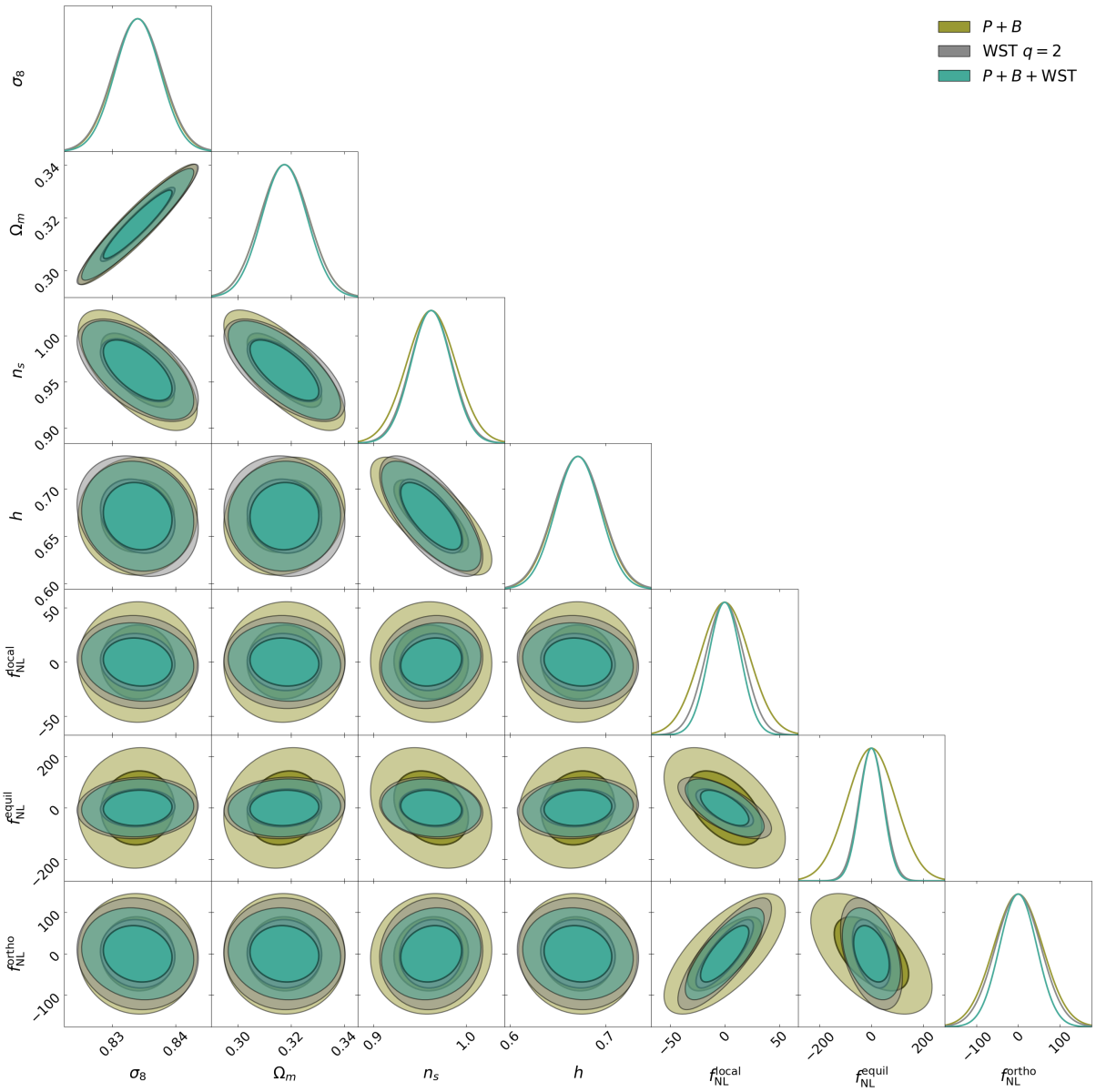

Figure 4.4: Expected distribution of the parameters around the fiducial QUIJOTE-PNG cosmology, with covariance given by the Fisher matrix computed on the non-linear matter field at $z = 1$. The statistics compared in the figure are $P + B$ (olive), WST coefficients $S_n^2(j, l)$ with $n = 0$, 1, 2 (grey), and the full combination $P + B +$ WST (aquamarine).

for $M_{\rm min}$), while for the PNG parameters a mixed outcome is obtained: $f_{\rm NL}^{\rm local}$ constraints are tighter (by $\approx$ 10–15%) for P_0+P_2+B, $f_{\rm NL}^{\rm equil}$ are tighter (by $\approx$ 20–30%) for WST, and $f_{\rm NL}^{\rm ortho}$ constraints are of similar quality for $q=2$ WST and about 30% better for $q=0.8$ WST.

Overall, the halo results confirm the matter findings: WST with $q=0.8$ and with $q=2$ yields comparable constraints across parameters. This indicates that once second-order coefficients are included, varying the exponent q does not significantly alter the total extracted information.

A direct comparison between the information content of WST and that of marked statistics for PNG and cosmological parameters is provided in appendix A.1, where very similar performance is observed.

4.6 Discussion

To identify where the information on PNG parameters resides within the WST, figures 4.8 and 4.9 display the sensitivity of WST coefficients to each parameter. For each parameter α, the absolute difference between the WST coefficients computed in the sets with α^+ and α^- is plotted, normalized to the fiducial cosmology (see Tab. 4.1 for the parameter shifts). figure 4.8 refers to dark matter at $z=1$ in real space, and figure 4.9 to halos at $z=0$ in redshift space. In each panel, the upper row contains PNG parameters and the lower row the remaining ones, while the left and right columns correspond to $q=0.8$ and $q=2$, respectively.

The WST coefficients from 0 to 75 are ordered in decreasing length scale of the outermost convolution. Thus, the leftmost coefficient is the second-order one with $j_2=4$, $j_1=3$, $l=0$, followed by $j_2=4$, $j_1=3$, $l=1$, and so on up to $j_2=4$, $j_1=0$, $l=4$ (end of the first light brown horizontal segment). These are followed by the first-order coefficients with $j=4$, $l=0,\cdots,4$ (dark brown segment), then by the second-order coefficients with $j_1<j_2=3$, etc.

No qualitative difference emerges between the left and right columns in each panel, indicating that the intuition developed in section 4.2.2 for $q=2$, namely, that first-order coefficients are related to the power spectrum, extends broadly to $q=0.8$. This explains why, for matter, first-order coefficients are almost insensitive to PNG parameters (figure 4.8, upper row, dark brown segments), especially at large scales: the $f_{\rm NL}$ dependence of the matter power spectrum is a suppressed one-loop effect and therefore manifests mainly at small scales. The same reasoning explains why second-order coeffi-

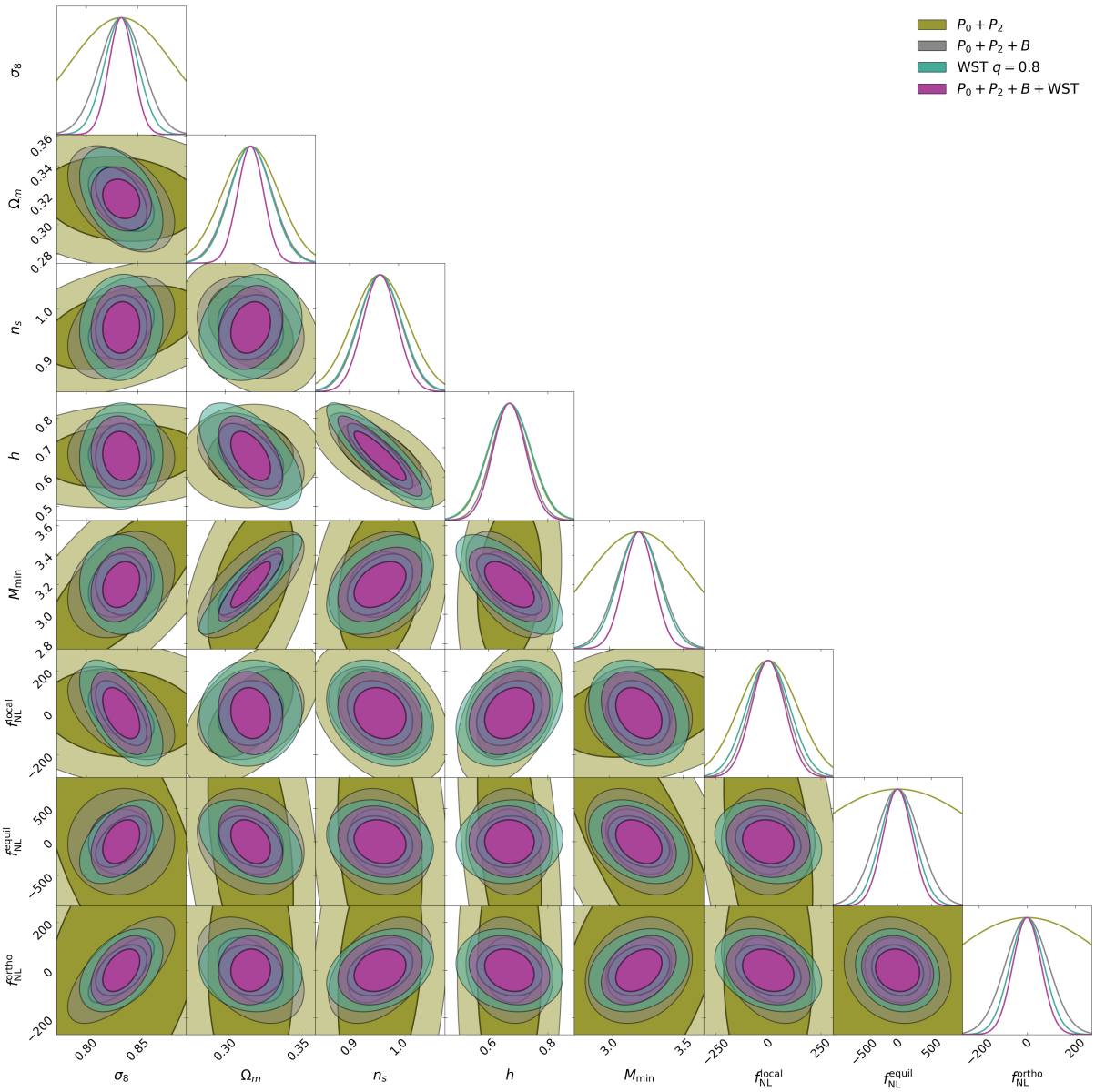

Figure 4.5: Expected distribution of the parameters around the fiducial QUIJOTE-PNG cosmology, with covariance given by the Fisher matrix computed on the non-linear halo field at $z = 0$. The statistics compared in the figure are the power spectrum monopole and quadrupole $P_0 + P_2$ (olive), $P_0 + P_2 + B$ (grey), WST coefficients $S_n^{0.8}(j, l)$ with $n = 0$, 1, 2 (aquamarine), and the full combination $P_0 + P_2 + B +$ WST (violet).

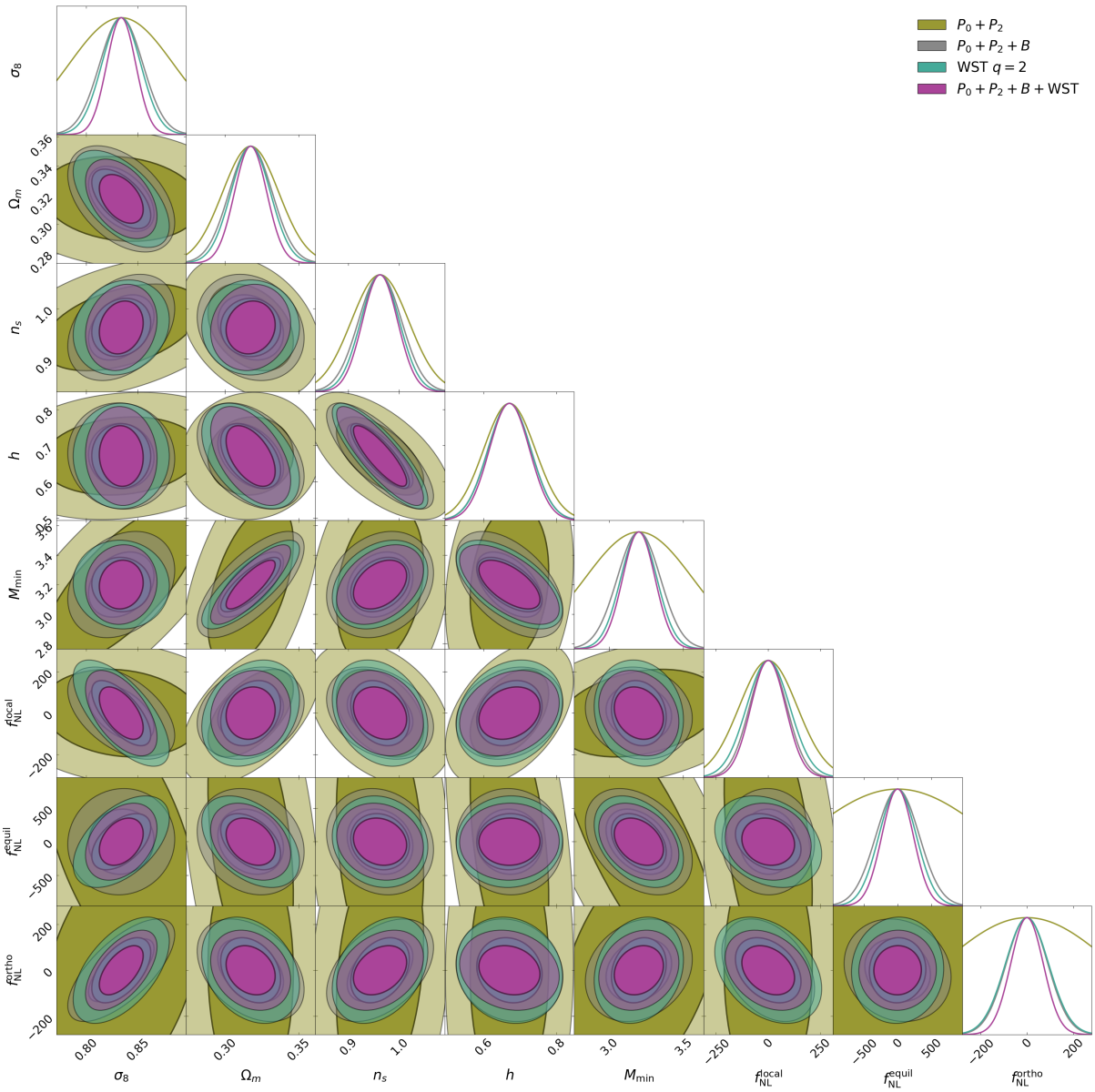

Figure 4.6: Expected distribution of the parameters around the fiducial QUIJOTE-PNG cosmology, with covariance given by the Fisher matrix computed on the non-linear halo field at $z = 0$. The statistics compared in the figure are the power spectrum monopole and quadrupole $P_0 + P_2$ (olive), $P_0 + P_2 + B$ (grey), WST coefficients $S_n^2(j, l)$ with $n = 0$, 1, 2 (aquamarine), and the full combination $P_0 + P_2 + B +$ WST (violet).

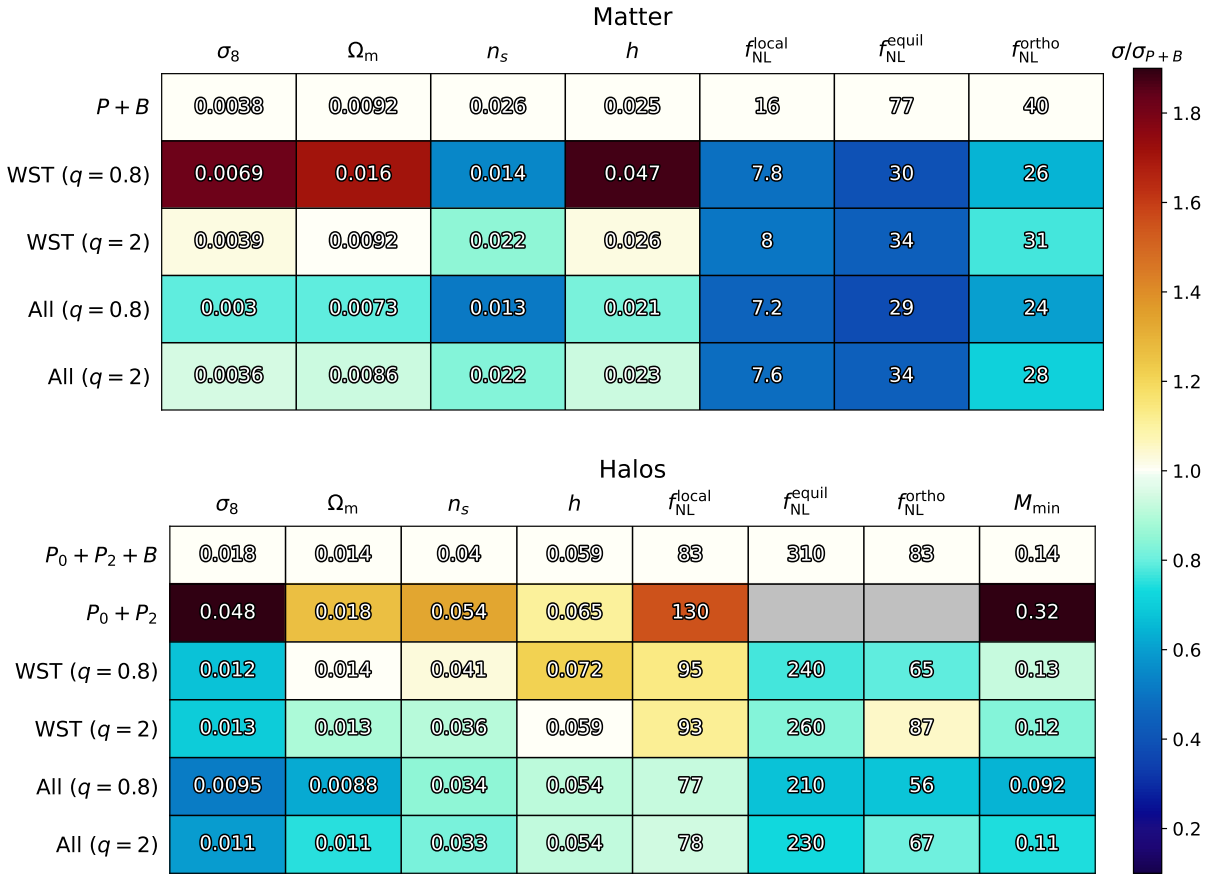

Figure 4.7: The Cramer-Rao 1σ bounds for different combinations of summary statistics (power spectrum, bispectrum and WST, "all" referring to the three of them together). Every statistic is measured up to $k_{\rm max} = 0.5\,h\,{\rm Mpc}^{-1}$ in the QUIJOTE N-body simulations at $z = 1$ (top panel), or in the corresponding halo catalogues at $z = 0$ (bottom panel), for a volume of $1\,(h^{-1}{\rm Gpc})^3$. The colour scale indicates the ratio of each error bar with respect to its power spectrum + bispectrum equivalent, highlighting the significant additional information captured by the WST in most cases. Note that each PNG shape is analyzed jointly with cosmological parameters and independently from the two others PNG shapes. For cosmological parameters, the largest bound of the three analyses is reported (differences being small anyway).

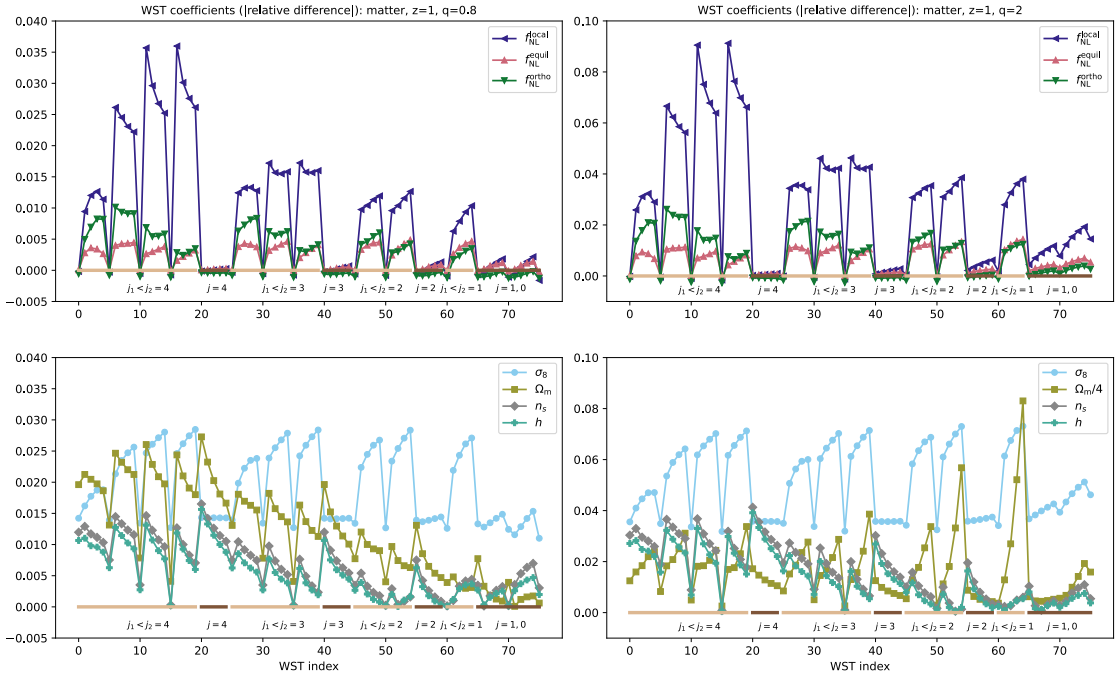

Figure 4.8: Variation of WST coefficients for matter in real space with respect to different parameters, normalized to the fiducial values, for the matter field at $z = 1$. Left: $q = 0.8$, Right: $q = 2$. Upper row: NG parameters; lower row: all other parameters. The green segments on the x-axis identify second order WST coefficients, whereas blue segments identify first order ones. There is essentially no NG signal in first order WST coefficients because the power spectrum of the matter field, to which such coefficients are sensitive, does not depend on f_{NL} at lowest order. See the main text for details.

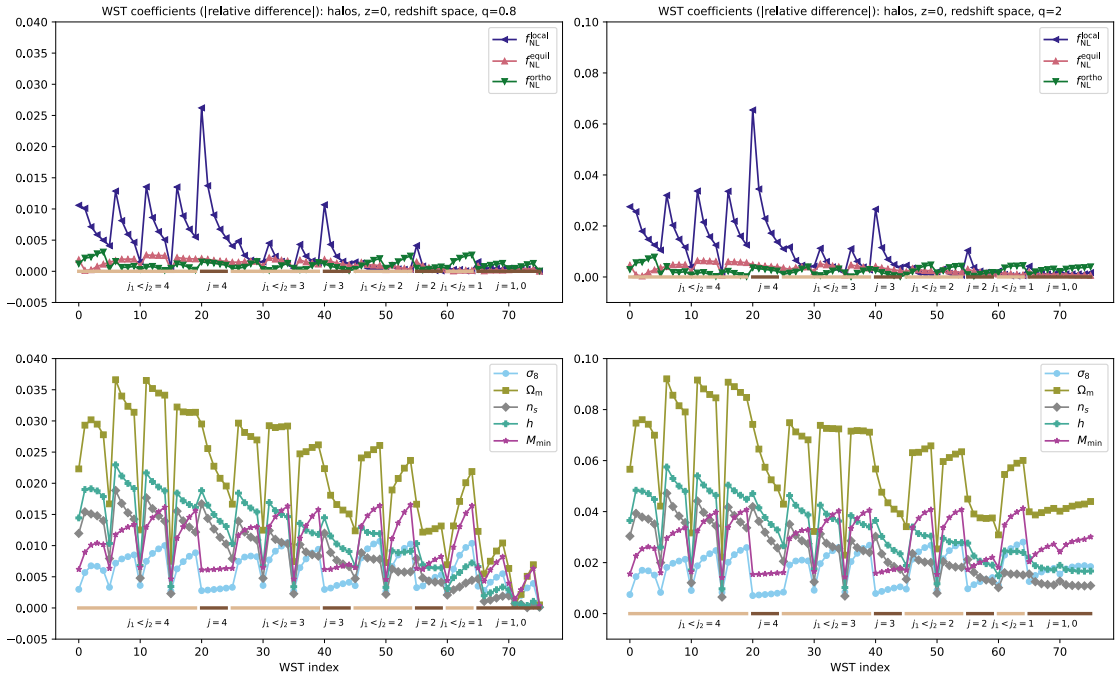

Figure 4.9: Variation of WST coefficients for the halo field in redshift space with respect to different parameters, normalized to the fiducial values. Now, first order WST coefficients pick up significant primordial NG signal for the local shape, due to the scale dependent bias signature in the power spectrum on large scales. See the main text for more explanations. Left: $q = 0.8$, Right: $q = 2$.

cients with $l = 0$ are almost insensitive to PNG parameters, since the sum in $U(j, l)(\vec{x})$ of equation 4.13 includes a single term, making the second-order coefficient also tied to the power spectrum. Consequently, for matter, PNG information primarily enters through second-order coefficients with $l \neq 0$. By contrast, for the other parameters, information is already present in the linear-order power spectrum; correspondingly, the signal is visible in first-order WST coefficients, as seen in the second row of figure 4.8.

It is also important to note that information content depends on both derivatives and covariances of the WST coefficients. The first row of figure 4.10 shows the correlation coefficients

$$r_{ij} \equiv \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, \tag{4.21}$$

where C_{ij} is the covariance between the i-th and j-th coefficients. r_{ij} ranges from -1 (complete anticorrelation) to 1 (complete correlation), with 0 indicating no correlation. The correlation patterns differ markedly between $q = 0.8$ and $q = 2$. For $q = 2$, the correlation among first-order coefficients is very small at large scales and increases at smaller ones, reflecting the behavior of power spectrum bins. First-order coefficients are also correlated with second-order ones for $l = 0$, reinforcing the interpretation that the latter are connected to the power spectrum. A striking feature is the strong correlation

among second-order coefficients with $l \neq 0$, together with very small correlation between these and first-order coefficients, especially at large scales. This separation is absent for $q = 0.8$, consistent with the conclusion (see section 4.5.1) that for $q = 2$ the PS information is confined to first order while for $q \neq 2$ it is distributed across orders. The strong correlation among second-order coefficients for $q = 2$ suggests that a reduced subset may suffice to capture the same information.

For halos, the differences between PNG sensitivities of first- and second-order coefficients, and between correlation patterns for $q = 2$ and $q = 0.8$, largely disappear for the local shape. In this case, first-order WST coefficients are sensitive to $f_{\rm NL}$ because the power spectrum exhibits scale-dependent bias on large scales [144]. This can be understood via the contribution to the density field

$$\delta_h = b_1 \delta_m + b_\phi f_{\rm NL}^{\rm local} \phi + \epsilon \,, \tag{4.22}$$

where b_1 and b_ϕ are tracer-dependent bias parameters (for matter, $b_1 = 1$ and $b_\phi = 0$), ϕ is the gravitational potential, and ϵ is a stochastic term, uncorrelated with δ_m and ϕ, that describes shot noise in the halo field. The latter is responsible for the reduced sensitivity compared to matter, especially at small scales, and for the decorrelation between WST coefficients in the $q = 2$ case seen in figure 4.10.

4.7 Conclusions

This work investigated the WST as a tool for PNG studies and compared its performance with that of a P+B analysis. The matter and halo fields, in real and redshift space, were analyzed from the publicly available Quijote and QuijotePNG N-body simulation suites, producing Fisher forecasts for the amplitude parameters $f_{\rm NL}^{\rm local}$, $f_{\rm NL}^{\rm equil}$, and $f_{\rm NL}^{\rm ortho}$, along with standard cosmological parameters.

For a fair comparison, WST and P+B analyses were implemented on the same fields. In particular, a sharp cut at $k_{\rm max} = 0.5 \;\; h\,{\rm Mpc}^{-1}$ was applied to minimize possible small-scale artifacts in the WST. Two values of the exponent q were compared, namely $q = 0.8$ and $q = 2$, with the latter allowing a clearer interpretation: first-order scattering coefficients are equivalent to the power spectrum (figure 4.2), so all additional information arises from second order.

WST (for both $q = 0.8$ and $q = 2$) outperforms the power spectrum alone for PNG and cosmological parameters, as well as for the nuisance parameter $M_{\rm min}$. For $f_{\rm NL}^{\rm local}$ in

halos, the improvement is about 27%. When combined with B, halo constraints from WST are weaker for $f_{\rm NL}^{\rm local}$ (by $\approx 15\%$), but stronger for $f_{\rm NL}^{\rm equil}$ ($\approx 25\%$) and for $f_{\rm NL}^{\rm ortho}$ ($\approx 28\%$ for $q = 0.8$).

Results for the standard (non-PNG) cosmological parameters align with previous analyses [30, 31], which also find WST to outperform the power spectrum alone. For PNG parameters, the main focus of this work, the WST enhances the extraction of PNG information from LSS data relative to a standard P+B analysis. For QUIJOTEPNG simulations, the level of improvement is competitive with the best results obtained using other summary statistics, such as the marked power spectrum and bispectrum [145].

Future work should move in two directions. On the data side, realism should be increased to match upcoming surveys, including modeling of galaxy bias, survey systematics, and related effects. On the theory side, improved interpretation of WST coefficients is desirable. The $q = 2$ choice provides a clean theoretical description with essentially no information loss relative to other q values; extending to more general families of wavelets tailored to PNG extraction from LSS, for example by exploiting angular structure to mitigate redshift-space distortions and leveraging underlying symmetries such as extended Galilean invariance [51, 93, 94, 146–148], may reduce the number of coefficients required in analyses.

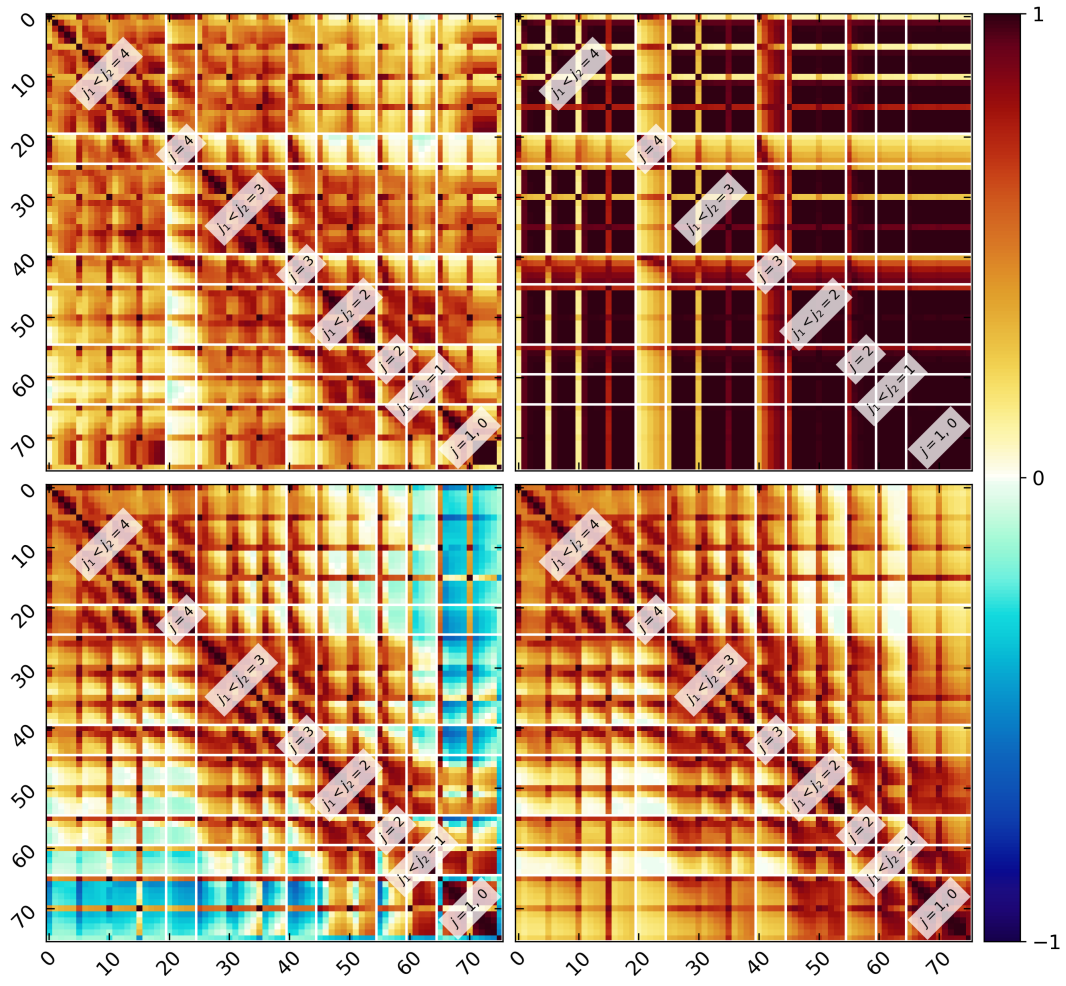

Figure 4.10: The correlation matrix of the WST coefficients. Upper row: matter, lower row: halos. Left: $q = 0.8$, Right: $q = 2$.

Chapter 5

Field-level Inference with the Bootstrap of LSS

The contents of this chapter have been published in the paper

Matteo Peron, Takahiro Nishimichi, Massimo Pietroni and Atsushi Taruya (October 2025). "Renormalized perturbation theory at field-level: the LSS bootstrap in Grid-SPT." *Journal of Cosmology and Astroparticle Physics* **2025**(10): 098. doi:10.1088/1475-7516/2025/10/098.[1]

which has been included, integrally or in part, according to the journal's copyright policy.

5.1 Introduction

The large-scale structure (LSS) of the Universe offers a stringent testing ground for fundamental physics. To fully exploit data from surveys such as DESI [1] and Euclid [2, 3], it is increasingly important to develop methods that optimize the extraction of cosmological information from LSS, particularly in ways sensitive to physics beyond the standard ΛCDM model. Various scenarios of "new physics" have been proposed, yet there is currently no theoretical rationale to single out one over the others. Moreover, in most plausible cases the predicted signatures are subtle, underscoring the need for a model-independent framework able to identify small anomalies. To this end, the "LSS bootstrap" approach discussed in section 2.3.3 was formulated in [51] and further de-

[1]The contents of the article are accessible at `https://www.doi.org/10.1088/1475-7516/2025/10/098`.

veloped in [96, 149]. This symmetry-based program assumes no specific form for the Lagrangian or the equations of motion. Instead, at each order in perturbation theory (PT), the most general form of the dark matter (DM) density and velocity fields, as well as of biased tracers (galaxies, DM halos, ...), compatible with the assumed symmetries, is expressed in terms of the linear field and a finite set of time-dependent coefficients. Detecting deviations of these coefficients from their ΛCDM values would signal new physics.

While state-of-the-art analyses often concentrate on low-order correlators such as the power spectrum [150, 151] and the bispectrum [5, 7], recent work has explored novel summary statistics to access non-Gaussian information encoded in higher-order correlators [11, 12, 15, 30, 31, 34, 109, 152, 153] (see also [154] for a review). A natural next step is to perform inference directly at the field level, thereby retaining the full, uncompressed information content. This approach, developed in [35–47] and further investigated in [48, 49, 155–158], provides an ideal setting for probing both linear and nonlinear imprints of new physics.

This work pursues the construction of a field-level perturbative framework for model-independent cosmological inference, emphasizing precision constraints on departures from ΛCDM.

To this end, the `GridSPT` code [159–161], an implementation of Eulerian PT on discretized grids (see section 2.3 and [85] for a review), is extended to include the LSS bootstrap parametrization, enabling a systematic exploration of nonlinear deviations from the standard model.

Working at the field level on a grid necessarily introduces an ultraviolet (UV) cutoff. The linear fields that serve as building blocks of higher PT orders contain wavenumbers only up to the grid's Nyquist frequency or, as discussed below, up to even lower values when avoiding spurious effects such as *aliasing*, see appendix B.1. A meticulous treatment of this artificial cutoff dependence is thus essential to obtain physically meaningful, cutoff-independent results.

A Wilsonian perspective is adopted (for a similar philosophy, see [162] and, more recently, [163, 164], which also include biased tracers and non-Gaussian initial conditions): the starting point is the definition of the perturbative model at a high momentum cutoff scale $\Lambda_{\rm uv}$. The model is built from the linear field coarse-grained at $\Lambda_{\rm uv}$, and consists of a truncated PT series plus a set of "counterterms". As in the Effective Field Theory of LSS (EFTofLSS) [99–101], the counterterms' form is fixed by the symmetries of the system (rotational invariance, Extended Galilean Invariance (EGI), and, for the matter and velocity fields, mass and momentum conservation). Their number is dictated by the

adopted order in both the PT expansion and the derivative expansion. The counterterm coefficients encode physics beyond the cutoff, i.e., from scales $q > \Lambda_{\rm uv}$ not described by the model, and are therefore not computable; their values must be fixed by *renormalization conditions*, via comparison between theory and measurements. By contrast, the *running* of these parameters, i.e., their evolution as the cutoff is lowered to $\Lambda < \Lambda_{\rm uv}$, is computable in PT. Hence, the requirement that physical observables be cutoff-independent links the UV model to any model defined at a lower scale Λ. This renormalization is carried out explicitly up to fifth order in perturbation theory, leading to properly renormalized field-level models suitable for determining the bootstrap parameters. The procedure not only ensures theoretical consistency, but also enables efficient computations on smaller grids.

In particular, the crucial role of higher-derivative terms in precision inference is highlighted. These terms, which naturally emerge in the renormalization procedure, can be computed analytically within the present framework. Their inclusion is essential to capture small-scale physics accurately and to ensure a robust determination of the bootstrap parameters required for model-independent cosmological tests.

The analysis is restricted to the case in which the bootstrap parameters are scale independent. This choice excludes models with strongly scale-dependent growth, such as those with extra scalar degrees of freedom whose Compton wavelength is comparable to LSS scales. This restriction does not apply to cosmologies with massive neutrinos in the phenomenologically allowed mass range, where all relevant scale-dependent effects can be accurately absorbed into the linear power spectrum (see, e.g., [165–168]).

The paper is organized as follows. section 5.2 presents the bootstrap PT kernel at second order, in both Fourier and configuration space. section 5.3 reviews `GridSPT` and describes the modifications implemented to compute the bootstrap kernels. section 5.4 discusses the need for a momentum cutoff when formulating field theory on a grid and shows in detail how models defined at different cutoffs are related (running). section 5.5 compares the PT models with N-body simulations and explains how to extract the bootstrap parameters. section 5.6 presents the main results and verifies the proper renormalization of the models. Finally, section 5.7 discusses the implications and conclusions. Appendix B.2 comments on the impact of the bootstrap parameter when included also in the third-order PT contribution, and Appendix B.3 provides details on the implementation of artificial noise in the dark matter simulations.

5.2 Field level bootstrap

This section reviews the perturbative model implemented in the analysis, based on the "LSS bootstrap" approach (see [51, 96, 169] for details).

5.2.1 First order

Perturbations are expressed in terms of the density contrast $\delta(\vec{x},\, a)$ and the rescaled velocity divergence,

$$\theta(\vec{x},\, a) \equiv \partial_i u^i(\vec{x},\, a)\,, \tag{5.1}$$

with

$$u^i(\vec{x},\, a) \equiv -\frac{v^i(\vec{x},\, a)}{faH(a)}\,, \tag{5.2}$$

where $v^i(\vec{x}, a)$ is the peculiar velocity, $f = \mathrm{d}\log D_+/\mathrm{d}\log a$ is the growth rate, and $D_+(a)$ is the linear growth factor, assumed to be scale independent. The fields are then expanded according to the standard perturbative Ansatz [85]:

$$\delta^{[N],\mathrm{PT}}(\vec{x},\, a) = \sum_{n=1}^{N} \delta^{(n)}(\vec{x},\, a)\,, \quad \theta^{[N],\mathrm{PT}}(\vec{x},\, a) = \sum_{n=1}^{N} \theta^{(n)}(\vec{x},\, a)\,. \tag{5.3}$$

In what follows, lowercase parentheses (e.g., $\delta^{(n)}$) denote the n-th order contribution, while square brackets (e.g., $\delta^{[N],\mathrm{PT}}$) indicate truncation at order N, i.e., the sum of all m-th order terms with $1 \leq m \leq N$.

The first-order solutions on the growing mode read:

$$\delta^{(1)}(\vec{x},\, a) = \theta^{(1)}(\vec{x},\, a) \equiv \varphi(\vec{x},\, a)\,, \tag{5.4}$$

with $\varphi(\vec{x},\, a) = D_+(a)\varphi_0(\vec{x})$ the linear density field, a Gaussian realization whose variance is given by the linear power spectrum $P(k)$ of the chosen cosmology.

5.2.2 Second order

In Fourier space, the second-order fields can be written as convolutions:

$$\delta^{(2)}(\vec{k},\, a) = \mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2} F_2(\vec{q}_1, \vec{q}_2; a)\varphi(\vec{q}_1,\, a)\varphi(\vec{q}_2,\, a)\,, \tag{5.5}$$

$$\theta^{(2)}(\vec{k},\, a) = \mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2} G_2(\vec{q}_1, \vec{q}_2; a)\varphi(\vec{q}_1,\, a)\varphi(\vec{q}_2,\, a)\,, \tag{5.6}$$

where

$$\mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n} \equiv \int \frac{d^3q_1 \cdots d^3q_n}{(2\pi)^{3(n-1)}} \delta_D(\vec{k} - \vec{q}_1 \cdots - \vec{q}_n)\,, \tag{5.7}$$

and the convolution kernels are

$$F_2(\vec{q}_1, \vec{q}_2; a) = \beta(\vec{q}_1, \vec{q}_2) + \frac{a_\gamma^{(2)}(a)}{2}\,\gamma(\vec{q}_1, \vec{q}_2)\,, \tag{5.8}$$

$$G_2(\vec{q}_1, \vec{q}_2;\, a) = \beta(\vec{q}_1, \vec{q}_2) + \frac{d_\gamma^{(2)}(a)}{2}\,\gamma(\vec{q}_1, \vec{q}_2)\,, \tag{5.9}$$

with mode-coupling functions

$$\beta(\vec{q}_1, \vec{q}_2) = \frac{|\vec{q}_1 + \vec{q}_2|^2 \vec{q}_1 \cdot \vec{q}_2}{2q_1^2 q_2^2}\,, \tag{5.10}$$

$$\gamma(\vec{q}_1, \vec{q}_2) = 1 - \frac{(\vec{q}_1 \cdot \vec{q}_2)^2}{q_1^2 q_2^2}\,. \tag{5.11}$$

The coefficients of the $\beta(\vec{q}_1, \vec{q}_2)$ couplings in equations (5.8) and (5.9) are fixed to unity by EGI, while the bootstrap coefficients $a_\gamma^{(2)}$ and $d_\gamma^{(2)}$ are left unconstrained by symmetries and, in general, are time dependent.

As an example, in ΛCDM the equations for the second-order bootstrap coefficients read [51]:

$$\frac{d\,a_\gamma^{(2)}(a)}{d\ln a} = f\left(2 - 2a_\gamma^{(2)}(a) + d_\gamma^{(2)}(a)\right)\,, \tag{5.12}$$
$$\frac{d\,d_\gamma^{(2)}(a)}{d\ln a} = f\left(\frac{3}{2}\frac{\Omega_m(a)}{f^2}\left(a_\gamma^{(2)}(a) - d_\gamma^{(2)}(a)\right) - d_\gamma^{(2)}(a)\right)\,.$$

Deviations from these solutions parameterize departures from ΛCDM in a model-independent manner. A standard PT practice is to adopt the Einstein–de Sitter (EdS) limit,[2] obtained by setting $\Omega_m/f^2 = 1$ in the above equations, which yields the time-independent values [85]

$$a^{(2)}_{\gamma,\,\mathrm{EdS}} = \frac{10}{7}\,, \quad d^{(2)}_{\gamma,\,\mathrm{EdS}} = \frac{6}{7}\,. \tag{5.13}$$

These values serve as initial conditions for models that reduce to EdS at early times.

The bootstrap approach extends beyond second order with an increasing number of cosmology-dependent coefficients. Explicit expressions up to third order in the Eulerian formalism are given in Appendix B.2 and in [51, 169], and up to sixth order in the equivalent Lagrangian formalism in [96]. The program also extends to biased tracers' density contrast; in that case, mass and momentum conservation need not hold and the kernels contain more free parameters, corresponding to those in the perturbative bias expansion (see, e.g., [95]). The present analysis is restricted to the matter density field in real space. Moreover, all coefficients of kernels of order higher than second are set to their EdS values, leaving only the second-order bootstrap coefficient $a_\gamma^{(2)}$ free to vary.

Any detection of a deviation of a bootstrap parameter from its ΛCDM value would signal new physics. In realistic scenarios, several bootstrap parameters would generally deviate, implying a larger signal than the conservative setup considered here.

In configuration space, the second-order density field is given by

$$\delta^{(2)}(\vec{x}) = \varphi_\beta^{(2)}(\vec{x}) + \frac{a_\gamma^{(2)}}{2}\varphi_\gamma^{(2)}(\vec{x})\,, \tag{5.14}$$

where time arguments are omitted and the second-order fields are

[2] For analyses beyond the EdS approximation, see [169–171].

$$\varphi_\beta^{(2)}(\vec{x}) \equiv \partial_i\varphi(\vec{x}) \left(\frac{\partial_i}{\partial^2}\varphi(\vec{x})\right) + \left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right)\left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right), \tag{5.15}$$

$$\varphi_\gamma^{(2)}(\vec{x}) \equiv (\varphi(\vec{x}))^2 - \left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right)\left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right). \tag{5.16}$$

5.3 GridSPT and bootstrap

5.3.1 The GridSPT code

The numerical implementation used here is based on the grid-based Eulerian PT code GridSPT [159–161]. The code performs standard PT calculations at the field level, generating numerical realizations of higher-order density and velocity fields at each grid point. Notably, by applying a novel nEPT (nth-order Eulerian PT) scheme, GridSPT delivers improved predictions for summary statistics, surpassing those from conventional order-by-order PT calculations [172, 173].

The core of GridSPT is a real-space recursion under the EdS approximation. For $n \geq 2$, the n-th order Eulerian density and velocity fields are constructed in configuration space from lower-order counterparts as [161]:

$$\begin{pmatrix}\delta^{(n)}\\ \theta^{(n)}\end{pmatrix} = \frac{2}{(2n+3)(n-1)} \begin{pmatrix} n+\frac{1}{2} & 1 \\ \frac{3}{2} & n\end{pmatrix} \sum_{m=1}^{n-1} \begin{pmatrix} \partial_i\delta^{(m)}\frac{\partial_i}{\partial^2}\theta^{(n-m)} + \delta^{(m)}\theta^{(n-m)} \\ \frac{1}{2}\partial^2\left(\frac{\partial_i}{\partial^2}\theta^{(m)}\frac{\partial_i}{\partial^2}\theta^{(n-m)}\right)\end{pmatrix}. \tag{5.17}$$

Given the linear fields $\delta^{(1)}(\vec{x})$ and $\theta^{(1)}(\vec{x})$ in equation (5.4), the algorithm computes the nonlinear source terms on the right-hand side of equation (5.17) using Fast Fourier Transforms. The differential operators in the sources are implemented via

$$\frac{\partial_i}{\partial^2}\varphi(\vec{x}) = -i\int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}}\,\frac{k_i}{k^2}\varphi(\vec{k}), \tag{5.18}$$

$$\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x}) = -\int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}}\,\frac{k_ik_j}{k^2}\varphi(\vec{k}). \tag{5.19}$$

5.3.2 Bootstrapping `GridSPT`

The `GridSPT` fields obtained from equation (5.17) rely on the EdS approximation for the PT kernels. To compute the $n = 2$ bootstrap density field, equation (5.5) becomes

$$\delta^{(2)}(\vec{x}) = \delta^{(2)}_{\rm EdS}(\vec{x}) + \frac{1}{2}\varepsilon_\gamma \, a^{(2)}_{\gamma,\rm EdS} \, \varphi^{(2)}_\gamma(\vec{x}) \,, \tag{5.20}$$

where the parameter

$$\varepsilon_\gamma \equiv \frac{a^{(2)}_\gamma}{a^{(2)}_{\gamma,\rm EdS}} - 1 \tag{5.21}$$

quantifies the relative deviation from the EdS limit. At $z = 1$, for the ΛCDM model considered here, $\varepsilon_{\gamma,\Lambda\rm CDM} \simeq -7.8 \cdot 10^{-4}$.

The field $\delta^{(2)}_{\rm EdS}(\vec{x})$ is taken from `GridSPT`, while $\varphi^{(2)}_\gamma$, defined in equation (5.16), is built from the same linear field $\varphi(\vec{x})$ used for $\delta^{(2)}_{\rm EdS}(\vec{x})$. To reduce the tensorial complexity, equation (5.16) is rewritten as

$$\varphi^{(2)}_\gamma(\vec{x}) = (\varphi(\vec{x}))^2 - \frac{1}{2}\partial^2 \left[\left(\frac{\partial_i}{\partial^2}\varphi(\vec{x}) \right) \left(\frac{\partial_i}{\partial^2}\varphi(\vec{x}) \right) \right] + \left(\frac{\partial_i}{\partial^2}\varphi(\vec{x}) \right) \partial_i \varphi(\vec{x}) \,, \tag{5.22}$$

for an irrotational velocity field. Since the analysis focuses on the parameter $a^{(2)}_\gamma$, all other bootstrap parameters are set to their EdS values; hence, for $n > 2$,

$$\delta^{(n)}(\vec{x}) = \delta^{(n)}_{\rm EdS}(\vec{x}) \,. \tag{5.23}$$

Strictly speaking, this is not fully consistent, as the relations in [51] imply that higher-order PT kernels also depend on $a^{(2)}_\gamma$. An explicit check, however, shows that this dependence is subdominant; including it in the third-order field produces negligible changes (Appendix B.2).

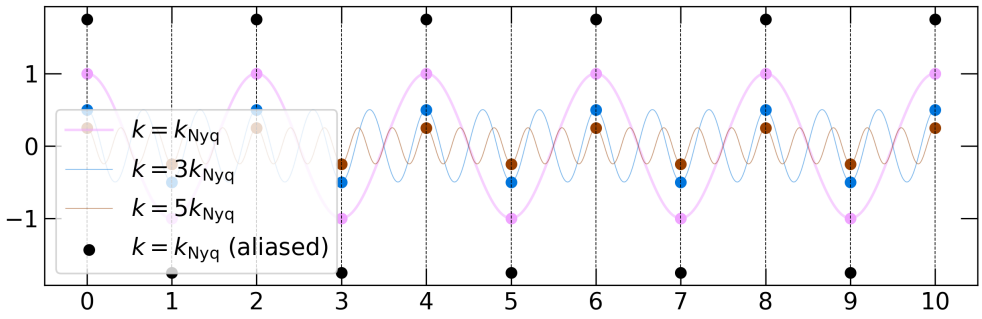

Figure 5.1: Simple illustration of how aliasing works. A detector samples a signal at each unit interval in a domain of length $L = 10$, so $N_g = 10$, meaning that the Nyquist frequency is $k_{\rm Nyq} = \frac{\pi N_g}{L} = \pi$. A wavemode at the Nyquist frequency (pink) is correctly sampled, however frequencies at $k = (2n+1)k_{\rm Nyq}$ (e.g., blue and brown) cannot be resolved and appear to the detector as $k = k_{\rm Nyq}$ as well. The result is that the measured signal at $k_{\rm Nyq}$ is the sum of all the contributions for $n > 0$ (black).

5.4 Field on the grid: regularization and renormalization

5.4.1 The need for a momentum cutoff

As discussed in [161], perturbative computations on a grid require care to avoid spurious "aliasing" effects in the resulting fields. These arise from products of linear fields at the same grid point or, in field-theoretic language, from *composite operators*. On a finite cubic grid of volume L^3 and N_g^3 points, a mode $\vec{k}$ is indistinguishable from $\vec{k} + 2\,\vec{n}\,k_{\rm Nyq}$, where $k_{\rm Nyq} = \pi N_g/L$ is the Nyquist frequency and $\vec{n} = (n_x, n_y, n_z)$ with $n_{x,y,z} \in \mathbb{Z}$. Modes with $\vec{n} \neq \vec{0}$, not supported by the grid, appear at lower (aliased) modes (figure 5.1). In Fourier space, N-th order PT fields are given by convolutions involving up to N linear fields at all supported modes. To ensure that the convolution momenta do not exceed the Nyquist limit, a low-pass filter is applied to the linear fields at an N-dependent scale Λ, given by the Orszag rule [161, 174]:

$$k \leq \Lambda = \frac{2}{N+1} k_{\rm Nyq} = \frac{2}{N+1} \frac{\pi N_g}{L} \,. \tag{5.24}$$

An isotropic top-hat filter in Fourier space is adopted, and the Λ-dependence of the fields is indicated explicitly. For the linear fields, $\varphi_\Lambda(\vec{x})$ denotes the Fourier transform of

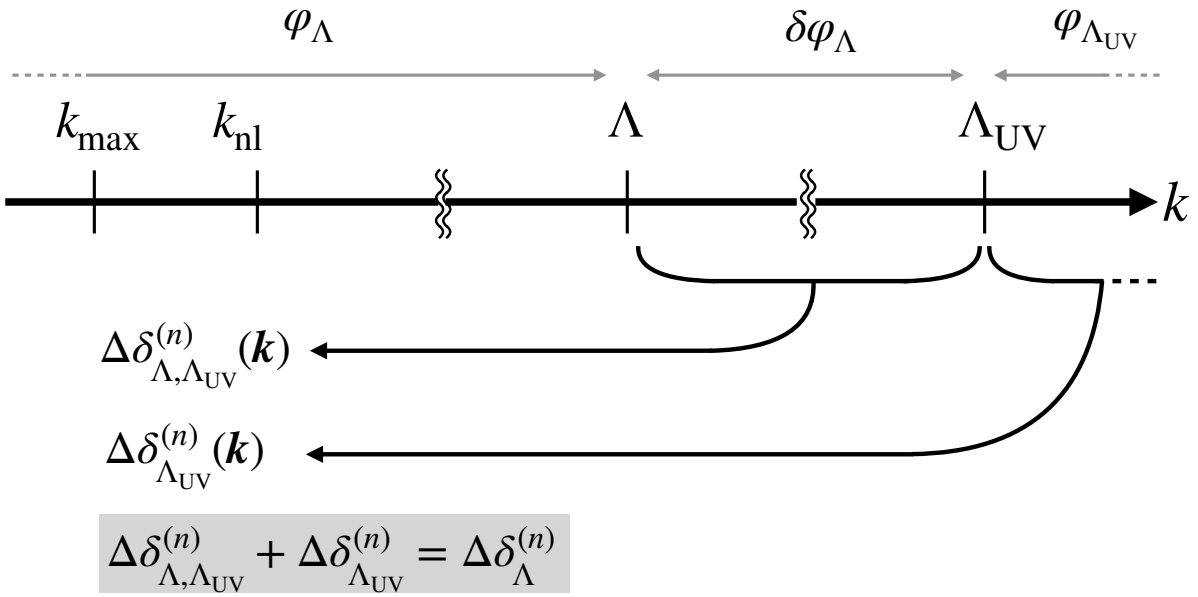

Figure 5.2: Schematic illustration of the relevant scales and fields defined in the paper. The theory is defined at a scale $\Lambda_{\rm uv} \gg k_{\rm nl} \gg k_{\rm max}$, as the sum of the PT contribution, equation (5.28), and the UV counterterms, equation (5.30). By considering a scale $\Lambda < \Lambda_{\rm uv}$, one can define a new theory, again as the sum of a PT part (expressed in terms of the new IR fields φ_Λ) and UV counterterms. The latter are given by the sum of the UV counterterms of the original theory and the new contributions of equation (5.36), obtained by integrating out the $\delta\varphi_\Lambda$ fields.

$$\varphi_\Lambda(\vec{k}) = \varphi(\vec{k})\,\Theta\left(\Lambda - k\right)\,, \tag{5.25}$$

while, for $n > 1$, $\delta^{(n)}_\Lambda(\vec{x})$ and $\theta^{(n)}_\Lambda(\vec{x})$ denote n-th order fields built from n filtered linear fields $\varphi_\Lambda(\vec{x})$.

Even ignoring aliasing, the grid spacing ultimately sets a cutoff (the Nyquist frequency). Thus, in field-theoretic terms, the theory on the grid is unavoidably *regularized* by an unphysical cutoff; the cutoff dependence must be handled carefully. The task is to define Λ-independent quantities from Λ-dependent fields and couplings, and to relate them to observables, or in other words, to *renormalize* the theory. Practically, controlling the cutoff dependence allows the use of smaller cutoffs, and hence smaller grids, improving computational efficiency.

5.4.2 Integrating out UV modes

In a Wilsonian picture [175, 176], the field theory is first defined at a UV scale $\Lambda_{\rm uv}$ and then related to theories at lower cutoffs Λ by integrating out fluctuations with wavenumbers between Λ and $\Lambda_{\rm uv}$. The following hierarchy is assumed (figure 5.2):

$$k_{\rm max} \ll k_{\rm nl} \ll \Lambda_{\rm uv}\,. \tag{5.26}$$

A distinction is crucial between the observed external wavenumber $\vec{k}$ and the internal wavenumbers $\vec{q}_i$ coupled to it. While one restricts to $k < k_{\rm nl}$, the internal $\vec{q}_i$ can be larger than $k_{\rm nl}$; since PT fails at $q > k_{\rm nl}$, UV counterterms must correct the faulty PT couplings to the mode $\vec{k}$.

The UV theory follows EFTofLSS [99–101]. Consider PT models truncated at order N; in Fourier space and omitting time dependence,

$$\delta^{[N]}(\vec{k}) = \delta^{[N],{\rm PT}}_{\Lambda_{\rm uv}}(\vec{k}) + \Delta\delta^{[N]}_{\Lambda_{\rm uv}}(\vec{k})\,, \tag{5.27}$$

with the truncated PT series

$$\delta^{[N],{\rm PT}}_{\Lambda_{\rm uv}}(\vec{k}) \equiv \sum_{n=1}^{N} \delta^{(n)}_{\Lambda_{\rm uv}}(\vec{k})\,, \tag{5.28}$$

and n-th order term

$$\delta^{(n)}_{\Lambda_{\rm uv}}(\vec{k}) = \mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n} F_n(\vec{q}_1,\cdots,\vec{q}_n)\varphi_{\Lambda_{\rm uv}}(\vec{q}_1)\cdots\varphi_{\Lambda_{\rm uv}}(\vec{q}_n)\,. \tag{5.29}$$

Here F_n are the symmetrized PT kernels [85] (e.g., equation (5.5) at $n = 2$), and $\varphi_{\Lambda_{\rm uv}}$ has support only for $q \leq \Lambda_{\rm uv}$ (equation (5.25)). The second contribution sums the UV *counterterms* up to $n = N$:

$$\Delta\delta^{[N]}_{\Lambda_{\rm uv}}(\vec{k}) \equiv \sum_{n=1}^{N} \Delta\delta^{(n)}_{\Lambda_{\rm uv}}(\vec{k})\,. \tag{5.30}$$

These counterterms encode the effect of modes with $q > \Lambda_{\rm uv}$ on the observed scales, so that the PT expansion converges to the truth (e.g., a high-resolution N-body result or data), up to nonperturbative corrections $O(k^2/k^2_{\rm nl})$, which are coefficients to be fixed by renormalization conditions (section 5.4.5). Given the hierarchy (5.26), residual $O(k^2/\Lambda^2_{\rm uv})$

effects are negligible.

Lower a cutoff $\Lambda < \Lambda_{\rm uv}$ and enforce cutoff independence:

$$\delta_\Lambda^{[N],\rm PT}(\vec{k}) + \Delta\delta_\Lambda^{[N]}(\vec{k}) = \delta_{\Lambda_{\rm uv}}^{[N],\rm PT}(\vec{k}) + \Delta\delta_{\Lambda_{\rm uv}}^{[N]}(\vec{k}) = \delta^{[N]}(\vec{k})\,. \tag{5.31}$$

Split the UV field as

$$\varphi_{\Lambda_{\rm uv}}(\vec{q}) = \varphi_\Lambda(\vec{q}) + \delta\varphi_\Lambda(\vec{q})\,, \tag{5.32}$$

with $\delta\varphi_\Lambda$ supported on $\Lambda < q < \Lambda_{\rm uv}$:

$$\begin{aligned}\varphi_\Lambda(\vec{q}) &= \varphi_{\Lambda_{\rm uv}}(\vec{q})\Theta(\Lambda - q) = \varphi(\vec{q})\Theta(\Lambda - q)\,,\\ \delta\varphi_\Lambda(\vec{q}) &= \varphi_{\Lambda_{\rm uv}}(\vec{q})\Theta(q - \Lambda) = \varphi(\vec{q})\Theta(\Lambda_{\rm uv} - q)\Theta(q - \Lambda)\,.\end{aligned} \tag{5.33}$$

Implementing equation (5.32) in equation (5.29) yields

$$\delta_{\Lambda_{\rm uv}}^{(n)}(\vec{k}) = \delta_\Lambda^{(n)}(\vec{k}) + \Delta\delta_{\Lambda,\Lambda_{\rm uv}}^{(n)}(\vec{k})\,, \tag{5.34}$$

with the IR PT term

$$\delta_\Lambda^{(n)}(\vec{k}) = \mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n} F_n(\vec{q}_1,\cdots,\vec{q}_n)\varphi_\Lambda(\vec{q}_1)\cdots\varphi_\Lambda(\vec{q}_n)\,, \tag{5.35}$$

and the mixed UV/IR contribution

$$\Delta\delta_{\Lambda,\Lambda_{\rm uv}}^{(n)}(\vec{k}) = \sum_{p=1}^{n} \begin{pmatrix} n \\ p \end{pmatrix} \mathcal{I}_{\vec{k};\vec{q}_1,\cdots,\vec{q}_n} F_n(\vec{q}_1,\cdots,\vec{q}_n)\delta\varphi_\Lambda(\vec{q}_1)\cdots\delta\varphi_\Lambda(\vec{q}_p)\varphi_\Lambda(\vec{q}_{p+1})\cdots\varphi_\Lambda(\vec{q}_n)\,. \tag{5.36}$$

Defining

$$\delta_\Lambda^{[N],\rm PT}(\vec{k}) \equiv \sum_{n=1}^{N} \delta_\Lambda^{(n)}(\vec{k})\,, \tag{5.37}$$

the counterterms at scale Λ are

$$\Delta\delta^{[N]}_{\Lambda}(\vec{k}) = \Delta\delta^{[N]}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) + \Delta\delta^{[N]}_{\Lambda_{\rm uv}}(\vec{k})\,, \tag{5.38}$$

with

$$\Delta\delta^{[N]}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) \equiv \sum_{n=1}^{N} \Delta\delta^{(n)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\,. \tag{5.39}$$

Here $\Delta\delta^{[N]}_{\Lambda_{\rm uv}}$ is not computable in PT and is fixed by renormalization conditions, while $\Delta\delta^{[N]}_{\Lambda,\Lambda_{\rm uv}}$ vanishes at $\Lambda = \Lambda_{\rm uv}$ and is perturbatively calculable for any Λ via equation (5.36).

5.4.3 Structure of the counterterms

The lowest orders clarify the counterterms' structure. At $n = 1$,

$$\Delta\delta^{(1)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) = \int d^3q\, \delta_D(\vec{k}-\vec{q})\delta\varphi_\Lambda(\vec{q}) = \delta\varphi_\Lambda(\vec{k}) = 0\,, \tag{5.40}$$

which vanishes for $k < \Lambda$. At $n = 2$, only the term with two UV fields survives:

$$\Delta\delta^{(2)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) = \mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2} F_2(\vec{q}_1,\vec{q}_2)\delta\varphi_\Lambda(\vec{q}_1)\delta\varphi_\Lambda(\vec{q}_2)\,. \tag{5.41}$$

Treating $\delta\varphi_\Lambda$ stochastically (i.e., integrating them out) gives

$$\langle\Delta\delta^{(2)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\rangle = \int \frac{d^3q}{(2\pi)^3} F_2(\vec{q},-\vec{q})P_{\Lambda,\Lambda_{\rm uv}}(q) = 0\,, \tag{5.42}$$

by momentum conservation; and, for Gaussian initial conditions,

$$\langle\Delta\delta^{(2)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\varphi_\Lambda(\vec{k}')\rangle = 0\,. \tag{5.43}$$

Its power spectrum is

$$\langle\Delta\delta^{(2)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\Delta\delta^{(2)}_{\Lambda,\Lambda_{\rm uv}}(-\vec{k})\rangle' = P^{22}_{\Lambda,\Lambda_{\rm uv}}(k)\,, \tag{5.44}$$

with

$$P^{22}_{\Lambda,\Lambda_{\rm uv}}(k) \equiv 2\int \frac{d^3q}{(2\pi)^3} F_2(\vec{q},\vec{k}-\vec{q})^2 P_{\Lambda,\Lambda_{\rm uv}}(q) P_{\Lambda,\Lambda_{\rm uv}}(|\vec{k}-\vec{q}|) \simeq \frac{127}{294}k^4 \int \frac{d^3q}{(2\pi)^3}\left(\frac{P_{\Lambda,\Lambda_{\rm uv}}(q)}{q^2}\right)^2 , \tag{5.45}$$

for $k \ll \Lambda$. Thus, to leading order,

$$\Delta\delta^{(2)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) \simeq \frac{k^2}{k^2_{\rm nl}}\,\epsilon^{(2)}_{\Lambda,\Lambda_{\rm uv}}, \tag{5.46}$$

where $\epsilon^{(2)}_{\Lambda,\Lambda_{\rm uv}}$ is uncorrelated with φ_Λ and has constant variance

$$P^{\epsilon}_{\Lambda,\Lambda_{\rm uv}} = \lim_{k/\Lambda\to 0}(k^4_{\rm nl}/k^4) P^{22}_{\Lambda,\Lambda_{\rm uv}}(k)$$

.

At $n = 3$,

$$\Delta\delta^{(3)}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) = \Delta\delta^{(3),a}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) + \Delta\delta^{(3),b}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\,, \tag{5.47}$$

with

$$\Delta\delta^{(3),a}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) \equiv 3\,\mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2,\vec{q}_3} F_3(\vec{q}_1,\vec{q}_2,\vec{q}_3)\delta\varphi_\Lambda(\vec{q}_1)\delta\varphi_\Lambda(\vec{q}_2)\varphi_\Lambda(\vec{q}_3)\,, \tag{5.48}$$

$$\Delta\delta^{(3),b}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) \equiv \mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2,\vec{q}_3} F_3(\vec{q}_1,\vec{q}_2,\vec{q}_3)\delta\varphi_\Lambda(\vec{q}_1)\delta\varphi_\Lambda(\vec{q}_2)\delta\varphi_\Lambda(\vec{q}_3)\,, \tag{5.49}$$

and

$$\begin{aligned}
&\langle\Delta\delta^{(3),a}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\rangle = 0\,,\\
&\langle\Delta\delta^{(3),a}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\varphi_\Lambda(-\vec{k})\rangle' \simeq -\,c^{(3)}_{\Lambda,\Lambda_{\rm uv}}\frac{k^2}{k^2_{\rm nl}}P_\Lambda(k)\,,
\end{aligned} \tag{5.50}$$

with

$$c^{(3)}_{\Lambda,\Lambda_{\rm uv}} \equiv \frac{61}{630}\int\frac{d^3q}{(2\pi)^3}\frac{P_{\Lambda,\Lambda_{\rm uv}}(q)}{q^2}\,. \tag{5.51}$$

Hence,

$$\Delta\delta^{(3),a}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) \simeq -\left(c^{(3)}_{\Lambda,\Lambda_{\rm uv}} + \eta^{(3)}_{\Lambda,\Lambda_{\rm uv}}\right)\frac{k^2}{k^2_{\rm nl}}\varphi_\Lambda(\vec{k})\,, \tag{5.52}$$

where $\eta^{(3)}_{\Lambda,\Lambda_{\rm uv}}$ is a scale-independent stochastic coefficient. The remaining term behaves as

$$\Delta\delta^{(3),b}_{\Lambda,\Lambda_{\rm uv}}(\vec{k}) \simeq \frac{k^2}{k^2_{\rm nl}}\,\epsilon^{(3)}_{\Lambda,\Lambda_{\rm uv}}, \tag{5.53}$$

with $\epsilon^{(3)}_{\Lambda,\Lambda_{\rm uv}}$ uncorrelated with φ_Λ.

5.4.4 Quadratic higher derivative terms

For $n > 3$, counterterms with $n - p = 2$ produce quadratic contributions in φ_Λ that correlate with $\varphi^{(2)}_\gamma$ and thus enter the determination of the bootstrap coefficient in equation (5.20). At $n = 4$, integrating out two $\delta\varphi_\Lambda$ yields in configuration space:

$$\begin{aligned} &c^{(n)}_{\varphi^2;\Lambda,\Lambda_{\rm uv}}\frac{\partial^2}{k^2_{\rm nl}}\varphi^2_\Lambda(\vec{x})\,, \quad c^{(n)}_{\varphi_\beta;\Lambda,\Lambda_{\rm uv}}\frac{\partial^2}{k^2_{\rm nl}}\varphi^{(2)}_{\beta,\Lambda}(\vec{x})\,,\\ &c^{(n)}_{\varphi_\gamma;\Lambda,\Lambda_{\rm uv}}\frac{\partial^2}{k^2_{\rm nl}}\varphi^{(2)}_{\gamma,\Lambda}(\vec{x})\,, \quad c^{(n)}_{\tilde{\gamma};\Lambda,\Lambda_{\rm uv}}\frac{\partial^2}{k^2_{\rm nl}}\,\varphi^{(2)}_{\tilde{\gamma},\Lambda}(\vec{x})\,. \end{aligned} \tag{5.54}$$

Their Λ-running, computable in PT, is

$$c^{(4)}_{X;\Lambda,\Lambda_{\rm uv}} = \frac{2096}{33957\,\pi^2}\,C_X\,k^2_{\rm nl}\int_\Lambda^{\Lambda_{\rm uv}} P(\Lambda')d\Lambda'\,, \tag{5.55}$$

with $X = \{\varphi^2,\,\varphi_\gamma,\,\varphi_\beta,\,\varphi_{\tilde{\gamma}}\}$ and

$$C_{\varphi^2} = 1\,, \quad C_{\varphi_\gamma} = \frac{4223}{10480}\,, \quad C_{\varphi_\beta} = \frac{32879}{20960}\,, \quad C_{\tilde{\gamma}} = \frac{1503}{1310}\,. \tag{5.56}$$

These are higher-derivative terms and thus suppressed by $O(k^2/k^2_{\rm nl})$ in Fourier space.

5.4.5 Running and renormalization

The running of $\Delta\delta^{[N]}_\Lambda(\vec{k})$ follows from that of $\Delta\delta^{[N]}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})$:

$$\Lambda\frac{d}{d\Lambda}\Delta\delta^{[N]}_{\Lambda}(\vec{k}) = \Lambda\frac{d}{d\Lambda}\Delta\delta^{[N]}_{\Lambda,\Lambda_{\rm uv}}(\vec{k})\,, \tag{5.57}$$

and is therefore perturbatively computable. Renormalization conditions are imposed as boundary data at a chosen scale $\Lambda = \bar{\Lambda}$ and reference momentum $\vec{k} = \bar{\vec{k}}$. One convenient choice is to fix conditions at $\bar{\Lambda} = \Lambda_{\rm uv}$. Symmetry arguments then imply, up to leading order in k,

$$\Delta\delta^{(2)}_{\Lambda_{\rm uv}}(\vec{k}) \simeq \frac{k^2}{k^2_{\rm nl}}\,\epsilon^{(2)}_{\Lambda_{\rm uv}}\,, \tag{5.58}$$

$$\Delta\delta^{(3)}_{\Lambda_{\rm uv}}(\vec{k}) \simeq \frac{k^2}{k^2_{\rm nl}}\,\epsilon^{(3)}_{\Lambda_{\rm uv}} - \left(c^{(3)}_{\Lambda_{\rm uv}} + \eta^{(3)}_{\Lambda_{\rm uv}}\right)\frac{k^2}{k^2_{\rm nl}}\varphi_{\Lambda_{\rm uv}}(\vec{k}) + O\left(\frac{k^4}{k^4_{\rm nl}}\right)\,. \tag{5.59}$$

The same structure holds at any Λ with $k \ll k_{\rm nl} \ll \Lambda$:

$$\Delta\delta^{(2)}_{\Lambda}(\vec{k}) \simeq \frac{k^2}{k^2_{\rm nl}}\,\epsilon^{(2)}_{\Lambda}\,, \tag{5.60}$$

$$\Delta\delta^{(3)}_{\Lambda}(\vec{k}) \simeq \frac{k^2}{k^2_{\rm nl}}\,\epsilon^{(3)}_{\Lambda} - \left(c^{(3)}_{\Lambda} + \eta^{(3)}_{\Lambda}\right)\frac{k^2}{k^2_{\rm nl}}\varphi_{\Lambda}(\vec{k}) + O\left(\frac{k^4}{k^4_{\rm nl}}\right)\,. \tag{5.61}$$

From equation (5.38),

$$c^{[3]}_{\Lambda} = c^{(3)}_{\Lambda} = c^{(3)}_{\Lambda,\Lambda_{\rm uv}} + c^{(3)}_{\Lambda_{\rm uv}}\,, \tag{5.62}$$

with running (from equation (5.51))

$$\Lambda\frac{d}{d\Lambda}c^{(3)}_{\Lambda} = \Lambda\frac{d}{d\Lambda}c^{(3)}_{\Lambda,\Lambda_{\rm uv}} = -\frac{61}{1260\,\pi^2}\Lambda\,k^2_{\rm nl}P(\Lambda)\,, \tag{5.63}$$

hence

$$c^{(3)}_{\Lambda} = c^{(3)}_{\Lambda_{\rm uv}} + \frac{61}{1260\,\pi^2}\,k^2_{\rm nl}\int_{\Lambda}^{\Lambda_{\rm uv}} d\Lambda' P(\Lambda')\,. \tag{5.64}$$

At $N = 5$, the sound-speed coefficient receives

$$\Delta\delta^{[5]}_{\Lambda}(\vec{k}) = \Delta\delta^{(3)}_{\Lambda}(\vec{k}) + \Delta\delta^{(5)}_{\Lambda}(\vec{k}) = -c^{[5]}_{\Lambda}\frac{k^2}{k^2_{\rm nl}}\varphi_{\Lambda}(\vec{k}) + \cdots\,, \tag{5.65}$$

with $c_\Lambda^{[5]} \equiv c_\Lambda^{(3)} + c_\Lambda^{(5)}$ and running

$$c_\Lambda^{(5)} = c_{\Lambda_{\rm uv}}^{(5)} + 15 \lim_{\vec{k}\to 0} \frac{k_{\rm nl}^2}{k^2} \int \frac{d^3q}{(2\pi)^3} \frac{d^3p}{(2\pi)^3} F_5(\vec{k},\vec{q},-\vec{q},\vec{p},-\vec{p}) P_{\Lambda,\Lambda_{\rm uv}}(q) P_{\Lambda,\Lambda_{\rm uv}}(p)\,. \tag{5.66}$$

A physically motivated renormalization condition sets, for each N,

$$c_{\Lambda=0}^{[N]} = \bar{c} \equiv \lim_{\vec{k}\to 0} \left(1 - \frac{\langle \bar{\delta}(\vec{k}) \varphi(-\vec{k}) \rangle'}{P(k)} \right) \frac{k_{\rm nl}^2}{k^2}\,, \tag{5.67}$$

with $\bar{\delta}(\vec{k})$ the data field (e.g., from simulations). Combining equations (5.64) and (5.66), and imposing $c_{\Lambda=0}^{[3]} = c_{\Lambda=0}^{[5]}$, one finds

$$c_\Lambda^{[5]} = c_\Lambda^{[3]} + 15 \lim_{\vec{k}\to 0} \frac{k_{\rm nl}^2}{k^2} \int \frac{d^3q}{(2\pi)^3} \frac{d^3p}{(2\pi)^3} F_5(\vec{k},\vec{q},-\vec{q},\vec{p},-\vec{p}) P(q) P(p) \Theta(\Lambda - q)^2 \Theta(\Lambda - p)^2\,. \tag{5.68}$$

In the formal limit $N \to \infty$ and $\Lambda \to \infty$, the purely nonperturbative sound speed remains:

$$c_s^2 = \lim_{N\to\infty} \lim_{\Lambda\to\infty} c_\Lambda^{[N]}\,. \tag{5.69}$$

For any finite N, running from the measured $\Lambda = 0$ value (equation (5.67)) up to large Λ is perturbative and thus calculable. The running of the higher-derivative coefficients follows from equation (5.55):

$$\Lambda \frac{d}{d\Lambda} c_{X;\Lambda,\Lambda_{\rm uv}}^{(4)} = -\frac{2096}{33957\,\pi^2}\, C_X\, \Lambda\, k_{\rm nl}^2 P(\Lambda)\,. \tag{5.70}$$

5.5 Setting up the analysis

5.5.1 Likelihood and data

The bootstrap-PT model is compared with ΛCDM simulations to assess the recoverability of the nonlinear bootstrap parameter ε_γ, a first step toward model-independent searches

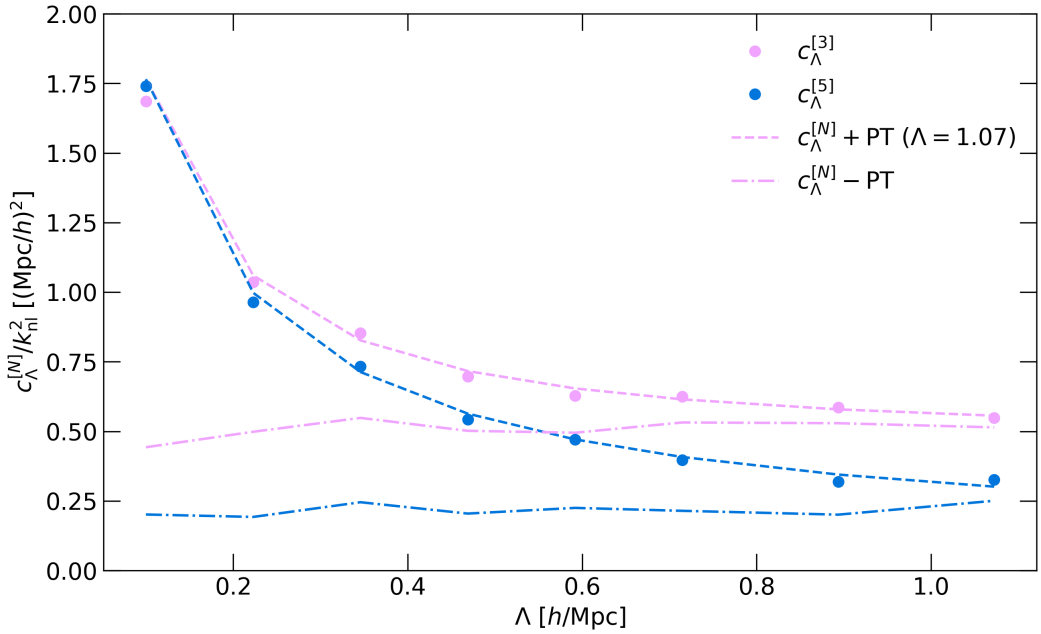

Figure 5.3: Running of $c_\Lambda^{[3]}$ (pink lines) and $c_\Lambda^{[5]}$ (blue lines) as measured from N-body simulations (dots). The running predicted by PT from equations (5.64) and (5.66) is shown by dashed lines. To facilitate comparison, the PT curves are shifted to match the N-body ones at $\Lambda = \Lambda_{\rm uv} = 1.07\ h\,{\rm Mpc}^{-1}$. The difference between the N-body measurement and the (unshifted) PT computation is shown by dot–dashed lines; their Λ-independent values approximate $c_{\Lambda_{\rm uv}}^{[3]}$ and $c_{\Lambda_{\rm uv}}^{[5]}$.

for new physics from real data. A Gaussian likelihood is assumed:

$$\log P_N[\delta|\varphi_0](k_{\max};\Lambda,\{\alpha^i_\Lambda\}) = -\frac{1}{2}\sum_n^{n_{\max}}\left[\frac{\left|\delta(\vec{k}_n)-\delta^{[N]}(\vec{k}_n)\right|^2}{L^3\,p_\epsilon(k_n)}\right]+\text{const}\,, \tag{5.71}$$

where $\vec{k}_n = \frac{2\pi}{L}\vec{n}$, $n_{\max} = k_{\max}\,L/(2\pi)$, and constant noise with

$$p_\epsilon(k_n) = \frac{1}{\bar{n}}\,, \tag{5.72}$$

is adopted. The set $\{\alpha^i_\Lambda\}$ denotes the model parameters, e.g., $\{c^{[N]}_\Lambda, \varepsilon_\gamma, \dots\}$ (with dots possibly including stochastic and higher-derivative terms). The data $\delta(\vec{k}_n)$ are computed from the overdensity field of dark matter particles measured from an N-body simulation with the same initial conditions $\varphi(\vec{k})$ used to build $\delta^{[N]}(\vec{k}_n)$ via `GridSPT`; i.e., both share the same random seeds and cosmological parameters.

A cosmological N-body simulation from the ongoing `Dark Quest II` project (Nishimichi *et al.*, in prep.) is used. It employs $3\,000^3$ particles in a comoving volume of $(1\,h^{-1}\text{Gpc})^3$. The assumed cosmology is the best-fitting flat ΛCDM model to the Planck temperature and polarization spectra [177]. Massive neutrinos are ignored (set to $\sum m_\nu = 0$) for simplicity in comparing with perturbation theory.[3]

Initial conditions are generated at $z = 91$ using second-order Lagrangian PT [141, 179], with a Gaussian field following the `CLASS` power spectrum [180]. Particle evolution to $z = 0$ is performed with the `GINKAKU` Tree–Particle Mesh code, built on the FDPS library [181, 182],[4] with the Particle–Mesh extension from `GreeM` [183–185]. Tree forces are accelerated by `Phantom-GRAPE` [186, 187] with AVX-512 instructions [188].

The $z = 1$ snapshot is analyzed. The matter density field is constructed by cloud-in-cells (CIC) assignment [189] on a $2\,000^3$ grid ($0.5\,h^{-1}$Mpc resolution).

Artificial noise is added to the dark matter density field to mimic a realistic point-source catalog such as galaxies (Appendix B.3). The procedure is:

1. assuming a value for the average density of sources $\bar{n} = 2\cdot 10^{-3}\,(h\,\text{Mpc}^{-1})^3$, which means 2 million points per cubic comoving $h^{-1}\,$Gpc, to mimic a Euclid-like survey

[3] The Dark Quest II simulation suite treats massive neutrinos using the linear-response approach [178]. Scale-dependent linear growth from massive neutrinos can be incorporated consistently in PT. A simplified case without scale-dependent growth is considered here.

[4] https://github.com/FDPS/FDPS

[2];

2. computing the number count of points per grid cell as $\tilde{N}(\vec{x}) = (1 + \delta_m(\vec{x}))\bar{n}\,\Delta x^3$, where $\delta_m(\vec{x})$ is the dark matter density field and $\Delta x \equiv L/N_g$ is the size of a grid cell;

3. sampling a new number count for each cell $N(\vec{x})$ by assuming $\tilde{N}(\vec{x})$ as the average of a Poisson distribution, $N(\vec{x}) \sim \text{Poisson}(\tilde{N}(\vec{x}))$;

4. recalculating the noisy data by applying the definition $\delta(\vec{x}) = \frac{N(\vec{x})}{\bar{N}} - 1$, where $\bar{N} \equiv L^3\bar{n}$ is the average value of $N(\vec{x})$.

Ten independent noisy realizations are generated and analyzed below.

On the scales of interest, the mean number of particles per mode can be estimated as $\bar{N} \sim \bar{n}\,(2\pi/k_{\rm max})^3$, which yields $\bar{N} \sim 680$ for $k_{\rm max} = 0.09\,h\,{\rm Mpc}^{-1}$ and $\bar{N} \sim 180$ for $k_{\rm max} = 0.14\,h\,{\rm Mpc}^{-1}$ with the choices adopted here (see section 5.6). Since $\bar{N} \gg 1$, the Poisson distribution is well approximated by a Gaussian, justifying the form of the likelihood in equation (5.71).

5.5.2 Models

The analysis focuses on two PT models, $N = 3$ and $N = 5$. Proper renormalization is assessed by varying Λ and $k_{\rm max}$. The renormalized $N = 3$ model is

$$\delta^{[3]}(\vec{k}) = \delta^{[3]}_{\Lambda,\rm EdS}(\vec{k}) + \frac{1}{2}\varepsilon_\gamma\, a^{(2)}_{\rm EdS}\,\varphi^{(2)}_{\gamma,\Lambda}(\vec{x}) + \Delta\delta^{[3]}_\Lambda(\vec{k})\,, \tag{5.73}$$

where $\delta^{[3]}_{\Lambda,\rm EdS}(\vec{k})$ is computed with `GridSPT`. Up to $O(k^4/k^4_{\rm nl})$,

$$\Delta\delta^{[3]}_\Lambda(\vec{k}) = -c^{[3]}_\Lambda\frac{k^2}{k^2_{\rm nl}}\varphi_\Lambda(\vec{k}) + \frac{k^2}{k^2_{\rm nl}}\left(\epsilon^{[3]}_\Lambda - \eta^{[3]}_\Lambda\,\varphi_\Lambda(\vec{k})\right)\,, \tag{5.74}$$

with $\epsilon^{[3]}_\Lambda \equiv \epsilon^{(2)}_\Lambda + \epsilon^{(3)}_\Lambda$, $c^{[3]}_\Lambda = c^{(3)}_\Lambda$, and $\eta^{[3]}_\Lambda = \eta^{(3)}_\Lambda$.

For $N = 5$,

$$\delta^{[5]}(\vec{k}) = \delta^{[5]}_{\Lambda,\rm EdS}(\vec{k}) + \frac{1}{2}\varepsilon_\gamma\, a^{(2)}_{\rm EdS}\,\varphi^{(2)}_{\gamma,\Lambda}(\vec{x}) + \Delta\delta^{[5]}_\Lambda(\vec{k})\,, \tag{5.75}$$

with `GridSPT` in the EdS limit and

$$\Delta\delta^{[5]}_\Lambda(\vec{k}) = -\,c^{[5]}_\Lambda \frac{k^2}{k^2_{\rm nl}}\varphi_\Lambda(\vec{k}) + \frac{k^2}{k^2_{\rm nl}}\sum_X c^{[5]}_{X,\Lambda} X_\Lambda(\vec{k}) + \frac{k^2}{k^2_{\rm nl}}\left(\epsilon^{[5]}_\Lambda - \eta^{[5]}_\Lambda\,\varphi_\Lambda(\vec{k})\right) + O\left(\frac{k^2}{k^2_{\rm nl}}\tilde{\eta}^{[5]}_{X,\Lambda}X_\Lambda(\vec{k}), \frac{k^2}{k^2_{\rm nl}}Y^{(3)}_\Lambda(\vec{k})\right), \tag{5.76}$$

where $X = \{\varphi^2,\, \varphi_\gamma,\, \varphi_\beta,\, \varphi_{\tilde{\gamma}}\}$ denotes quadratic operators (section 5.4.4), $c^{[5]}_\Lambda = c^{(3)}_\Lambda + c^{(5)}_\Lambda$, $\epsilon^{[5]}_\Lambda \equiv \sum_{n=2}^5 \epsilon^{(n)}_\Lambda$, $c^{[5]}_{X,\Lambda} = c^{(4)}_{X,\Lambda}$, and $\eta^{[5]}_\Lambda \equiv \sum_{n=3}^5 \eta^{(n)}_\Lambda$.

5.5.3 Estimators of the parameters

Consider the Maximum a Posteriori (MAP) values of $c^{[N]}_\Lambda$ and ε_γ that maximize equation (5.71). Setting first derivatives to zero yields, at $N = 3$,

$$\bar{c}^{[3]}_\Lambda = -\frac{\sum_n^{n_{\rm max}} {\rm Re}\left[\left(\delta(\vec{k}_n) - \delta^{[3]}_{\Lambda,{\rm EdS}}(\vec{k})\right)\varphi_\Lambda(-\vec{k}_n)\right]}{\sum_n^{n_{\rm max}} \frac{k_n^2}{k^2_{\rm nl}}\left|\varphi_\Lambda(\vec{k}_n)\right|^2}, \tag{5.77}$$

$$\bar{\varepsilon}^{[3]}_{\gamma,\Lambda} = \frac{2}{a^{(2)}_\gamma}\frac{\sum_n^{n_{\rm max}} {\rm Re}\left[\left(\delta(\vec{k}_n) - \delta^{[3]}_{\Lambda,{\rm EdS}}(\vec{k}) + \bar{c}^{[3]}_\Lambda \frac{k_n^2}{k^2_{\rm nl}}\varphi_\Lambda(\vec{k}_n)\right)\varphi^{(2)}_{\gamma,\Lambda}(-\vec{k}_n)\right]}{\sum_n^{n_{\rm max}}\left|\varphi^{(2)}_{\gamma,\Lambda}(\vec{k}_n)\right|^2}, \tag{5.78}$$

where stochastic terms in equation (5.74) are neglected, being uncorrelated with φ_Λ and $\varphi^{(2)}_{\gamma,\Lambda}$. At $N = 5$,

$$\bar{c}^{[5]}_\Lambda = -\frac{\sum_n^{n_{\rm max}} {\rm Re}\left[\left(\delta(\vec{k}_n) - \delta^{[5]}_\Lambda(\vec{k}_n) - \text{“higher der.”}\right)\varphi_\Lambda(-\vec{k}_n)\right]}{\sum_n^{n_{\rm max}} \frac{k_n^2}{k^2_{\rm nl}}\left|\varphi_\Lambda(\vec{k}_n)\right|^2}, \tag{5.79}$$

$$\bar{\varepsilon}^{[5]}_{\gamma,\Lambda} = \frac{2}{a^{(2)}_\gamma}\frac{\sum_n^{n_{\rm max}} {\rm Re}\left[\left(\delta(\vec{k}_n) - \delta^{[5]}_\Lambda(\vec{k}_n) + \bar{c}^{[5]}_\Lambda \frac{k_n^2}{k^2_{\rm nl}}\varphi_\Lambda(\vec{k}_n) - \text{“higher der.”}\right)\varphi^{(2)}_{\gamma,\Lambda}(-\vec{k}_n)\right]}{\sum_n^{n_{\rm max}}\left|\varphi^{(2)}_{\gamma,\Lambda}(\vec{k}_n)\right|^2}, \tag{5.80}$$

where “higher derivative” denotes nonstochastic $O(k_n^2/k^2_{\rm nl}X_\Lambda)$ and $O(k_n^2/k^2_{\rm nl}Y^{(3)}_\Lambda)$ terms in equation (5.76); the latter do not correlate with $\varphi^{(2)}_{\gamma,\Lambda}$ and are therefore neglected.

As an initial check of renormalization, parameters are extracted from grids with different Λ and compared to the PT running. Figure 5.3 shows the simulation measurements

(solid) and PT predictions (dashed) for $\bar{c}_\Lambda^{[3]}$ (magenta) and $\bar{c}_\Lambda^{[5]}$ (blue). For visual comparison, the PT curves are shifted to match the N-body values at $\Lambda = \Lambda_{\rm uv} = 1.07\,h\,{\rm Mpc}^{-1}$; the differences (dot–dashed) approximate the Λ-independent $c^{[3]}_{\Lambda_{\rm uv}}$ and $c^{[5]}_{\Lambda_{\rm uv}}$.

A measurement of the EFTofLSS sound speed from N-body nonlinear power spectra [190] yields, at $z = 1$ (WMAP7-like cosmology), $\hat{c}_s^2 \simeq 0.4\,({\rm h}^{-1}\,{\rm Mpc})^2$ using

$$\hat{c}_s^2 = -\lim_{k\to 0} \frac{P_{\rm nl}(k) - P(k) - P_{\rm 1loop}(k)}{2k^2 P(k)}\,. \tag{5.81}$$

This should be compared with $c^{[3]}_{\Lambda_{\rm uv}}$; from figure 5.3, $c^{[3]}_{\Lambda_{\rm uv}}/k_{\rm nl}^2 \simeq 0.5\,({\rm h}^{-1}\,{\rm Mpc})^2$, in reasonable agreement given methodological and cosmological differences (and noting equation (5.81) requires an extrapolation to $k \to 0$, whereas equation (5.77) uses modes up to $k_{\rm max}$).

In evaluating $\bar{c}_\Lambda^{[5]}$, higher-derivative terms are neglected; the agreement with PT indicates that their impact on the running of $c_\Lambda^{[5]}$ is subleading. For the small parameter $\bar{\varepsilon}^{[5]}_{\gamma,\Lambda}$, however, additive higher-derivative contributions become relevant. These contributions are therefore computed analytically and added to the data-inferred value via

$$\begin{aligned}\Delta\varepsilon^{[5],X}_{\gamma,\Lambda} &= -\frac{2}{a_\gamma^{(2)}}\frac{c^{[5]}_{X;\Lambda,\Lambda_{\rm uv}}}{k_{\rm nl}^2}\,\frac{\sum_n^{n_{\rm max}} k_n^2\,{\rm Re}\left[X(\vec{k}_n)\varphi_\gamma^{(2)}(-\vec{k}_n)\right]}{\sum_n^{n_{\rm max}}\left|\varphi_\gamma^{(2)}(\vec{k}_n)\right|^2} \\ &\to -\frac{2}{a_\gamma^{(2)}}\frac{c^{[5]}_{X;\Lambda,\Lambda_{\rm uv}}}{k_{\rm nl}^2}\,\frac{\int^{k_{\rm max}}\frac{d^3k}{(2\pi)^3}k^2 P_{X\gamma}(k)}{\int^{k_{\rm max}}\frac{d^3k}{(2\pi)^3}P_{\gamma\gamma}(k)}\,,\end{aligned} \tag{5.82}$$

with $c^{[5]}_{X;\Lambda,\Lambda_{\rm uv}}$ from equation (5.55) and $P_{X\gamma}(k) \equiv \langle X(\vec{k})\varphi_\gamma^{(2)}(-\vec{k})\rangle'$. Only the perturbative running from $\Lambda_{\rm uv}$ to Λ is included (i.e., $c^{[5]}_{X,\Lambda_{\rm uv}} = 0$).

The largest wavenumber $k_{\rm max}$ entering the likelihood is a crucial choice. Since $\bar{\varepsilon}^{[3]}_\gamma$ scales as a (formal) linear-order quantity (ratio of 2-loop to 1-loop objects), whereas $\bar{\varepsilon}^{[5]}_\gamma$ is (formally) 1-loop, one expects $k_{\rm max}$ comparable to linear-PT reach for $N = 3$ and to 1-loop reach for $N = 5$. This expectation is quantified next.

5.6 Results

Summarizing section 5.5, two renormalized models are considered. In both cases,

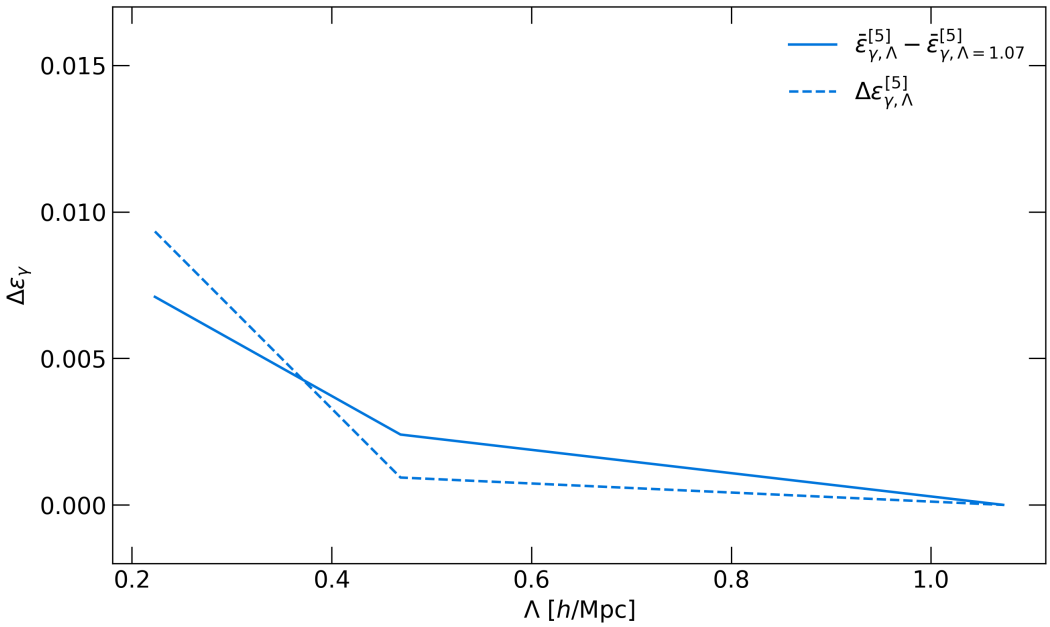

Figure 5.4: Running of ε_γ at $N = 5$ from N-body MAP estimates (solid; shown as the difference with the value at $\Lambda = \Lambda_{\rm uv} = 1.07\ h\,{\rm Mpc}^{-1}$) and from PT running via equation (5.70) inserted into equation (5.82) (dashed).

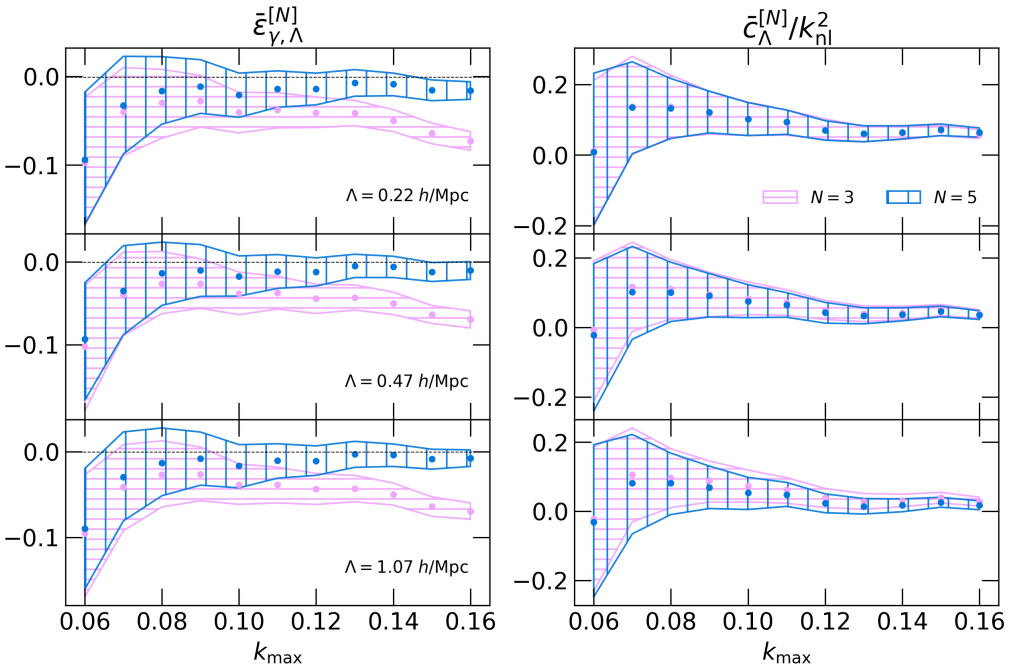

Figure 5.5: $k_{\rm max}$ dependence of the maximum-likelihood parameters $\bar{\varepsilon}^{[N]}_{\gamma,\Lambda}$ and $\bar{c}^{[N]}_{\Lambda}/k^2_{\rm nl}$ for $N = 3$ (pink) and $N = 5$ (blue). For $N = 5$, higher-derivative contributions are not included here.

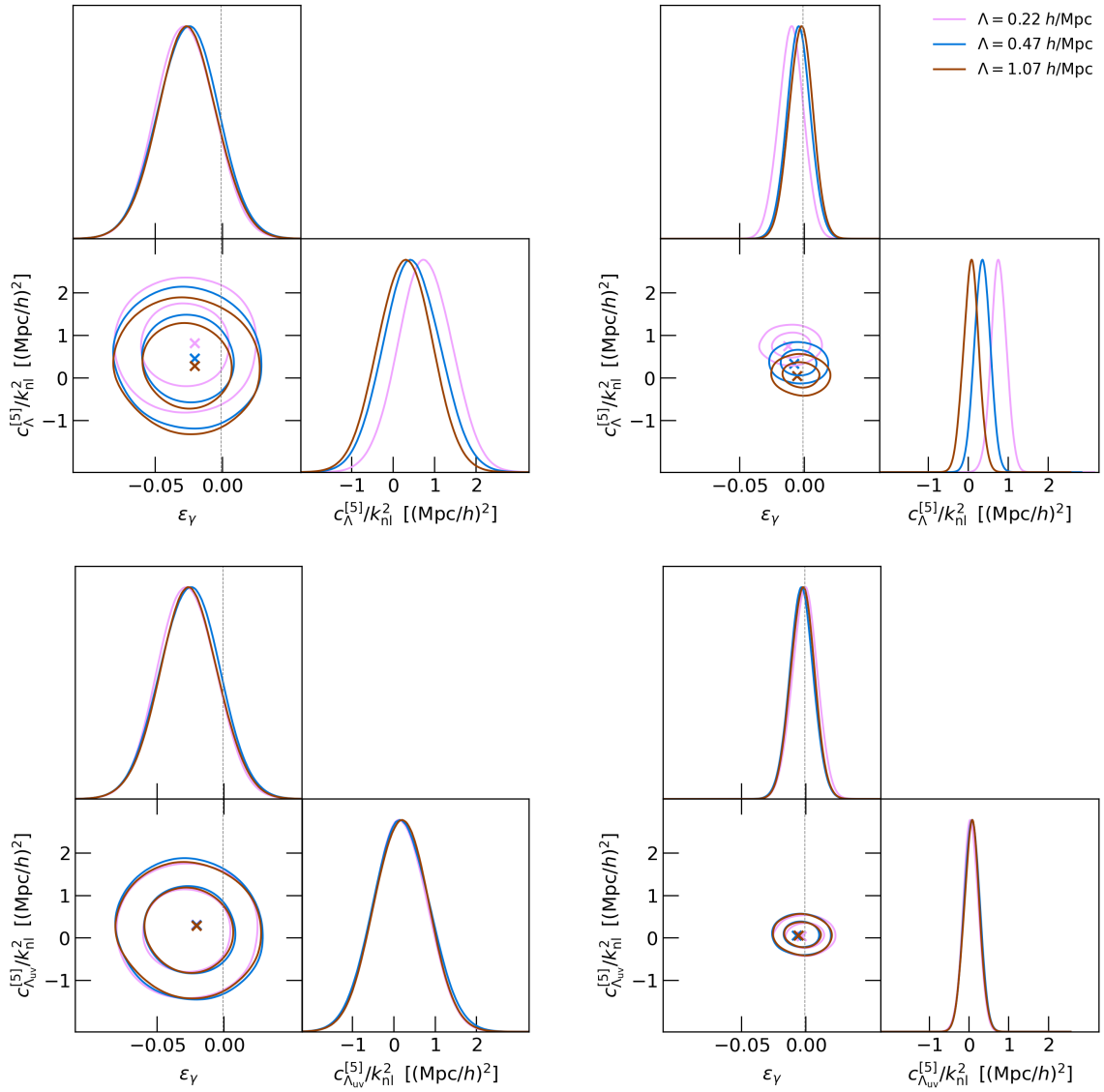

Figure 5.6: Results in terms of Λ-dependent (upper row) and Λ-independent parameters (lower row). Left: $N = 3$. Right: $N = 5$. Crosses indicate MAP values from equations (5.77)–(5.80) using the N-body data without added noise. The $k_{\rm max}$ values used are 0.09 $h\,\mathrm{Mpc}^{-1}$ ($N = 3$) and 0.14 $h\,\mathrm{Mpc}^{-1}$ ($N = 5$). To facilitate comparison, contours for ϵ_γ versus $c_\Lambda^{[5]}/k_{\rm nl}^2$ are shown also for $N = 3$, relating $c_\Lambda^{[3]}/k_{\rm nl}^2$ to $c_\Lambda^{[5]}/k_{\rm nl}^2$ via equation (5.68). For a fair comparison with the noiseless MAP values, contours are derived by stacking and averaging the mean of the ten MCMC samples with different shot noise realizations.

$\Lambda\ [h\,\mathrm{Mpc}^{-1}]$	0.223	0.469	1.072
$N_g\ (N = 3)$	141	298	682
$N_g\ (N = 5)$	212	447	1024

Table 5.1: Correspondence between Λ and N_g.

$$\delta^{[N]}(\vec{k}) = \delta^{[N]}_{\Lambda,\mathrm{EdS}}(\vec{k}) + \frac{1}{2}\varepsilon_\gamma\, a^{(2)}_{\mathrm{EdS}}\, \varphi^{(2)}_{\gamma,\Lambda}(\vec{x}) - c^{[N]}_\Lambda \frac{k^2}{k^2_{\mathrm{nl}}}\varphi_\Lambda(\vec{k})\,, \tag{5.83}$$

and, for $N = 5$, the higher-derivative contributions of equation (5.82) are added analytically to the inferred ε_γ. The irrelevance of the stochastic terms in equations (5.74) and (5.76) has been checked explicitly.

Three values of the linear cutoff are adopted: $\Lambda/(h\,\mathrm{Mpc}^{-1}) = 0.22,\, 0.47,\, 1.07$ (table 5.1); the largest effectively acts as Λ_{uv}. The k_{max} dependence of ε_γ is inspected and k_{max} is set where the deviation from the truth reaches 1-σ. As illustrated in figure 5.5, this yields $k^{N=3}_{\mathrm{max}} = 0.09\,h\,\mathrm{Mpc}^{-1}$ and $k^{N=5}_{\mathrm{max}} = 0.14\,h\,\mathrm{Mpc}^{-1}$ for $\Lambda \gtrsim 0.4\,h\,\mathrm{Mpc}^{-1}$. The corresponding values of ε_γ and $c^{[5]}_\Lambda/k^2_{\mathrm{nl}}$ are shown in the top panels of figure 5.6 (without higher-derivative corrections to ε_γ). Posteriors are averaged over the ten shot-noise realizations.

Figure 5.6 (bottom) exhibits Λ-independent results after converting $c^{[5]}_\Lambda/k^2_{\mathrm{nl}}$ at $\Lambda = 0.22,\, 0.47\ h\,\mathrm{Mpc}^{-1}$ to $c^{[5]}_{\Lambda_{\mathrm{uv}}}/k^2_{\mathrm{nl}}$ using the perturbative running of equations (5.64) and (5.66), and after including the higher-derivative correction to ε_γ from equation (5.82). The alignment among different Λ choices confirms proper renormalization. Quantitatively, the precision on ε_γ improves by a factor ~ 2.4 when moving from $N = 3$ to $N = 5$, driven by the larger k_{max} accessible at higher perturbative order.

5.7 Discussion and conclusions

A first step toward field-level inference beyond ΛCDM in the nonlinear LSS regime has been presented, focusing on a model-independent precision test via the second-order bootstrap coefficient for matter in real space. A careful treatment of the UV cutoff induced by discretization is central to the framework. By defining the theory at a high cutoff $\Lambda_{\mathrm{uv}} \gg k_{\mathrm{nl}}$ and running down to lower Λ, higher-derivative corrections are controlled by a single physical scale k_{nl}, while $O(k^2/\Lambda^2_{\mathrm{uv}})$ effects are irrelevant. At any Λ, the model is given by a truncated PT series in terms of φ_Λ plus counterterms: UV physics ($k > \Lambda$) is absorbed into counterterm coefficients, while IR physics ($k < \Lambda$) is described by `GridSPT` extended to the bootstrap parametrization. Lowering Λ in a controlled manner enables coarser grids and hence computational savings.

Two renormalized models were studied, at $N = 3$ and $N = 5$. The resulting k_{max} values are consistent with those from EFTofLSS correlator analyses [191]. For $N = 5$, including higher-derivative terms quadratic in the linear fields is essential for Λ-independent, unbiased recovery of the bootstrap coefficient. The non-renormalization theorem for

galileon operators [192] does not apply here, since the coefficients renormalized are those of higher-derivative operators like $(\partial^2/k_{\rm nl}^2)\varphi_{\gamma,\Lambda}$.

The methodology readily extends to more realistic scenarios. Incorporating biased tracers is straightforward within this framework by adding bias operators to `GridSPT`, noting that bias parameters are in general renormalized starting at $O(k^0)$, not $O(k^2)$ [163, 164, 192]. Redshift-space distortions, already treated perturbatively in `GridSPT` [161] (see also [193, 194] for field-level approaches), can be included, while accurate modeling of Fingers-of-God remains a known challenge [195–197]. Application to real data will ultimately require full sampling of cosmological parameters and marginalization over the amplitudes and phases of the linear fields, which can be carried out with Hamiltonian Monte Carlo and related techniques [47, 198–200].

Chapter 6

Ongoing Work

The analysis presented in chapter 5 is centered on a simplified scenario, where the observable field of density fluctuations $\delta(\vec{x})$ is defined in real space, with the initial configuration of the field fixed to that of the N-body realization used for the data and then forward-modeled to later times via Eulerian perturbation theory. The aim of this chapter is to lay the groundwork to overcome these two simplifications and to extend the analysis to a more general and realistic setting. Section 6.1 presents the need to move the analysis from real space to redshift space, to approach a more realistic scenario closer to galaxy clustering survey data; section 6.2 provides an extension of the work from the statistical inference side, detailing the need to sample and marginalize over initial conditions using Hamiltonian Monte Carlo (HMC) sampling methods.

6.1 Field-level inference in redshift space

To accurately infer quantities of interest from LSS galaxy clustering data, it is essential to model not only the physical processes that lead to the formation of the gravitationally bound structures of the cosmic web (see section 2.3), but also of systematic observational effects that intervene in real observational surveys. Of these, *redshift space distortions* (RSD) are the major systematic contribution. Section 2.3.2 contains a standard treatment of RSD, with the transformation from the real space field δ to redshift space δ_S presented in equations (2.48) and (2.55). Standard perturbation theory (SPT) can be used to express δ_S order-by-order n in terms of convolutions of n linear fields $\delta^{(1)}$ and mode-mixing kernels Z_n, which can be constructed from the real space kernels for δ, $F_{m\leq n}$, and velocity divergence θ, $G_{m\leq n}$. While SPT produces a well-defined perturbative model at large scales, non-perturbative effects like the so-called "Fingers of God"

(FoG), represent the main challenge in performing inference in redshift space; FoG, being a fully non-perturbative effect, see Scoccimarro [195], can either be included in a model through phenomenological treatments, for example as in Taruya, Nishimichi, and Saito [196], or Eggemeier et al. [197], or through an effective field theory (EFT) approach, see Ivanov, Simonović, and Zaldarriaga [151], Nishimichi et al. [191], Senatore and Zaldarriaga [201], and D'Amico et al. [202]. Since the work presented in chapter 5 makes use of the EFTofLSS, the second approach provides the most natural extension to RSD, introducing one-loop counterterms like the quadrupole contribution $-c_2\mu^2(k/k_{\rm nl})^2\delta^{(1)}$ and the FoG contribution $-c_{4,\rm FoG}(f\mu\, k/k_{\rm nl})^4(1+f\mu^2)\delta^{(1)}$, along the same lines of equations 16 and 17 of Nishimichi et al. [191].

The numerical perturbation theory code `GridSPT` [159] was used to provide model for the real space density field; in Taruya, Nishimichi, and Jeong [161], the authors expand the code to also compute the redshift space field perturbatively via a novel iterative algorithm that keeps the calculation of the fields on the same regular grid without any interpolation. Analogously to equation (5.20), the LSS bootstrap introduces the parameter $d_\gamma^{(2)}$ within the velocity divergence at second order

$$\theta^{(2)} = \theta^{(2)}_{\rm EdS} + \frac{1}{2}\varepsilon_{d_\gamma} d^{(2)}_{\gamma,\rm EdS}\varphi^{(2)}_\gamma\,, \tag{6.1}$$

so ε_{d_γ}, defined from $d_\gamma^{(2)}$ analogously to equation (5.21), has to be inferred together with the density parameter $\varepsilon_{a_\gamma} \equiv \varepsilon_\gamma$ in a redshift space analysis. Then, if the goal is to test deviations from general relativity, a possibility is to parametrize the linear growth rate f with a parameter ε_f,

$$f = f_{\Lambda\rm CDM}(1+\varepsilon_f)\,, \tag{6.2}$$

where $f_{\Lambda\rm CDM}$ is given by solving the linear theory in a ΛCDM background [169]. Maximum a posteriori estimators for these new parameters can be readily derived by augmenting the `GridSPT` algorithm with the new parameters and taking the maximum of a log-posterior like the one in equation (5.71), the explicit results are shown in appendix B.4. The transformation to redshift space also introduces anisotropic stochasticity within the field, which must be modeled within the variance of the likelihood [112].

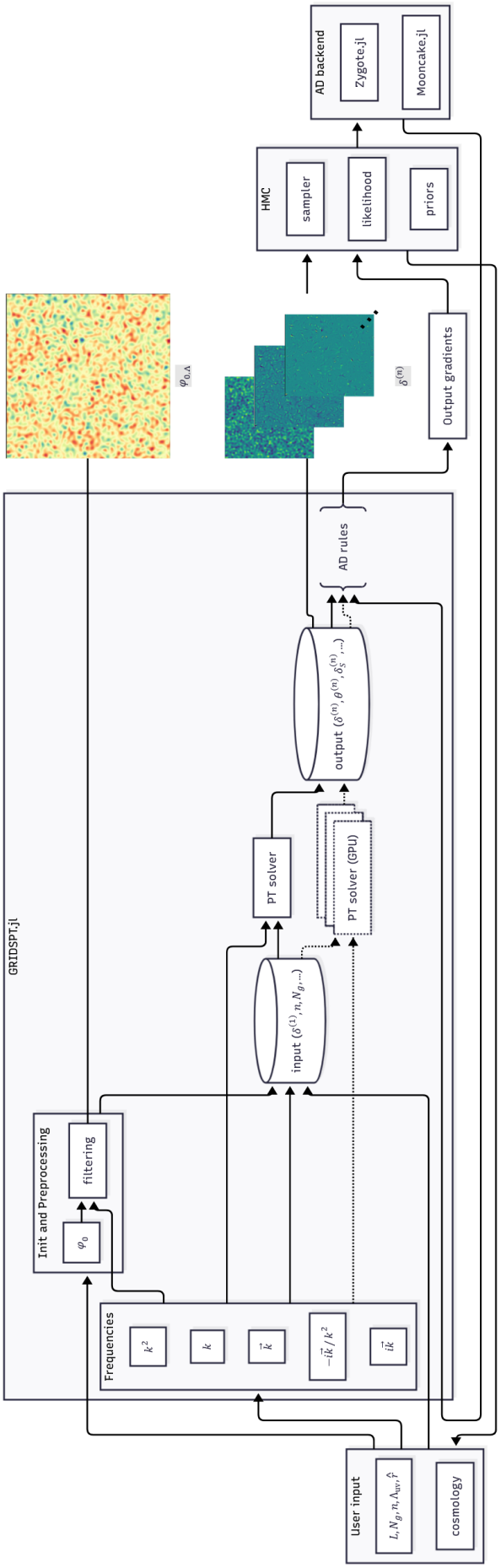

Figure 6.1: Flow chart diagram showing the schematic of the `GridSPT.jl` code. Similar to the original implementation of `GridSPT`, user inputs are processed through a perturbation theory (PT) solver and output fields are returned at every order up to n. `GridSPT.jl` includes more functionalities, like pre-computation of the Fourier modes, generation of the initial field, GPU acceleration and optimized automatic differentiation rules. The figure also shows a possible conceptual scheme for implementing Hamiltonian Monte Carlo.

6.2 High-dimensional inference of the initial conditions

Extraction of cosmological information from the observed density fluctuations and carried out by statistical inference on correlators like the power spectrum and bispectrum is generally done by assuming a specific cosmological model, often the ΛCDM model, for the background, and an inflationary scenario that seeds the initial perturbations. Taking equation (2.78) as an example, cosmological perturbation theory provides a straightforward way to model the nonlinear evolution of the perturbations in terms of the linear power spectrum P_{11} alone, itself parametrized in terms of A_s and n_s, the amplitude and spectral index of the primordial power spectrum. A_s and n_s represent two of the cosmological parameters that are jointly inferred from the data, and together with a handful of other parameters provided by the background theory, such as matter density $\Omega_{m,0}$ and the Hubble constant H_0,[1] inference can be carried out by sampling moderately low-dimensional posterior distributions with standard Metropolis-Hastings sampling algorithms. For field-level inference the complexity of the problem is significantly higher, because the theory is built on linear order fields $\delta^{(1)}(\vec{k}, a) = D(a)\varphi_0(\vec{k})$, where $\varphi_0(\vec{k})$ is a random field where each mode $\vec{k}$ is sampled according to a normal distribution in standard inflationary models. This means that the initial density field can be written as

$$\varphi_0(\vec{k}) \propto \mathcal{A}(\vec{k}) e^{i\theta(\vec{k})}\,, \tag{6.3}$$

where the amplitudes $\mathcal{A}(\vec{k})$ are distributed according to a Rayleigh distribution and the phases $\theta(\vec{k})$ are uniformly distributed in the range $[-\pi, \pi)$. Since the initial linear power spectrum $P_L(k) \equiv P_{11}(k, a = a_{\text{in}})$ is proportional to the ensemble average of the initial linear field, $\left\langle \varphi_0(\vec{k})\varphi_0(-\vec{k}') \right\rangle = (2\pi)^3 \delta_D(\vec{k} + \vec{k}') P_L(k)$, the complex exponential carrying the phases simplifies, and the above links directly the random amplitudes to the linear power spectrum:

$$\left\langle \mathcal{A}(\vec{k})\mathcal{A}^*(\vec{k}') \right\rangle \propto (2\pi)^3 \delta_D(\vec{k} - \vec{k}') P_L(k)\,. \tag{6.4}$$

This highlights how, in correlator-based inference, the randomness of the initial field is effectively absorbed in a function $P_L(k)$ that can be parametrized from first principles. However this is not the case at field-level, the randomness of $\varphi_0(\vec{k})$ has to be explicitly modeled with appropriate parametrization, for example by taking the real and imaginary parts of φ_0, since both are normally distributed. In practice, the fields need to

[1] In case of galaxy clustering, a number of bias parameters is also considered.

be discretized onto a regular cubic grid with N_g grid cells per side, and N_g is usually of $O(10^2 - 10^3)$, meaning that the random field will be composed by $O(10^6 - 10^9)$ voxels, which correspond to the number of extra parameters that need to be introduced in the inference problem. Such high-dimensional posteriors can only be sampled effectively with HMC sampling or similar advanced Markov chain Monte Carlo methodologies. Seminal work in this direction has been done during the development of the BORG algorithm [35, 203]. In chapter 5 the analysis was carried out by fixing φ_0 at the same realization of the initial field used to generate the N-body data, so the likelihood presented in equation 5.71 is a low-dimensional, conditional likelihood that can be readily sampled by standard random walk samplers; the goal now is to generalize the results by sampling over $\varphi_0(\vec{k})$, extending the set of parameters to be inferred from $\{\alpha^i_\Lambda\}$ to $\{\alpha^i\} = \{\alpha^i_\Lambda\} \cup \{\mathrm{Re}[\varphi_0](\vec{k}_i)\} \cup \{\mathrm{Im}[\varphi_0](\vec{k}_i)\}$, with $i = 1, \cdots, N_g^3/2$ because $\varphi_0(\vec{k})$ is the Fourier transform of a real field, and obtain a final posterior distribution on the parameters of interest by marginalizing the initial random state $\{\mathrm{Re}[\varphi_0](\vec{k}_i)\} \cup \{\mathrm{Im}[\varphi_0](\vec{k}_i)\}$. As expressed in section 3.2, Hamiltonian Monte Carlo requires that the posterior distribution is differentiable over the parameters; in other words, the models used in section 5.5.2 to evolve $\varphi_0(\vec{k})$ forward to $\delta^{[N]}(\vec{k})$ must have a computable gradient. Analytical computation of $\vec{\nabla}_{\{\alpha^i\}}\delta^{[N]}(\vec{k})$ is possible but cumbersome and inefficient given the number of parameters involved. A more practical approach is to leverage *automatic differentiation* (AD): automatic differentiation treats any computer program as a composition of simple operations with known derivative rule, so the derivative of the program can be obtained by evaluating and multiplying a large number of simple derivatives according to the chain rule; see appendix B.5 and Krämer [204] for a more technical treatment. `GridSPT` can be readily adapted to accomodate AD due to the lack of non-differentiable operations and results independent on control flow (e.g., "if-else" statements), and since most operations are performed on the regular grid voxel-by-voxel (element-wise multiplications and sums of dense 3D arrays) or require common algorithms like the fast Fourier transform, the program can also be massively parallelized on the GPU for increased performance. For these reasons, a GPU-accelerated, auto-differentiable version of `GridSPT`, `GridSPT.jl`, has been developed in preparation for the analysis. On top of performance improvements and AD, `GridSPT.jl` expands the original implementation by generalizing the computation of the redshift space density field δ_S up to arbitrary order, and providing an interface to generate initial random fields with user-defined cosmological information. Figure 6.1 presents a conceptual schematic of the code applied to HMC.[2].

Once the code is ready for production, the actual implementation of HMC will require careful consideration. Commonly used algorithms for Hamiltonian Monte Carlo,

[2]Testing and progress on the code is available at `https://github.com/MatPeron/GridSPT.jl-tests`.

like the "No U-Turn Sampler" (NUTS), make use of "mass matrix adaptation", a technique that performs an initial exploration of the posterior and tunes the mass matrix of the kinetic term to efficiently explore the distribution. This works well as long as the posterior covariance is not ill-conditioned, which may happen when sampling degenerate parameters like A_s and the linear galaxy bias b_1. Degenerate parameters can be taken care of by either imposing some informative priors or introducing Gibbs steps into the sampling, an approach used also by Kostić et al. [43]. Augmentations to the HMC sampler like the "microcanonical Hamiltonian Monte Carlo sampler" [205, 206] and samplers augmented via multi-layer perceptrons [207] could be a further consideration to deal both with challenging posteriors and improve sampling efficiency.

Chapter 7

Conclusions

The central aim of this thesis has been to design, analyze, and connect principled approaches that close the gap between existing techniques and optimal information extraction of late-time structure formation, focusing specifically on the discovery of new physics. The chapters proceed along two complementary axes:

1. **Optimal summaries of the data:** building and testing summary statistics that capture as much cosmological information as possible beyond the two-point function, focusing specifically on wavelet scattering transforms (WST) for primordial non-Gaussianity (PNG).

2. **Field-level forward modeling:** formulating a renormalized perturbative forward model at the field level, based on the the bootstrap approach [51, 96, 149] to parameterize deviations from the standard ΛCDM model.

The first axis addresses the question of *what* should be measured: summary statistics that are both optimal (or near-optimal) for targeted parameter extraction, like $f_{\rm NL}$, and explainable. The second axis emphasizes *how* models should be developed to control systematics, such as discretization effects introduced by the grid, and infer as much information as possible on new physics from measurements. Both converge on a practical program: use the available theoretical tools to construct forward models that are renormalized at the resolution of the data for general cosmological inference, and apply summary statistics that are suitable for addressing specific problems.

7.1 Summary of the results

The introductory chapters 2 and 3 review the cosmology of large scale structure (LSS) and the core tools of statistical inference used throughout the thesis, including Bayesian parametric inference, likelihood-based forecasting with the Fisher matrix, and high-dimensional samplers such as Hamiltonian Monte Carlo (HMC). These preliminaries provide a foundation for the two main research threads.

- A summary-statistics approach based on WST, designed to compress the non-linear, multi-scale information relevant to PNG and late-time cosmological parameters, presented in chapter 4.
- A field-level, renormalized perturbation-theory (PT) framework in which the discretization scale is explicit and the resulting model is computationally tractable on finite grids, presented in chapter 5.

7.1.1 Wavelet scattering for PNG and cosmology

Scope and setup. The first research component investigates the WST as a tool for PNG studies and compares its performance with a conventional power spectrum plus bispectrum analysis (P+B). The analysis is carried out for matter and halo fields, in both real and redshift space, using the public Quijote and QuijotePNG suites, to produce Fisher forecasts for the amplitudes $f_{\rm NL}^{\rm local}$, $f_{\rm NL}^{\rm equil}$, and $f_{\rm NL}^{\rm ortho}$ along with standard cosmological parameters. For a fair comparison, a sharp $k_{\rm max} = 0.5\,h\,{\rm Mpc}^{-1}$ cutoff is imposed to the field before computing both WST and P+B, and we study $q \in \{0.8, 2\}$, noting that $q = 2$ offers a clean interpretation: first-order scattering coefficients become equivalent to the power spectrum, therefore all gains must arise from second order.

Main findings. Across PNG and standard parameters (as well as a halo nuisance parameter $M_{\rm min}$), WST outperforms the power spectrum alone. For $f_{\rm NL}^{\rm local}$ in halos, WST improves forecasts by nearly 27%. When combined with the bispectrum, halo constraints from WST are slightly weaker for $f_{\rm NL}^{\rm local}$ ($\approx 15\%$), but stronger for $f_{\rm NL}^{\rm equil}$ ($\approx 25\%$) and $f_{\rm NL}^{\rm ortho}$ ($\approx 28\%$ for $q = 0.8$). Results for standard cosmological parameters are consistent with previous findings that WST beats the power spectrum alone [30, 31]. For PNG, WST is competitive with the best performing alternative summaries, such as marked statistics [145].

Outlook. The WST offers a compact, interpretable statistic: with $q = 2$, it recovers the power spectrum at first order and adds new non-Gaussian information at second

order. This aligns well with the desiderata of near-sufficient compression while remaining practical for survey pipelines. On the theory side, there is room for further improvement: tailoring wavelet families to PNG extraction (e.g., leveraging angular structure to mitigate redshift-space distortions) and encoding symmetries such as extended Galilean invariance [51, 93, 94, 146–148] could reduce dimensionality further without sacrificing constraining power. On the data side, realism (galaxy bias, selection and window effects, masks, and other systematics) is the next step toward deployment for surveys such as DESI and Euclid.

7.1.2 Field-level inference with finite UV cutoff and renormalized PT

Scope and setup. The second research component presents a first step toward field-level inference beyond ΛCDM in the non-linear regime, via a model-independent precision test using the bootstrap approach for the clustering of matter in real space. A key ingredient is an explicit treatment of the UV cutoff induced by discretization. The theory is defined at a large cutoff $\Lambda_{\rm uv} \gg k_{\rm nl}$ and then scaled down to lower Λ, such that higher-derivative corrections are organized by a single physical scale $k_{\rm nl}$, while $O(k^2/\Lambda_{\rm uv}^2)$ terms remain negligible. At any Λ, the model is a truncated PT series in terms of fields φ_Λ plus counterterms: modes above Λ renormalize inferred operator coefficients, while modes below Λ are modeled using `GridSPT`, extended with the bootstrap parametrization.

Main findings. Two renormalized models, at orders $N = 3$ and $N = 5$, are considered. The resulting $k_{\rm max}$ values are consistent with those from EFTofLSS correlator analyses [191]. For $N = 5$, higher-derivative terms quadratic in the linear fields are essential to achieve Λ-independent, unbiased recovery of the bootstrap coefficient. The galileon non-renormalization theorem [192] is not applicable in this setting, because the renormalized coefficients in question are those of higher-derivative operators like $(\partial^2/k_{\rm nl}^2)\varphi_{\gamma,\Lambda}$.

Outlook. Generalizing the methodology is straightforward: biased tracers can be incorporated by adding the standard bias operator basis to `GridSPT`, keeping in mind that bias parameters are generally renormalized already at $O(k^0)$, not only at $O(k^2)$ [163, 164, 192]. Redshift space distortions (RSD) can be included (see [161] and field-level developments [193, 194]), though accurate Fingers-of-God (FoG) modeling remains an open challenge [195–197]. Ultimately, application to data requires joint sampling of cosmological parameters and marginalization over the amplitudes and phases of linear fields, which is naturally addressed with HMC and related methods [47, 198–200]. Lowering Λ in a controlled way enables coarser grids and significant computational savings without loss

of accuracy.

7.2 Next steps: toward RSD and marginalization over initial conditions

The final chapter of this thesis, chapter 6, outlines and initiates the generalization of the work on field-level inference to realistic survey observables. What follows details an approach to address some of the challenges expressed at the end of section 7.1.2:

1. **Redshift-space distortions.** The SPT model in redshift space is augmented within the bootstrap of LSS framework D'Amico et al. [51], and by the addition of EFT counterterms induced by the line-of-sight projection and FoG damping [191]. The perturbative RSD kernels available in `GridSPT` [161] provide a starting point, but further work is required to carry out renormalization along the lines of the previous work.

2. **Marginalization over initial fields.** In response to the need to sample a joint posterior over cosmological parameters and initial linear modes, with a Gaussian prior on the initial random configuration, a GPU-accelerated and automatically differentiable version of `GridSPT` has been developed; the new implementation is designed to perform HMC-based sampling of the high-dimensional posterior. As the code emerges from its testing phase, considerations on how mass matrix preconditioning and novel samplers can improve computational efficiency and overcome degenerate directions in parameter space become necessary. Marginalization over the linear modes is the final technical step required to go from the fixed initial conditions proof of concept to a full field-level inference pipeline.

These two steps are modular: RSD and initial conditions sampling and marginalization can be enabled independently and later coupled, facilitating staged validation and incremental complexity towards more and more realistic inference problems.

7.3 Limitations

It is important to emphasize several limitations. On the summary–statistics side, constraints derived from the WST rely on Fisher forecasts built on simulations with idealized

conditions. Covariances have been estimated without full survey realism, and the sharp $k_{\max}$ filtering adopted for fair comparison with P+B and to compare to similar results in the literature extends well beyond values of k that can be probed with standard perturbation theory. Performance depends on the choice of wavelet family and hyperparameters (see [31, 208] for different possibilities), and performance across cosmologies and tracers remains unknown.

Moving to field-level, extensions to biased tracers are conceptually straightforward but challenging on the probabilistic modeling side, as the stochastic noise terms entering the EFT likelihood [209, 210] can compromise its assumed Gaussianity and introduce biases or underestimate uncertainties on the inferred parameters, as was highlighted recently by Akitsu et al. [50]; truncation errors at finite perturbative order and possible systematics induced by prior choices must be characterized jointly and disentangled from the UV running induced by discretization. Sampling the joint posterior over cosmological parameters and initial modes remains computationally intensive and sensitive to mass–matrix preconditioning and automatic differentiation details; algorithmic choices can interact non-trivially with model degeneracies and stochastic terms [39, 42, 46, 47, 49]. Redshift-space modeling remains a leading source of systematic uncertainty; the pairwise-velocity probability density function is non-Gaussian across scales, complicating the simple damping ansatz for Fingers of God (FoG) effects and limiting the usable scale range; degeneracies with bias and selection further obscure parameter sensitivity [195–197]. While approaches are advancing, survey-ready prescriptions with uniform performance across tracers and redshifts are still under active development [47, 194].

7.4 Outlook

Several directions appear both natural and impactful. For WST, besides tailoring scattering filters (e.g., anisotropic wavelets) to account for RSD anisotropy and relevant symmetries, and integrating survey realism, future works could explore coupling WST summaries to likelihood-free or semi-analytic approaches such as Makinen et al. [211] or leverage learnable scattering architectures and hyperparameter tuning methods developed for machine learning applications [212–214].

For field-level inference, the priority is a renormalized redshift-space forward model that marries the bootstrap construction with EFT counterterms. In parallel, full marginalization over initial conditions and the need to test prior choices and a variety of scale cuts invites the technical challenge of developing SPT codes that are optimized with state of the art high performance computing techniques. HMC preconditioning that is

robust to the choice of initial guess and parametric degeneracies is also a concern, and methodologies along the lines of Jasche and Wandelt [35] could be followed. Recent works clarify how field-level posteriors relate to n-point information and provide diagnostics for optimality [42, 47, 49]. Hybrid perturbative models with force integration [46, 215] can be used in place of standalone perturbation theory as a forward model in field-level inference. Simulation-based approaches are still immature for a direct application to the uncompressed field, but recent works [123] show that there is interest in this direction; while being a black-box, the more powerful modeling capabilities provided by neural architectures could be leveraged to assess the impact of non-Gaussianity of the field-level likelihood, and so provide clues to pave the way to more principled approaches.

7.5 Closing remarks

This thesis has explored two complementary routes toward near-optimal information extraction from late-time structure formation. Wavelet scattering provides compact, interpretable summaries that capture non-Gaussian structure beyond two-point analyses and can be tuned to specific signals of interest. A renormalized, discretization-aware field-level program offers a principled path to unbiased inference at finite resolution and a transparent connection to theory, with clear extensions to biased tracers and redshift space. The two axes reinforce each other: summary statistics indicate where information concentrates and supply robust cross-checks, while field-level modeling provides the theoretical backbone that unifies disparate observables. With forthcoming surveys placing systematic control and modeling fidelity at center stage, the combination of controlled forward models and task-specific compression appears well positioned to translate statistical precision into robust constraints on primordial physics, gravity, and undiscovered new phenomena of the Universe.

Bibliography

[1] Amir Aghamousa et al. *The DESI Experiment Part I: Science,Targeting, and Survey Design*. In: (Oct. 2016).

[2] Luca Amendola et al. *Cosmology and fundamental physics with the Euclid satellite*. In: *Living Rev. Rel.* 21.1 (2018), p. 2.

[3] Y. Mellier et al. *Euclid. I. Overview of the Euclid mission*. In: (May 2024).

[4] Guido D'Amico et al. *Limits on primordial non-Gaussianities from BOSS galaxy-clustering data*. In: (Jan. 2022).

[5] Guido D'Amico et al. *The BOSS bispectrum analysis at one loop from the Effective Field Theory of Large-Scale Structure*. In: (June 2022).

[6] Oliver H. E. Philcox et al. *Cosmology with the redshift-space galaxy bispectrum monopole at one-loop order*. In: *Phys. Rev. D* 106.4 (2022), p. 043530.

[7] Mikhail M. Ivanov et al. *Cosmology with the galaxy bispectrum multipoles: Optimal estimation and application to BOSS data*. In: *Phys. Rev. D* 107.8 (2023), p. 083515.

[8] ChangHoon Hahn et al. SimBIG*: The First Cosmological Constraints from the Non-Linear Galaxy Bispectrum*. In: (Oct. 2023).

[9] Ravi K. Sheth. *The halo-model description of marked statistics*. In: *Mon. Not. Roy. Astron. Soc.* 364 (2005), p. 796.

[10] Ravi K. Sheth, Andrew J. Connolly, and Ramin Skibba. *Marked correlations in galaxy formation models*. In: (Nov. 2005).

[11] Oliver H. E. Philcox, Elena Massara, and David N. Spergel. *What does the marked power spectrum measure? Insights from perturbation theory*. In: *Phys. Rev. D* 102.4 (2020), p. 043516.

[12] Elena Massara et al. *Using the Marked Power Spectrum to Detect the Signature of Neutrinos in Large-Scale Structure*. In: *Phys. Rev. Lett.* 126.1 (2021), p. 011301.

[13] Elena Massara et al. *Cosmological Information in the Marked Power Spectrum of the Galaxy Field*. In: *Astrophys. J.* 951.1 (2023), p. 70.

[14] Dipak Munshi and Alan Heavens. *A New Approach to Probing Primordial Non-Gaussianity*. In: *Mon. Not. Roy. Astron. Soc.* 401 (2010), p. 2406.

[15] Azadeh Moradinezhad Dizgah et al. *Capturing non-Gaussianity of the large-scale structure with weighted skew-spectra*. In: *JCAP* 04 (2020), p. 011.

[16] Marcel Schmittfull and Azadeh Moradinezhad Dizgah. *Galaxy skew-spectra in redshift-space*. In: *JCAP* 03 (2021), p. 020.

[17] Jiamin Hou et al. SimBIG*: Cosmological Constraints from the Redshift-Space Galaxy Skew Spectra*. In: (Jan. 2024).

[18] Geraint Pratten and Dipak Munshi. *Non-Gaussianity in large-scale structure and Minkowski functionals*. In: *MNRAS* 423.4 (July 2012), pp. 3209–3226.

[19] Wei Liu, Aoxiang Jiang, and Wenjuan Fang. *Probing massive neutrinos with the Minkowski functionals of large-scale structure*. In: *J. Cosmology Astropart. Phys.* 2022.7, 045 (July 2022), p. 045.

[20] Sabino Matarrese, Francesco Lucchin, and Silvio A. Bonometto. *A Path-Integral Approach to Large-Scale Matter Distribution Originated by Non-Gaussian Fluctuations*. In: *ApJ* 310 (Nov. 1986), p. L21.

[21] Sabino Matarrese, Licia Verde, and Raul Jimenez. *The Abundance of high-redshift objects as a probe of non-Gaussian initial conditions*. In: *Astrophys. J.* 541 (2000), p. 10.

[22] James Robinson, Eric Gawiser, and Joseph Silk. *Constraining primordial non-gaussianity with the abundance of high redshift clusters*. In: *Astrophys. J.* 532 (2000), p. 1.

[23] Marilena LoVerde and Kendrick M. Smith. *The Non-Gaussian Halo Mass Function with f_NL, g_NL and \tau_NL*. In: *JCAP* 08 (2011), p. 003.

[24] Adrian E. Bayer et al. *Detecting neutrino mass by combining matter clustering, halos, and voids*. 2021.

[25] Gabriel Jung et al. *Quijote-PNG: The Information Content of the Halo Mass Function*. In: *Astrophys. J.* 957 (2023), p. 1.

[26] Marc Kamionkowski, Licia Verde, and Raul Jimenez. *The void abundance with non-gaussian primordial perturbations*. In: *J. Cosmology Astropart. Phys.* 2009.1, 010 (Jan. 2009), p. 010.

[27] Guido D'Amico et al. *Excursion sets and non-Gaussian void statistics*. In: *Phys. Rev. D* 83.2, 023521 (Jan. 2011), p. 023521.

[28] Stéphane Mallat. *Group Invariant Scattering*. In: *Communications on Pure and Applied Mathematics* 65.10 (2012), pp. 1331–1398.

[29] Mathieu Andreux et al. *Kymatio: Scattering Transforms in Python*. In: *arXiv e-prints*, arXiv:1812.11214 (Dec. 2018), arXiv:1812.11214.

[30] Georgios Valogiannis and Cora Dvorkin. *Towards an optimal estimation of cosmological parameters with the wavelet scattering transform*. In: *Phys. Rev. D* 105.10 (2022), p. 103534.

[31] Michael Eickenberg et al. *Wavelet Moments for Cosmological Parameter Estimation*. In: (Apr. 2022).

[32] Georgios Valogiannis and Cora Dvorkin. *Going beyond the galaxy power spectrum: An analysis of BOSS data with wavelet scattering transforms*. In: *Phys. Rev. D* 106.10 (2022), p. 103509.

[33] Bruno Régaldo-Saint Blancard et al. SimBIG*: Galaxy Clustering Analysis with the Wavelet Scattering Transform*. In: (Oct. 2023).

[34] Georgios Valogiannis, Sihan Yuan, and Cora Dvorkin. *Precise Cosmological Constraints from BOSS Galaxy Clustering with a Simulation-Based Emulator of the Wavelet Scattering Transform*. In: (Oct. 2023).

[35] Jens Jasche and Benjamin D. Wandelt. *Bayesian physical reconstruction of initial conditions from large scale structure surveys*. In: *Mon. Not. Roy. Astron. Soc.* 432 (2013), p. 894.

[36] F. S. Kitaura. *The initial conditions of the universe from constrained simulations*. In: *Monthly Notices of the Royal Astronomical Society* 429 (Feb. 2013), pp. L84–L88.

[37] H. Wang et al. *Reconstructing the Initial Density Field of the Local Universe: Methods and Tests with Mock Catalogs*. In: *The Astrophysical Journal* 772 (July 2013), p. 63.

[38] H. Wang et al. *ELUCID - Exploring the Local Universe with reConstructed Initial Density field I: Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics*. In: *The Astrophysical Journal* 794 (2014), p. 94.

[39] J. Jasche and G. Lavaux. *Physical Bayesian modelling of the non-linear matter distribution: new insights into the Nearby Universe*. In: *Astronomy & Astrophysics* 625 (2019), A64.

[40] G. Lavaux, J. Jasche, and F. Leclercq. *Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data*. In: (2019).

[41] F. Schmidt et al. *A rigorous EFT-based forward model for large-scale structure*. In: *Journal of Cosmology and Astroparticle Physics* 1901 (2019), p. 042.

[42] C. Modi, Y. Li, and D. Blei. *Reconstructing the Universe with Variational self-Boosted Sampling*. In: *Proceedings of the 39th International Conference on Machine Learning*. 2022.

[43] A. Kostić et al. *Consistency tests of field level inference with the EFT likelihood*. In: *Journal of Cosmology and Astroparticle Physics* 2023.07 (2023), p. 063.

[44] Adam Andrews et al. *Bayesian field-level inference of primordial non-Gaussianity using next-generation galaxy surveys*. In: *MNRAS* 520.4 (Apr. 2023), pp. 5746–5763.

[45] A. E. Bayer, U. Seljak, and C. Modi. *Field-Level Inference with Microcanonical Langevin Monte Carlo*. July 2023.

[46] L. Doeser et al. *Bayesian inference of initial conditions from non-linear cosmic structures using field-level emulators*. In: *Monthly Notices of the Royal Astronomical Society* 535 (Dec. 2024), pp. 1258–1277.

[47] Nhat-Minh Nguyen et al. *How much information can be extracted from galaxy clustering at the field level?* In: (Mar. 2024).

[48] G. Cabass, M. Simonović, and M. Zaldarriaga. *Cosmological information in perturbative forward modeling*. In: *Physical Review D* 109 (Feb. 2024), p. 043526.

[49] Fabian Schmidt. *On the Connection between Field-Level Inference and n-point Correlation Functions*. In: (Apr. 2025).

[50] Kazuyuki Akitsu et al. *Cosmology inference with perturbative forward modeling at the field level: a comparison with joint power spectrum and bispectrum analyses*. 2025.

[51] Guido D'Amico et al. *The large scale structure bootstrap: perturbation theory and bias expansion from symmetries*. In: *JCAP* 10 (2021), p. 069.

[52] S. Dodelson and F. Schmidt. *Modern Cosmology*. Academic Press, 2020.

[53] J.B. Hartle. *Gravity: An Introduction to Einstein's General Relativity*. Cambridge University Press, 2021.

[54] F. Bernardeau et al. *Large-scale structure of the Universe and cosmological perturbation theory*. In: *Phys. Rep.* 367.1-3 (Sept. 2002), pp. 1–248.

[55] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. *Large-scale galaxy bias*. In: *Phys. Rep.* 733 (Feb. 2018), pp. 1–193.

[56] C. L. Bennett et al. *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results*. In: *ApJS* 148 (Sept. 2003), pp. 1–27.

[57] D.N. Spergel et al. *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters*. In: *Astrophys.J.Suppl.* 148 (2003), pp. 175–194.

[58] E. Komatsu et al. *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*. In: *Astrophys.J.Suppl.* 192 (2011). 57 pages, 20 figures. Accepted for publication in ApJS. (v2) References added. The SZ section expanded with more analysis. The discrepancy between the KS and X-ray derived profiles has been resolved. (v3) New analysis of the SZ effect on individual clusters added (Section 7.3). The LCDM parameters have been updated using the latest recombination history code (RECFAST version 1.5), p. 18.

[59] G. Hinshaw et al. *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*. In: *Astrophys.J.Suppl.* 208 (2013), p. 19.

[60] N. Aghanim et al. *Planck 2018 results. I. Overview and the cosmological legacy of Planck*. In: *Astron. Astrophys.* 641 (2020), A1.

[61] Planck Collaboration et al. *Planck 2018 results. VI. Cosmological parameters*. In: *A&A* 641, A6 (Sept. 2020), A6.

[62] Jennifer K. Adelman-McCarthy et al. *The Fourth Data Release of the Sloan Digital Sky Survey*. In: *Astrophys. J. Suppl.* 162 (2006), pp. 38–48.

[63] Agne Semenaite et al. *Cosmological implications of the full shape of anisotropic clustering measurements in BOSS and eBOSS*. In: *Mon. Not. Roy. Astron. Soc.* 512.4 (2022), pp. 5657–5670.

[64] Rodrigo S. Gonçalves et al. *Measuring the cosmic homogeneity scale with SDSS-IV DR16 quasars*. In: *Journal of Cosmology and Astroparticle Physics* 2021.03 (Mar. 2021), p. 029.

[65] Jaewon Yoo and Yuki Watanabe. *Theoretical Models of Dark Energy*. In: *International Journal of Modern Physics D* 21.12, 1230002 (Dec. 2012), p. 1230002.

[66] M. Abdul Karim et al. *DESI DR2 results. II. Measurements of baryon acoustic oscillations and cosmological constraints*. In: *Physical Review D* 112.8 (Oct. 2025).

[67] Nils Schöneberg. *The 2024 BBN baryon abundance update*. In: *Journal of Cosmology and Astroparticle Physics* 2024.06 (June 2024), p. 006.

[68] Csaba Balazs et al. "A primer on dark matter". In: *Encyclopedia of Astrophysics (First Edition)*. Ed. by Ilya Mandel. First Edition. Oxford: Elsevier, 2026, pp. 17–32.

[69] G. Gamow. *Expanding Universe and the Origin of Elements*. In: *Physical Review* 70.7-8 (Oct. 1946), pp. 572–573.

[70] A. A. Penzias and R. W. Wilson. *A Measurement of Excess Antenna Temperature at 4080 Mc/s.* In: *ApJ* 142 (July 1965), pp. 419–421.

[71] Jérôme Martin, Christophe Ringeval, and Vincent Vennin. *Encyclopædia Inflationaris.* In: *Physics of the Dark Universe* 5 (Dec. 2014), pp. 75–235.

[72] Evan McDonough. *The cosmological heavy ion collider: Fast thermalization after cosmic inflation.* In: *Physics Letters B* 809 (2020), p. 135755.

[73] Tomohiro Fujita, Kyohei Mukaida, and Tenta Tsuji. *Reheating after axion inflation.* In: *Journal of Cosmology and Astroparticle Physics* 2025.07 (July 2025), p. 002.

[74] Nicola Barbieri et al. *Current Constraints on Cosmological Scenarios with Very Low Reheating Temperatures.* In: *Phys. Rev. Lett.* 135 (18 Oct. 2025), p. 181003.

[75] R. Brandenberger, H. Feldman, and V. Mukhanov. *Classical and Quantum Theory of Perturbations in Inflationary Universe Models.* 1993.

[76] Juan Martin Maldacena. *Non-Gaussian features of primordial fluctuations in single field inflationary models.* In: *JHEP* 0305 (2003), p. 013.

[77] Planck Collaboration et al. *Planck 2018 results - IX. Constraints on primordial non-Gaussianity.* In: *A&A* 641 (2020), A9.

[78] Y. Akrami et al. *Planck 2018 results. X. Constraints on inflation.* In: (July 2018).

[79] G. R. Blumenthal et al. *Formation of galaxies and large-scale structure with cold dark matter.* In: *Nature* 311 (Oct. 1984), pp. 517–525.

[80] Román Scoccimarro. *Cosmological Perturbations: Entering the Nonlinear Regime.* In: *ApJ* 487.1 (Sept. 1997), pp. 1–17.

[81] Peter Coles. *Large-Scale Structure.* In: *Perspectives on Radio Astronomy: Science with Large Antenna Arrays.* Ed. by M. P. van Haarlem. Jan. 2000, p. 53.

[82] Michael Kopp, Kyriakos Vattis, and Constantinos Skordis. *Solving the Vlasov equation in two spatial dimensions with the Schrödinger method.* In: *Phys. Rev. D* 96.12, 123532 (Dec. 2017), p. 123532.

[83] Sebastián Pueblas and Román Scoccimarro. *Generation of vorticity and velocity dispersion by orbit crossing.* In: *Phys. Rev. D* 80.4, 043504 (Aug. 2009), p. 043504.

[84] D. J. Heath. *The growth of density perturbations in zero pressure Friedmann-Lemaître universes.* In: *Monthly Notices of the Royal Astronomical Society* 179.3 (1977), pp. 351–358.

[85] F. Bernardeau et al. *Large-scale structure of the universe and cosmological perturbation theory.* In: *Phys. Rept.* 367 (2002), pp. 1–248.

[86] Matteo Fasiello, Tomohiro Fujita, and Zvonimir Vlah. *Perturbation theory of large scale structure in the* ΛCDM *Universe: Exact time evolution and the two-loop power spectrum*. In: *Phys. Rev. D* 106 (12 Dec. 2022), p. 123504.

[87] Wendy L. Freedman and Barry F. Madore. *The Hubble Constant*. In: *Annual Review of Astronomy and Astrophysics* 48.Volume 48, 2010 (2010), pp. 673–710.

[88] George Efstathiou. *To H0 or not to H0?* In: *Monthly Notices of the Royal Astronomical Society* 505.3 (June 2021), pp. 3866–3872.

[89] Bernard F Schutz. *Determining the Hubble constant from gravitational wave observations*. In: *Nature* 323.6086 (1986), pp. 310–311.

[90] Alvise Raccanelli et al. *Doppler term in the galaxy two-point correlation function: Wide-angle, velocity, Doppler lensing and cosmic acceleration effects*. In: *Physics of the Dark Universe* 19 (2018), pp. 109–123.

[91] N. Kaiser. *Clustering in real space and in redshift space*. In: *Mon.Not.Roy.Astron.Soc.* 227 (1987), pp. 1–27.

[92] M. H. Goroff et al. *Coupling of modes of cosmological mass density fluctuations*. In: *ApJ* 311 (Dec. 1986), pp. 6–14.

[93] Marco Peloso and Massimo Pietroni. *Ward identities and consistency relations for the large scale structure with multiple species*. In: *JCAP* 1404 (2014), p. 011.

[94] A. Kehagias and A. Riotto. *Symmetries and Consistency Relations in the Large Scale Structure of the Universe*. In: *Nucl.Phys.* B873 (2013), pp. 514–529.

[95] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. *Large-Scale Galaxy Bias*. In: (2016).

[96] Marco Marinucci, Kevin Pardede, and Massimo Pietroni. *Bootstrapping Lagrangian perturbation theory for the large scale structure*. In: *JCAP* 10 (2024), p. 051.

[97] Luca Amendola, Massimo Pietroni, and Miguel Quartin. *Fisher matrix for the one-loop galaxy power spectrum: measuring expansion and growth rates without assuming a cosmological model*. In: *JCAP* 11 (2022), p. 023.

[98] Luca Amendola et al. *Improving precision and accuracy in cosmology with model-independent spectrum and bispectrum*. In: *Journal of Cosmology and Astroparticle Physics* 2024.01 (Jan. 2024), p. 001.

[99] Daniel Baumann et al. *Cosmological Non-Linearities as an Effective Fluid*. In: *JCAP* 1207 (2012), p. 051.

[100] Massimo Pietroni et al. *Coarse-Grained Cosmological Perturbation Theory*. In: *JCAP* 1201 (2012), p. 019.

[101] John Joseph M. Carrasco, Mark P. Hertzberg, and Leonardo Senatore. *The Effective Field Theory of Cosmological Large Scale Structures*. In: *JHEP* 1209 (2012), p. 082.

[102] Valentin Assassi, Daniel Baumann, and Fabian Schmidt. *Galaxy bias and primordial non-Gaussianity*. In: *Journal of Cosmology and Astroparticle Physics* 2015.12 (Dec. 2015), p. 043.

[103] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004.

[104] A. Gelman et al. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

[105] Michael Betancourt. *A Conceptual Introduction to Hamiltonian Monte Carlo*. 2018.

[106] Tony Bonnaire et al. *Cosmology with cosmic web environments - I. Real-space power spectra*. In: *Astron. Astrophys.* 661 (2022), A146.

[107] Tony Bonnaire et al. *Cosmology with cosmic web environments - II. Redshift-space auto and cross-power spectra*. In: *Astron. Astrophys.* 674 (2023), A150.

[108] Arka Banerjee and Tom Abel. *Nearest neighbour distributions: New statistical measures for cosmological clustering*. In: *Mon. Not. Roy. Astron. Soc.* 500.4 (2020), pp. 5479–5499.

[109] William R. Coulton, Tom Abel, and Arka Banerjee. *Small-scale signatures of primordial non-Gaussianity in k-Nearest Neighbour cumulative distribution functions*. In: (Sept. 2023).

[110] Oliver Friedrich et al. *Primordial non-Gaussianity without tails – how to measure fNL with the bulk of the density PDF*. In: *Mon. Not. Roy. Astron. Soc.* 498.1 (2020), pp. 464–483.

[111] Cora Uhlemann et al. *Fisher for complements: Extracting cosmology and neutrino mass from the counts-in-cells PDF*. In: *Mon. Not. Roy. Astron. Soc.* 495.4 (2020), pp. 4006–4027.

[112] Julia Stadler, Fabian Schmidt, and Martin Reinecke. *Cosmology inference at the field level from biased tracers in redshift-space*. In: *JCAP* 10 (2023), p. 069.

[113] Justin Alsing et al. *Fast likelihood-free cosmology with neural density estimators and active learning*. In: *Mon. Not. Roy. Astron. Soc.* 488.3 (2019), pp. 4440–4458.

[114] Dezső Ribli et al. *Weak lensing cosmology with convolutional neural networks on noisy data*. In: *Mon. Not. Roy. Astron. Soc.* 490.2 (2019), pp. 1843–1860.

[115] Michelle Ntampaka et al. *A Hybrid Deep Learning Approach to Cosmological Constraints From Galaxy Redshift Surveys*. In: *ApJ* 889.2, 151 (2019), p. 151.

[116] Niall Jeffrey and Benjamin D. Wandelt. *Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks*. In: *34th Conference on Neural Information Processing Systems*. Nov. 2020.

[117] Pablo Villanueva-Domingo and Francisco Villaescusa-Navarro. *Learning Cosmology and Clustering with Cosmic Graphs*. In: *Astrophys. J.* 937.2 (2022), p. 115.

[118] T. Lucas Makinen et al. *The Cosmic Graph: Optimal Information Extraction from Large-Scale Structure using Catalogues*. In: *The Open Journal of Astrophysics* 5.1, 18 (Dec. 2022), p. 18.

[119] Helen Shao et al. *Robust Field-level Inference of Cosmological Parameters with Dark Matter Halos*. In: *Astrophys. J.* 944.1 (2023), p. 27.

[120] Natalí S. M. de Santi et al. *Robust Field-level Likelihood-free Inference with Galaxies*. In: *Astrophys. J.* 952.1 (2023), p. 69.

[121] Andrea Roncoli et al. *Domain Adaptive Graph Neural Networks for Constraining Cosmological Parameters Across Multiple Data Sets*. In: *37th Conference on Neural Information Processing Systems*. Nov. 2023.

[122] ChangHoon Hahn et al. *A forward modeling approach to analyzing galaxy clustering with S<span class=*. In: *Proc. Nat. Acad. Sci.* 120.42 (2023), e2218810120.

[123] Pablo Lemos et al. *SimBIG: Field-level Simulation-Based Inference of Galaxy Clustering*. In: *40th International Conference on Machine Learning*. Oct. 2023.

[124] Hume A. Feldman, Nick Kaiser, and John A. Peacock. *Power spectrum analysis of three-dimensional redshift surveys*. In: *Astrophys. J.* 426 (1994), pp. 23–37.

[125] J. R. Fergusson, M. Liguori, and E. P. S. Shellard. *General CMB and Primordial Bispectrum Estimation I: Mode Expansion, Map-Making and Measures of f_NL*. In: *Phys. Rev. D* 82 (2010), p. 023502.

[126] J. R. Fergusson, M. Liguori, and E. P. S. Shellard. *The CMB Bispectrum*. In: *JCAP* 12 (2012), p. 032.

[127] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard. *Rapid Separable Analysis of Higher Order Correlators in Large Scale Structure*. In: *Phys. Rev. D* 86 (2012), p. 063511.

[128] M. M. Schmittfull, D. M. Regan, and E. P. S. Shellard. *Fast Estimation of Gravitational and Primordial Bispectra in Large Scale Structures*. In: *Phys. Rev. D* 88.6 (2013), p. 063512.

[129] Johnathan Hung, James R. Fergusson, and E. P. S. Shellard. *Advancing the matter bispectrum estimation of large-scale structure: a comparison of dark matter codes*. In: (Feb. 2019).

[130] Joyce Byun et al. *Towards cosmological constraints from the compressed modal bispectrum: a robust comparison of real-space bispectrum estimators*. In: *JCAP* 03 (2021), p. 105.

[131] Gabriel Jung et al. *Quijote-PNG: Quasi-maximum Likelihood Estimation of Primordial Non-Gaussianity in the Nonlinear Dark Matter Density Field*. In: *Astrophys. J.* 940.1 (2022), p. 71.

[132] J. Hartlap, Patrick Simon, and P. Schneider. *Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix*. In: *Astron. Astrophys.* 464 (2007), p. 399.

[133] William R. Coulton et al. *Quijote-PNG: The Information Content of the Halo Power Spectrum and Bispectrum*. In: *Astrophys. J.* 943.2 (2023), p. 178.

[134] Gabriel Jung et al. *Quijote-PNG: Quasi-maximum Likelihood Estimation of Primordial Non-Gaussianity in the Nonlinear Halo Density Field*. In: *Astrophys. J.* 948.2 (2023), p. 135.

[135] William R. Coulton and Benjamin D. Wandelt. *How to estimate Fisher matrices from simulations*. In: (May 2023).

[136] Alan Heavens, Raul Jimenez, and Ofer Lahav. *Massive lossless data compression and multiple parameter estimation from galaxy spectra*. In: *Mon. Not. Roy. Astron. Soc.* 317 (2000), p. 965.

[137] Justin Alsing and Benjamin Wandelt. *Generalized massive optimal data compression*. In: *Mon. Not. Roy. Astron. Soc.* 476.1 (2018), pp. L60–L64.

[138] Francisco Villaescusa-Navarro et al. *The Quijote simulations*. In: *Astrophys. J. Suppl.* 250.1 (2020), p. 2.

[139] William R. Coulton et al. *Quijote-PNG: Simulations of Primordial Non-Gaussianity and the Information Content of the Matter Field Power Spectrum and Bispectrum*. In: *Astrophys. J.* 943.1 (2023), p. 64.

[140] Volker Springel. *The Cosmological simulation code GADGET-2*. In: *Mon.Not.Roy.Astron.Soc.* 364 (2005), pp. 1105–1134.

[141] M. Crocce, S. Pueblas, and R. Scoccimarro. *Transients from Initial Conditions in Cosmological Simulations*. In: *Mon.Not.Roy.Astron.Soc.* 373 (2006), pp. 369–381.

[142] Roman Scoccimarro et al. *Large-scale Bias and Efficient Generation of Initial Conditions for Non-Local Primordial Non-Gaussianity*. In: *Phys. Rev. D* 85 (2012), p. 083002.

[143] M. Davis et al. *The evolution of large-scale structure in a universe dominated by cold dark matter*. In: *ApJ* 292 (May 1985), pp. 371–394.

[144] Neal Dalal et al. *The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects*. In: *Phys. Rev. D* 77 (2008), p. 123514.

[145] Gabriel Jung et al. *Quijote-PNG: Optimizing the summary statistics to measure Primordial non-Gaussianity*. In: *arXiv e-prints*, arXiv:2403.00490 (Mar. 2024), arXiv:2403.00490.

[146] Marco Peloso and Massimo Pietroni. *Galilean invariance and the consistency relation for the nonlinear squeezed bispectrum of large scale structure*. In: *JCAP* 1305 (2013), p. 031.

[147] Paolo Creminelli et al. *Single-Field Consistency Relations of Large Scale Structure*. In: *JCAP* 1312 (2013), p. 025.

[148] Marco Peloso and Massimo Pietroni. *Galilean invariant resummation schemes of cosmological perturbations*. In: *JCAP* 1701.01 (2017), p. 056.

[149] Arhum Ansari et al. *Bootstrapping LSS perturbation theory beyond third order*. In: (Apr. 2025).

[150] Guido D'Amico et al. *The Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure*. In: *JCAP* 05 (2020), p. 005.

[151] Mikhail M. Ivanov, Marko Simonović, and Matias Zaldarriaga. *Cosmological Parameters from the BOSS Galaxy Power Spectrum*. In: *JCAP* 05 (2020), p. 042.

[152] Matteo Peron et al. *Constraining primordial non-Gaussianity from large scale structure with the wavelet scattering transform*. In: *JCAP* 07 (2024), p. 021.

[153] Marco Marinucci et al. *The constraining power of the Marked Power Spectrum: an analytical study*. In: (Nov. 2024).

[154] Elisabeth Krause et al. *A Parameter-Masked Mock Data Challenge for Beyond-Two-Point Galaxy Clustering Statistics*. In: (May 2024).

[155] M. McQuinn. *On the primordial information available to galaxy redshift surveys*. In: *Journal of Cosmology and Astroparticle Physics* 2021.06 (2021), p. 024.

[156] F. Leclercq and A. Heavens. *On the accuracy and precision of correlation functions and field-level inference in cosmology*. In: *Monthly Notices of the Royal Astronomical Society* 506 (Sept. 2021), pp. L85–L90.

[157] N. Porqueres et al. *Lifting weak lensing degeneracies with a field-based likelihood*. In: *Monthly Notices of the Royal Astronomical Society* 509 (Jan. 2022), pp. 3194–3202.

[158] L. Porth et al. *The information content of projected galaxy fields*. In: *Monthly Notices of the Royal Astronomical Society* 518 (Jan. 2023), pp. 3344–3356.

[159] Atsushi Taruya, Takahiro Nishimichi, and Donghui Jeong. *Grid-based calculation for perturbation theory of large-scale structure*. In: *Physical Review D* 98.10 (Nov. 2018), p. 103532.

[160] Atsushi Taruya, Takahiro Nishimichi, and Donghui Jeong. *Covariance of the matter power spectrum including the survey window function effect: N-body simulations versus fifth-order perturbation theory on grids*. In: *Physical Review D* 103.2 (Jan. 2021), p. 023501.

[161] Atsushi Taruya, Takahiro Nishimichi, and Donghui Jeong. *Grid-based calculations of redshift-space matter fluctuations from perturbation theory: UV sensitivity and convergence at the field level*. In: *Phys. Rev. D* 105.10 (2022), p. 103507.

[162] Sean M. Carroll, Stefan Leichenauer, and Jason Pollack. *A Consistent Effective Theory of Long-Wavelength Cosmological Perturbations*. In: (2013).

[163] Henrique Rubira and Fabian Schmidt. *Galaxy bias renormalization group*. In: *JCAP* 01 (2024), p. 031.

[164] Charalampos Nikolis, Henrique Rubira, and Fabian Schmidt. *The renormalization group for large-scale structure: primordial non-Gaussianities*. In: *JCAP* 08 (2024), p. 017.

[165] Shun Saito, Masahiro Takada, and Atsushi Taruya. *Impact of massive neutrinos on nonlinear matter power spectrum*. In: *Phys. Rev. Lett.* 100 (2008), p. 191301.

[166] Shun Saito, Masahiro Takada, and Atsushi Taruya. *Nonlinear power spectrum in the presence of massive neutrinos: perturbation theory approach, galaxy bias and parameter forecasts*. In: *Phys. Rev.* D80 (2009), p. 083528.

[167] Emanuele Castorina et al. *DEMNUni: The clustering of large-scale structures in the presence of massive neutrinos*. In: *JCAP* 07 (2015), p. 043.

[168] Mathias Garny and Petter Taule. *Two-loop power spectrum with full time- and scale-dependence and EFT corrections: impact of massive neutrinos and going beyond EdS*. In: *JCAP* 09 (2022), p. 054.

[169] Lorenzo Piga et al. *Constraints on modified gravity from the BOSS galaxy survey*. In: *JCAP* 04 (2023), p. 038.

[170] Massimo Pietroni. *Flowing with Time: a New Approach to Nonlinear Cosmological Perturbations*. In: *JCAP* 0810 (2008), p. 036.

[171] Yaniv Donath and Leonardo Senatore. *Biased Tracers in Redshift Space in the EFTofLSS with exact time dependence*. In: *JCAP* 10 (2020), p. 039.

[172] Zhenyuan Wang et al. *Perturbation theory remixed: Improved nonlinearity modeling beyond standard perturbation theory*. In: *Physical Review D* 107.10 (May 2023), p. 103534.

[173] Zhenyuan Wang et al. *Perturbation theory remixed. II. Improved modeling of nonlinear bispectrum.* In: *Physical Review D* 110.10 (Nov. 2024), p. 103548.

[174] Steven A. Orszag. *Atmospheric Predictability and Two-Dimensional Turbulence.* In: *Journal of the Atmospheric Sciences* 28.6 (1971), pp. 1074–1076.

[175] Kenneth G. Wilson. *The Renormalization Group: Critical Phenomena and the Kondo Problem.* In: *Rev. Mod. Phys.* 47 (1975), p. 773.

[176] Joseph Polchinski. *Renormalization and Effective Lagrangians.* In: *Nucl. Phys. B* 231 (1984), pp. 269–295.

[177] P. A. R. Ade et al. *Planck 2015 results. XIII. Cosmological parameters.* In: *Astron. Astrophys.* 594 (2016), A13.

[178] Yacine Ali-Haïmoud and Simeon Bird. *An efficient implementation of massive neutrinos in non-linear structure formation simulations.* In: *Monthly Notices of the Royal Astronomical Society* 428.4 (Nov. 2012), pp. 3375–3389.

[179] Roman Scoccimarro. *Transients from initial conditions: a perturbative analysis.* In: *Mon. Not. Roy. Astron. Soc.* 299 (1998), p. 1097.

[180] Diego Blas, Julien Lesgourgues, and Thomas Tram. *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes.* In: *JCAP* 1107 (2011), p. 034.

[181] Masaki Iwasawa et al. *Implementation and performance of FDPS: a framework for developing parallel particle simulation codes.* In: *PASJ* 68.4, 54 (Aug. 2016), p. 54.

[182] Daisuke Namekata et al. *Fortran interface layer of the framework for developing particle simulator FDPS.* In: *PASJ* 70.4, 70 (Aug. 2018), p. 70.

[183] Kohji Yoshikawa and Toshiyuki Fukushige. *PPPM and TreePM Methods on GRAPE Systems for Cosmological N-Body Simulations.* In: *PASJ* 57 (Dec. 2005), pp. 849–860.

[184] Tomoaki Ishiyama, Toshiyuki Fukushige, and Junichiro Makino. *GreeM: Massively Parallel TreePM Code for Large Cosmological N -body Simulations.* In: *PASJ* 61 (Dec. 2009), p. 1319.

[185] Tomoaki Ishiyama, Keigo Nitadori, and Junichiro Makino. *4.45 Pflops Astrophysical N-Body Simulation on K computer – The Gravitational Trillion-Body Problem.* In: *arXiv e-prints*, arXiv:1211.4406 (Nov. 2012), arXiv:1211.4406.

[186] Ataru Tanikawa et al. *N-body simulation for self-gravitating collisional systems with a new SIMD instruction set extension to the x86 architecture, Advanced Vector eXtensions.* In: *New A* 17.2 (Feb. 2012), pp. 82–92.

[187] Ataru Tanikawa et al. *Phantom-GRAPE: Numerical software library to accelerate collisionless N-body simulation with SIMD instruction set on x86 architecture*. In: *New A* 19 (Feb. 2013), pp. 74–88.

[188] Kohji Yoshikawa and Ataru Tanikawa. *Phantom-GRAPE : A Fast Numerical Library to Perform N-body Calculations*. In: *Research Notes of the American Astronomical Society* 2.4, 231 (Dec. 2018), p. 231.

[189] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. Taylor and Francis, 1981.

[190] Tobias Baldauf, Lorenzo Mercolli, and Matias Zaldarriaga. *Effective field theory of large scale structure at two loops: The apparent scale dependence of the speed of sound*. In: *Phys. Rev. D* 92.12 (2015), p. 123007.

[191] Takahiro Nishimichi et al. *Blinded challenge for precision cosmology with large-scale structure: results from effective field theory for the redshift-space galaxy power spectrum*. In: *Phys. Rev. D* 102.12 (2020), p. 123541.

[192] Valentin Assassi et al. *Renormalized Halo Bias*. In: *JCAP* 08 (2014), p. 056.

[193] Andrej Obuljen et al. *Modeling HI at the field level*. In: *Phys. Rev. D* 108.8 (2023), p. 083528.

[194] Julia Stadler et al. *Fast, Accurate and Perturbative Forward Modeling of Galaxy Clustering Part II: Redshift Space*. In: (Nov. 2024).

[195] Roman Scoccimarro. *Redshift-space distortions, pairwise velocities and nonlinearities*. In: *Phys.Rev.* D70 (2004), p. 083007.

[196] Atsushi Taruya, Takahiro Nishimichi, and Shun Saito. *Baryon Acoustic Oscillations in 2D: Modeling Redshift- space Power Spectrum from Perturbation Theory*. In: *Phys. Rev.* D82 (2010), p. 063522.

[197] Alexander Eggemeier et al. *Boosting galaxy clustering analyses with non-perturbative modelling of redshift-space distortions*. In: (Jan. 2025).

[198] R.M. Neal. *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Tech. rep. CRG-TR-93-1. Available online at: http://www.cs.utoronto.ca/∼radford/. Accessed 20 December 2007. Dept. of Computer Science, University of Toronto, 1993.

[199] R.M. Neal. *Slice sampling (with discussion)*. In: *Ann. Stat.* 31 (2003), pp. 705–767.

[200] Radford M. Neal. "MCMC Using Hamiltonian Dynamics". In: *Handbook of Markov Chain Monte Carlo*. Ed. by Steve Brooks et al. CRC Press, 2011, pp. 113–162.

[201] Leonardo Senatore and Matias Zaldarriaga. *Redshift Space Distortions in the Effective Field Theory of Large Scale Structures*. In: (Sept. 2014).

[202] Guido D'Amico et al. *Taming redshift-space distortion effects in the EFTofLSS and its application to data*. In: (Sept. 2021).

[203] Jasche, J. and Lavaux, G. *Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe*. In: *A&A* 625 (2019), A64.

[204] Nicholas Krämer. *A tutorial on automatic differentiation with complex numbers*. 2024.

[205] Jakob Robnik et al. *Microcanonical Hamiltonian Monte Carlo*. 2023.

[206] Jakob Robnik, Reuben Cohn-Gordon, and Uroš Seljak. *Metropolis Adjusted Microcanonical Hamiltonian Monte Carlo*. 2025.

[207] Daniel Levy, Matthew D. Hoffman, and Jascha Sohl-Dickstein. *Generalizing Hamiltonian Monte Carlo with Neural Networks*. 2018.

[208] E. Allys et al. *New interpretable statistics for large-scale structure analysis and generation*. In: *Phys. Rev. D* 102 (10 Nov. 2020), p. 103506.

[209] Giovanni Cabass and Fabian Schmidt. *The EFT likelihood for large-scale structure*. In: *Journal of Cosmology and Astroparticle Physics* 2020.04 (Apr. 2020), p. 042.

[210] Giovanni Cabass and Fabian Schmidt. *The likelihood for LSS: stochasticity of bias coefficients at all orders*. In: *Journal of Cosmology and Astroparticle Physics* 2020.07 (July 2020), p. 051.

[211] T. Lucas Makinen et al. *Hybrid Summary Statistics*. 2025.

[212] Bernd Bischl et al. *Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges*. In: *WIREs Data Mining and Knowledge Discovery* 13.2 (2023), e1484.

[213] Shanel Gauthier et al. *Parametric Scattering Networks*. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022, pp. 5739–5748.

[214] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. *Practical Bayesian optimization of machine learning algorithms*. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'12. Lake Tahoe, Nevada: Curran Associates Inc., 2012, pp. 2951–2959.

[215] Cornelius Rampf, Florian List, and Oliver Hahn. *BullFrog: multi-step perturbation theory as a time integrator for cosmological simulations*. In: *Journal of Cosmology and Astroparticle Physics* 2025.02 (Feb. 2025), p. 020.

Appendix A

Further material on the Wavelet Scattering Transform

A.1 Comparison with marked statistics

In [145], the marked power spectrum and marked bispectrum, computed from re-weighted density fields, were identified as promising observables of PNG on non-linear scales. In figure A.1, their reported 1-σ Fisher bounds, combining standard and marked power spectra and bispectra, are compared with the strongest WST-based constraints reported in figure 4.7. Very similar constraints on PNG are obtained when including marked statistics or the WST, indicating that both probe similar higher-order information beyond the bispectrum.

A.2 Mother wavelets in `Kymatio`

The C_l coefficients appearing in 4.4 are given by

$$C_l = \begin{cases} \frac{1}{(l+1)!!}\,, & l \text{ even} \\ \\ \frac{(2\pi)^{1/2}}{2^{(l+3)/2}\left(\frac{l+1}{2}\right)!}\,, & l \text{ odd}\,. \end{cases} \tag{A.1}$$

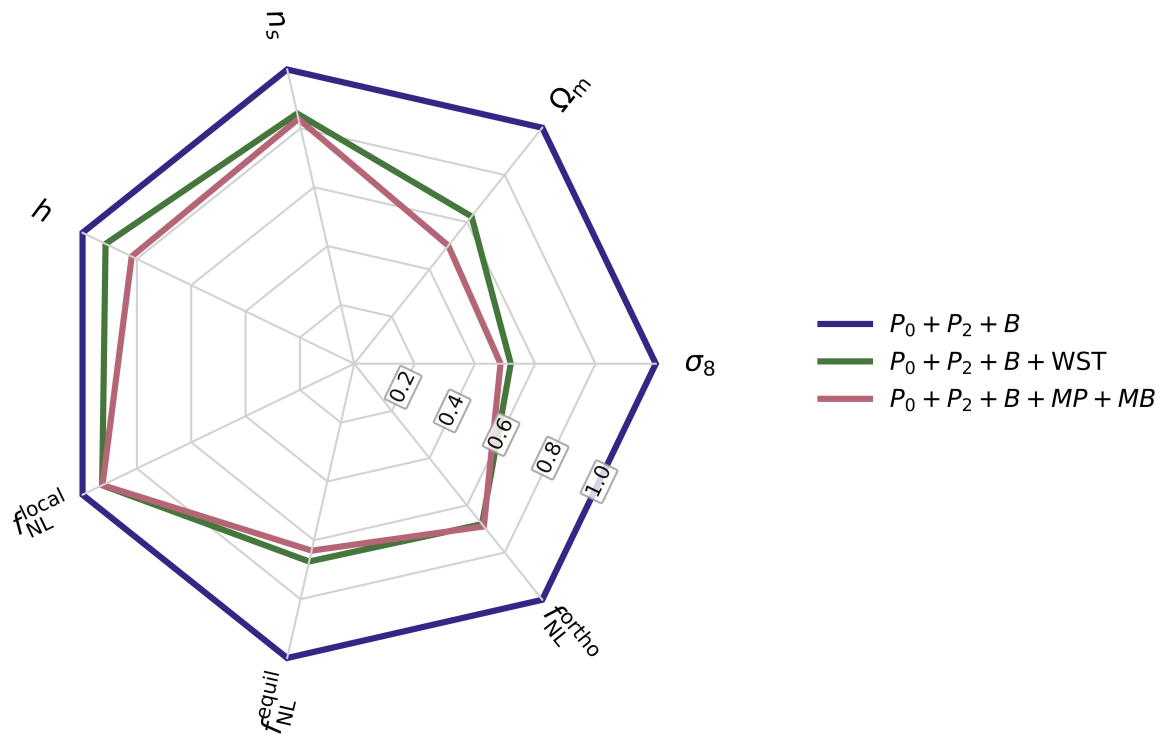

Figure A.1: Comparison of the 1-σ Fisher bounds on cosmological parameters and PNG amplitudes for different combinations of summary statistics (power spectrum, bispectrum, WST ($q = 0.8$), marked power spectrum and marked bispectrum), measured up to $k_{\max} = 0.5\,h\,\mathrm{Mpc}^{-1}$ in the QUIJOTE halo catalogues at $z = 0$.

A.3 Convergence and saturation

Given the simulation-based approach, numerical convergence must be checked. This is tested by comparing the variation in the Cramér–Rao bounds obtained when using a growing number $N_{\mathrm{der}} < 500$ of simulations to compute the numerical derivative in equation (4.19), with the result from the full suite ($N_{\mathrm{der}} = 500$). In figure A.2, the ratio of the two bounds is shown as a function of N_{der} for the WST, highlighting the f_{NL} configurations. If a line asymptotically converges to 1, the bound for that parameter is deemed numerically converged. The bounds converge rapidly, especially for f_{NL} in the halo case. For matter, convergence is slightly slower; however, the absolute difference never exceeds 3%, indicating high numerical stability. For similar tests demonstrating the stability of the P+B analysis, see [131, 133, 134, 139]. It is also informative to examine how the Cramér–Rao bounds improve as more WST coefficients are added. This is shown in figure A.3, where, as in figures 4.8, 4.9, and 4.10, the coefficients are ordered starting from $j = 4$ down to $j = 0$ (see section 4.6). The horizontal line at 1 indicates the corresponding bounds from $P+B$. Focusing on PNG parameters, additional information relative to $P+B$ appears already at intermediate scales (starting around the 40^{th} coefficient) and, for halos, shows clear saturation at small scales, supporting the

robustness of the bounds against spurious small-scale effects.

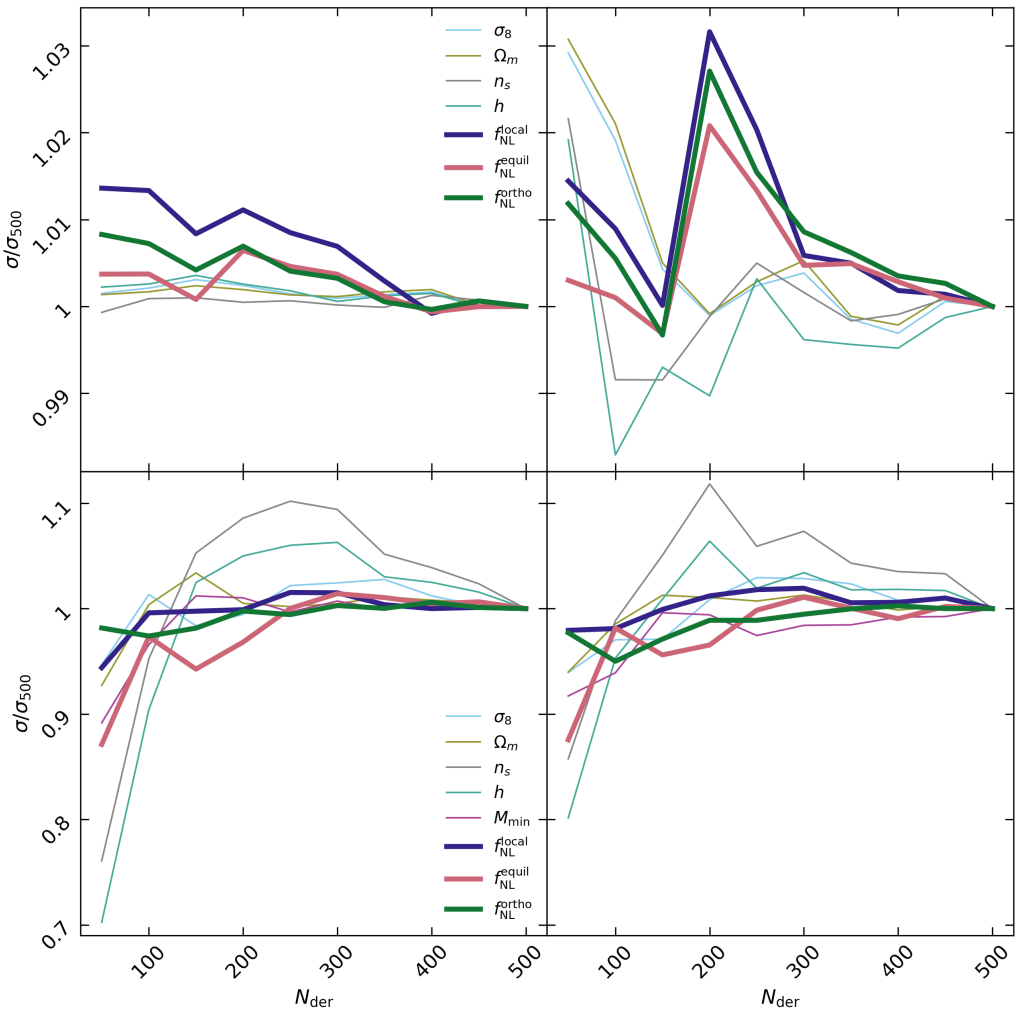

Figure A.2: Convergence of the Cramer-Rao 1-σ bounds on the cosmological parameters as more derivatives on the WST coefficients are used to compute the Fisher matrix. Upper row: matter, lower row: halos. Left: $q = 0.8$, Right: $q = 2$.

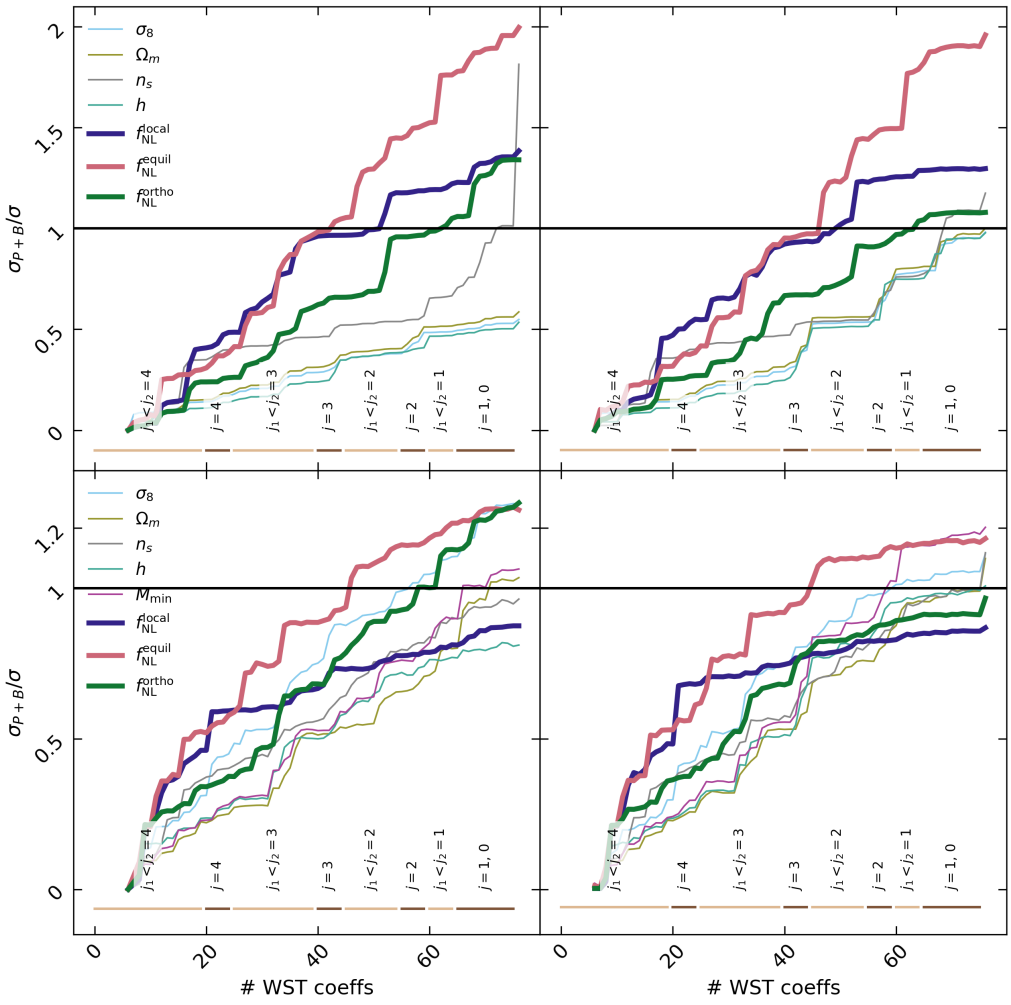

Figure A.3: Change in the Cramer-Rao 1σ bounds on the cosmological parameters obtained from as we add more WST coefficients to the Fisher matrix. The horizontal lines represent the bounds given by $P+B$. Upper row: matter, lower row: halos. Left: $q = 0.8$, Right: $q = 2$.

Appendix B

Proofs and derivations for field-level analysis

B.1 Aliasing and truncation

The Nyquist theorem asserts that a signal can be sampled only up to the Nyquist frequency $k_{\rm Nyq}$. Real signals also contain $k > k_{\rm Nyq}$, and while these frequencies cannot be resolved, their power contributes to the measured signal, producing aliasing. For a 3D field on a cubic, periodic box of volume L^3 with N_g grid subdivisions per side, $k_{\rm Nyq} = \pi N_g/L$.

Aliasing also appears in perturbative fields. Consider the product of two fields in Fourier space:

$$h(\vec{k}) = \int_{\vec{x}} \int_{\vec{q}_1,\vec{q}_2} e^{-i(\vec{k}-\vec{q}_{12})\cdot\vec{x}} f_{\vec{q}_1} g_{\vec{q}_2} \tag{B.1}$$

$$= \int_{\vec{q}_1,\vec{q}_2} \delta_D(\vec{k}-\vec{q}_{12}) f_{\vec{q}_1} g_{\vec{q}_2}. \tag{B.2}$$

If f and g are defined on a finite frequency domain, the sum $\vec{q}_{12}$ may exceed $k_{\rm Nyq}$, generating aliasing in h. A sharp low-pass filter at $\Lambda < k_{\rm Nyq}$ eliminates such contributions:

$$h_\Lambda(\vec{k}) = \int_{\vec{q}_1,\vec{q}_2} \delta_D(\vec{k}-\vec{q}_{12}) f_{\vec{q}_1,\Lambda} g_{\vec{q}_2,\Lambda} \ . \tag{B.3}$$

Residual aliasing when Λ is large is accounted for by adding images shifted by $\pm 2nk_{\rm Nyq}$:

$$h_\Lambda(\vec{k}) = \sum_{n=\mathbf{0}}^{\infty} \int_{\vec{q}_1,\vec{q}_2} \delta_D(\vec{k} \pm 2nk_{\rm Nyq} - \vec{q}_{12}) f_{\vec{q}_1,\Lambda} g_{\vec{q}_2,\Lambda} \ . \tag{B.4}$$

B.2 Second order bootstrap effect at third order

As briefly introduced in section 5.3, the effect of $a_\gamma^{(2)}$ at third order has been found to be negligible, so the results presented in this paper derive from an implementation of bootstrap only at second order. This appendix supplements the main results by providing more details about implementation and numerical proof of the marginal contributions of third order to the estimation of ε_γ.

Continuing on the same line as equations (5.5)–(5.16), at third order in Fourier space

$$\delta^{(3)}(\vec{k},\, a) = \mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2,\vec{q}_3} F_3(\vec{q}_1,\vec{q}_2,\vec{q}_3;\, a)\varphi(\vec{q}_1,a)\varphi(\vec{q}_2,a)\varphi(\vec{q}_3,a)\,, \tag{B.5}$$

$$\theta^{(3)}(\vec{k},\, a) = \mathcal{I}_{\vec{k};\vec{q}_1,\vec{q}_2,\vec{q}_3} G_3(\vec{q}_1,\vec{q}_2,\vec{q}_3;\, a)\varphi(\vec{q}_1,a)\varphi(\vec{q}_2,a)\varphi(\vec{q}_3,a)\,, \tag{B.6}$$

which leads to the definition of the following fields:

$$\begin{aligned}\varphi^{(3)}_{\beta\beta}(\vec{x}) \equiv \frac{1}{2}&\left[\partial_i\varphi(\vec{x})\left(\frac{\partial_i}{\partial^2}\varphi^{(2)}_\beta(\vec{x})\right) + \partial_i\varphi^{(2)}_\beta(\vec{x})\left(\frac{\partial_i}{\partial^2}\varphi(\vec{x})\right)\right] \\ &+ \left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right)\left(\frac{\partial_i\partial_j}{\partial^2}\varphi^{(2)}_\beta(\vec{x})\right)\,,\end{aligned} \tag{B.7}$$

$$\begin{aligned}\varphi^{(3)}_{\beta\gamma}(\vec{x}) \equiv \frac{1}{2}&\left[\partial_i\varphi(\vec{x})\left(\frac{\partial_i}{\partial^2}\varphi^{(2)}_\gamma(\vec{x})\right) + \partial_i\varphi^{(2)}_\gamma(\vec{x})\left(\frac{\partial_i}{\partial^2}\varphi(\vec{x})\right)\right] \\ &+ \left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right)\left(\frac{\partial_i\partial_j}{\partial^2}\varphi^{(2)}_\gamma(\vec{x})\right)\,,\end{aligned} \tag{B.8}$$

$$\varphi^{(3)}_{\gamma\beta}(\vec{x}) \equiv \varphi(\vec{x})\varphi^{(2)}_\beta(\vec{x}) - \left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right)\left(\frac{\partial_i\partial_j}{\partial^2}\varphi^{(2)}_\beta(\vec{x})\right)\,, \tag{B.9}$$

$$\varphi^{(3)}_{\gamma\gamma}(\vec{x}) \equiv \varphi(\vec{x})\varphi^{(2)}_\gamma(\vec{x}) - \left(\frac{\partial_i\partial_j}{\partial^2}\varphi(\vec{x})\right)\left(\frac{\partial_i\partial_j}{\partial^2}\varphi^{(2)}_\gamma(\vec{x})\right)\,, \tag{B.10}$$

$$\varphi^{(3)}_{\alpha_a\gamma}(\vec{x}) \equiv \left(\frac{\partial_i}{\partial^2}\varphi^{(2)}_\gamma(\vec{x})\right)\partial_i\varphi(\vec{x}) - \partial_i\varphi^{(2)}_\gamma(\vec{x})\left(\frac{\partial_i}{\partial^2}\varphi(\vec{x})\right)\,. \tag{B.11}$$

Finally:

$$\delta_3(\vec{x}) = \varphi^{(3)}_{\beta\beta}(\vec{x}) + \left(\frac{1}{2}a^{(3)}_{\gamma a,\mathrm{EdS}} - a^{(3)}_{\gamma b,\mathrm{EdS}} + 1\right)\varphi^{(3)}_{\gamma\beta}(\vec{x}) + \left(\frac{1}{8}a^{(3)}_{\gamma a,\mathrm{EdS}} - \frac{1}{4}a^{(3)}_{\gamma b,\mathrm{EdS}}\right)\varphi^{(3)}_{\alpha_a\gamma}(\vec{x}) + \left(\frac{1}{2}a^{(3)}_{\gamma b,\mathrm{EdS}} - \frac{1}{4}a^{(3)}_{\gamma a,\mathrm{EdS}} + a^{(2)}_{\gamma} - 1\right)\varphi^{(3)}_{\beta\gamma}(\vec{x}) + \left(\frac{1}{4}a^{(3)}_{\gamma a,\mathrm{EdS}} + \frac{1}{2}a^{(3)}_{\gamma b,\mathrm{EdS}}\right)\varphi^{(3)}_{\gamma\gamma}(\vec{x})\,, \tag{B.12}$$

$$\theta_3(\vec{x}) = \varphi^{(3)}_{\beta\beta}(\vec{x}) + \left(\frac{1}{2}d^{(3)}_{\gamma a,\mathrm{EdS}} - d^{(3)}_{\gamma b,\mathrm{EdS}} + 1\right)\varphi^{(3)}_{\gamma\beta}(\vec{x}) + \left(\frac{1}{8}d^{(3)}_{\gamma a,\mathrm{EdS}} - \frac{1}{4}d^{(3)}_{\gamma b,\mathrm{EdS}}\right)\varphi^{(3)}_{\alpha_a\gamma}(\vec{x}) + \left(\frac{1}{2}d^{(3)}_{\gamma b,\mathrm{EdS}} - \frac{1}{4}d^{(3)}_{\gamma a,\mathrm{EdS}} + d^{(2)}_{\gamma} - 1\right)\varphi^{(3)}_{\beta\gamma}(\vec{x}) + \left(\frac{1}{4}d^{(3)}_{\gamma a,\mathrm{EdS}} + \frac{1}{2}d^{(3)}_{\gamma b,\mathrm{EdS}}\right)\varphi^{(3)}_{\gamma\gamma}(\vec{x})\,. \tag{B.13}$$

Setting all the third order bootstrap parameters to their EdS value is equivalent to modifying the third order of `GridSPT` in the following way:

$$\delta^{(3)}(\vec{x}) = \delta^{(3)}_{\mathrm{EdS}}(\vec{x}) + \varepsilon_\gamma a^{(2)}_{\gamma,\mathrm{EdS}}\left(\varphi^{(3)}_{\beta\gamma}(\vec{x}) + \frac{1}{2}\varphi^{(3)}_{\gamma\gamma}(\vec{x})\right)\,. \tag{B.14}$$

Assuming the velocity field to be irrotational, the tensor products appearing in the fields $\varphi^{(3)}_{\beta\gamma}$ and $\varphi^{(3)}_{\gamma\gamma}$ can be simplified to make numerical computation feasible. This leads to the following expressions:

$$\varphi^{(3)}_{\beta\gamma}(\vec{x}) = \frac{1}{2}\partial^2\left(\frac{\partial_i}{\partial^2}\varphi(\vec{x})\frac{\partial_i}{\partial}\varphi^{(2)}_{\gamma}(\vec{x})\right)\,, \tag{B.15}$$

$$\varphi^{(3)}_{\gamma\gamma}(\vec{x}) = \varphi(\vec{x})\varphi^{(2)}_{\gamma}(\vec{x}) - \frac{1}{2}\partial^2\left(\frac{\partial_i}{\partial^2}\varphi(\vec{x})\frac{\partial_i}{\partial}\varphi^{(2)}_{\gamma}(\vec{x})\right) + \frac{1}{2}\partial_i\varphi(\vec{x})\frac{\partial_i}{\partial}\varphi^{(2)}_{\gamma}(\vec{x}) + \partial_i\varphi^{(2)}_{\gamma}(\vec{x})\frac{\partial_i}{\partial^2}\varphi(\vec{x})\,. \tag{B.16}$$

Figure B.1 compares the countours obtained when adding the contribution to third order. while the error bars appear to improve the effect is very marginal, showing how higher orders give minimal contributions to the detection of alternative cosmologies using the bootstrap approach.

B.3 On injecting shot noise in a continuous field

In cosmology, the noise associated with samples of observed galaxies or other cosmological probes follows a Poisson distribution, because galaxies do (excluding small scale

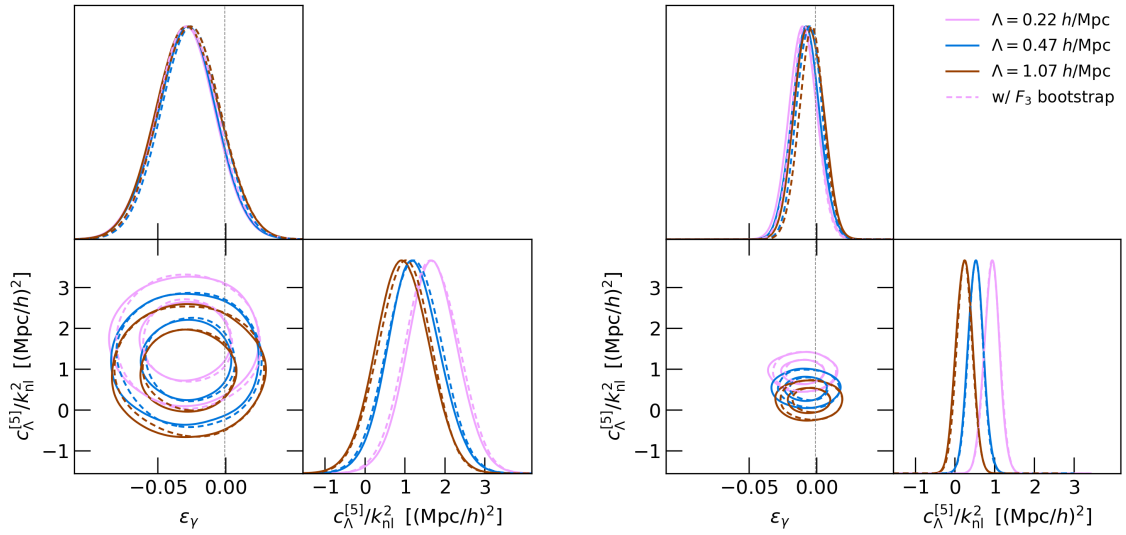

Figure B.1: Comparison between adding the second order bootstrap only up to second order (solid lines) or also to third order (dashed lines). The left (right) column is obtained at order $N = 3$ ($N = 5$).

hydrodynamical physics and other non-gravitational effects) sample the underlying matter density distribution $\rho(\vec{x})$. In principle this type of noise can be mimicked by starting from the number count:

$$\tilde{N}(\vec{x}) \equiv \rho_m(\vec{x})\, \Delta^3 x = (1 + \delta_m(\vec{x}))\, \bar{n}\, \Delta^3 x\,, \tag{B.17}$$

where $\delta_m(\vec{x}) \equiv \frac{\rho_m(\vec{x})}{\langle \rho_m \rangle} - 1$ is the density contrast of dark matter and $\Delta^3 x$ is a finite volume element. This number can be used as the position-dependent average of a Poisson distribution from which to sample an actual number of particles. If the density at $\vec{x}$ is high then more particles are likely to be sampled, and viceversa:

$$N(\vec{x}) \sim P(N \mid \tilde{N}) = \frac{\tilde{N}^N e^{-\tilde{N}}}{N!}\,; \tag{B.18}$$

note how the distribution of N is conditional on $\tilde{N}$, because the latter is also a random variable. Given $N(\vec{x})$, the noisy density contrast $\delta(\vec{x})$ can be obtained by simply rescaling the counts by their mean over every position $\vec{x}$ so that $\langle \delta(\vec{x}) \rangle = 0$. It is known that in galaxies or other tracers, the noise is independent from the signal and additive. To show that this is the case here as well, one just needs to prove the following:

$$\langle \delta_m(\vec{x})\, [\delta(\vec{x}') - \delta_m(\vec{x}')] \rangle \stackrel{?}{=} 0\,. \tag{B.19}$$

Using the definitions of expected value and the axioms of probability (and omitting the dependence on $\vec{x}$ for the sake of clarity):

$$\begin{aligned}
\left\langle \left(\frac{\rho_m}{\langle \rho_m \rangle} - 1 \right) \left(\frac{\rho}{\bar{n}} - 1 \right) \right\rangle &= \left\langle \left(\frac{\rho_m}{\langle \rho_m \rangle} - 1 \right) \left(\frac{\rho_m}{\langle \rho_m \rangle} - 1 \right) \right\rangle \\
\frac{1}{\langle \rho_m \rangle \bar{n}} \frac{1}{V^2} \langle \tilde{N}\, N \rangle - 1 &= \frac{1}{\langle \rho_m \rangle^2} \frac{1}{V^2} \langle \tilde{N}\, \tilde{N} \rangle - 1 \\
\frac{1}{\bar{n}} \iint \mathrm{D}\tilde{N}\, \mathrm{D}N\, \tilde{N} N P(\tilde{N},\, N) &= \frac{1}{\langle \rho_m \rangle} \int \mathrm{D}\tilde{N}\, \tilde{N}^2 P(\tilde{N}) \\
\int \mathrm{D}\tilde{N}\, \tilde{N} \left[\int \mathrm{D}N\, N P(N \mid \tilde{N}) \right] P(\tilde{N}) &= \int \mathrm{D}\tilde{N}\, \tilde{N}^2 P(\tilde{N})\,.
\end{aligned} \tag{B.20}$$

Notice that $\langle \rho_m \rangle = \bar{n}$ follows naturally from equation (B.17). To complete the proof, the integral within square brackets must be equal to $\tilde{N}$:

$$\begin{aligned}
\tilde{N} &= \int \mathrm{D}N\, N P(N \mid \tilde{N}) \\
&= \sum_{N=0}^{\infty} N \frac{\tilde{N}^N e^{-\tilde{N}}}{N!} \\
&= \tilde{N} e^{-\tilde{N}} \sum_{N=1}^{\infty} \frac{\tilde{N}^{N-1}}{(N-1)!}\,,
\end{aligned} \tag{B.21}$$

where the $N = 0$ term vanishes. The sum is the Taylor series of the exponential $e^{\tilde{N}}$, which shows that the noise is indeed additive and uncorrelated when the field is sampled this way. This result has also been tested numerically by sampling ten realizations of the noise and displaying the Fourier analogue of the two terms on the left and right of equation (B.20) in figure B.2. The individual realizations tend to scatter around the theoretical expectation at large scales, while they converge to it when the number of Fourier modes increases, possibly an effect similar to cosmic variance. Plotting the difference between the power spectra of the noisy field and the starting dark matter field, one expects that its value is constant and goes as $1/\bar{n}$ [124]. Indeed figure B.3 shows that this is the case. Again, the same behavior as for the previous figure is seen, however at very small scales the noise shows a systematic increase in power compared to theoretical expectation, likely due to aliasing induced in δ_m when interpolating the N-body dark matter particles onto a regular grid.

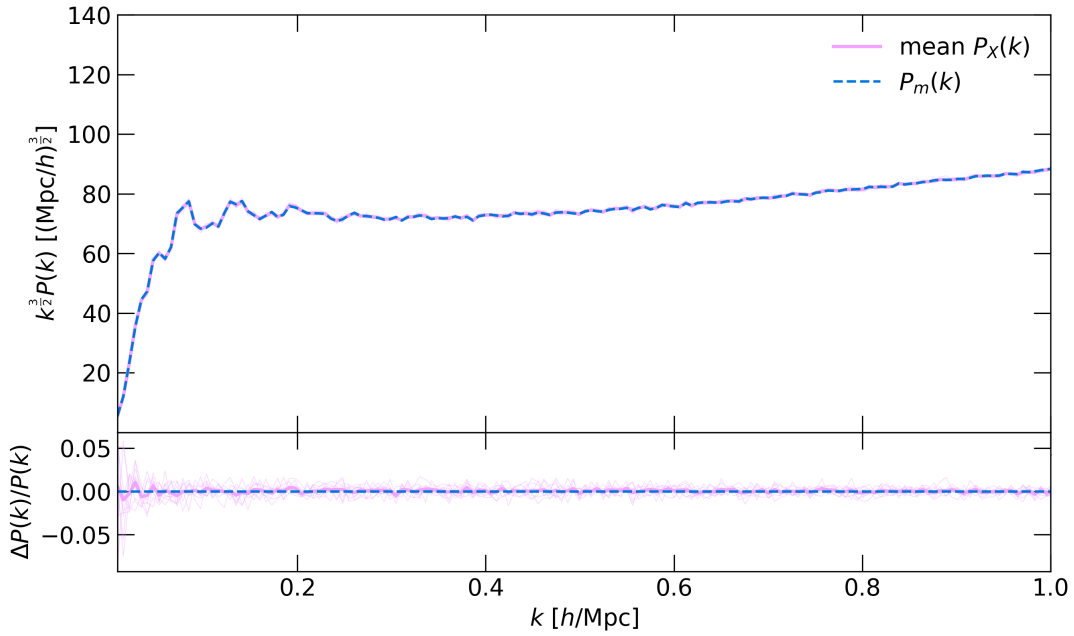

Figure B.2: Comparison between the cross power spectrum $P_X(k) \equiv \frac{\delta_D(k-k')}{(2\pi)^3}\langle\delta_m(\vec{k})\delta(\vec{k}')\rangle$ with and without noise in ten independent realizations of the noise (pink thin lines), and the auto power spectrum of dark matter $P_m(k)$ (blue dashed line). The mean of the ten realization is shown as the pink thick line.

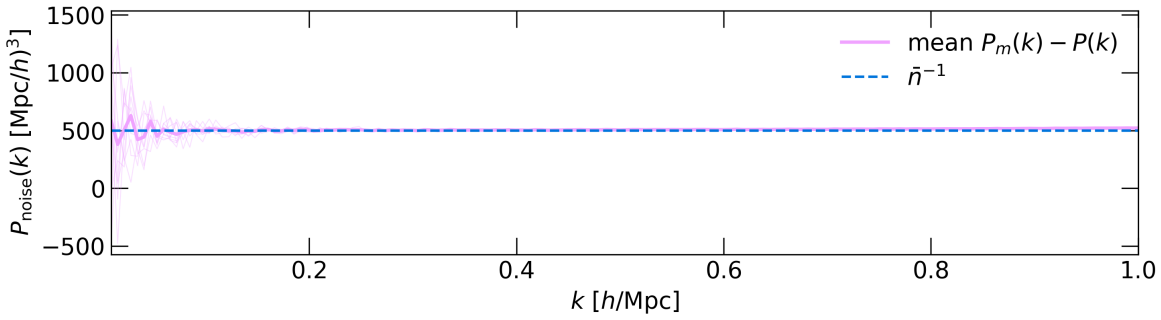

Figure B.3: Comparison between the difference between the auto power spectra of dark matter $P_m(k)$ and the noisy counterpart $P(k)$ in ten independent realizations of the noise (pink thin lines), and the theoretical value of noise in case its distribution is Poissonian (blue dashed line). The mean of the ten realization is shown as the pink thick line.

B.4 Third order estimators in redshift space

What follows respects the notation introduced in chapter 5 and in Taruya, Nishimichi, and Jeong [161]. $\mathcal{F}$ denotes the Fourier transform of a configuration space quantity, the overbar over a quantity denotes the complex conjugate, and the following new fields are introduced:

$$\Phi_1^{(3)} \equiv \mathcal{F}\left[\varphi_\gamma^{(2)} u_z^{(1)}\right] , \tag{B.22}$$

$$\Phi_2^{(3)} \equiv \mathcal{F}\left[\left(\delta^{(1)} + f\nabla_z u_z^{(1)}\right)\frac{\nabla_z}{\nabla^2}\varphi_\gamma^{(2)}\right] , \tag{B.23}$$

$$\Phi_3^{(3)} \equiv \mathcal{F}\left[f u_z^{(1)} \nabla_z \frac{\nabla_z}{\nabla^2}\varphi_\gamma^{(2)}\right] , \tag{B.24}$$

$$\Phi_{a_\gamma}^{(3)} \equiv \varphi_\gamma^{(2)} + ifk_z\Phi_1^{(3)} , \tag{B.25}$$

$$\Phi_{d_\gamma}^{(3)} \equiv f\left(\frac{k_z}{k}\right)^2 \varphi_\gamma^{(2)} + ifk_z\Phi_2^{(3)} + ifk_z\Phi_3^{(3)} . \tag{B.26}$$

The estimators for the quantities introduced in section 6.1 follows. For ε_{a_γ}:

$$\begin{aligned}\bar{\varepsilon}_{a_\gamma,\Lambda} = \frac{2}{a_{\gamma,\mathrm{EdS}}^{(2)}} \sum_{k\le k_{\max}} \frac{1}{\left|\Phi_{a_\gamma}^{(3)}\right|^2} \mathrm{Re}\Big[\overline{\Phi_{a_\gamma}^{(3)}}\Big\{-\delta_{\mathrm{EdS}}^{(2)} - ifk_z\mathcal{F}\left[\delta_{\mathrm{EdS}}^{(2)} u_z^{(1)}\right] \quad + \\ + \ \delta - \left(\delta_S^{[3]} - \delta^{(2)} - ifk_z\mathcal{F}\left[\delta^{(2)}u_z^{(1)}\right]\right)\Big\}\Big] .\end{aligned} \tag{B.27}$$

For ε_{d_γ}:

$$\begin{aligned}\bar{\varepsilon}_{d_\gamma,\Lambda} = \frac{2}{d_{\gamma,\mathrm{EdS}}^{(2)}} \sum_{k\le k_{\max}} \frac{1}{\left|\Phi_{d_\gamma}^{(3)}\right|^2} \mathrm{Re}\Big[\overline{\Phi_{d_\gamma}^{(3)}}\Big\{-ifk_z u_{\mathrm{z,EdS}}^{(2)} \quad - \\ - ifk_z\mathcal{F}\left[\left(\delta^{(1)} + f\nabla_z u_z^{(1)}\right)u_{\mathrm{z,EdS}}^{(2)}\right] \quad - \\ - ifk_z\mathcal{F}\left[f u_z^{(1)}\nabla_z u_{\mathrm{z,EdS}}^{(2)}\right] \quad + \\ + \delta - \left(\delta_S^{[3]} - ifk_z u_z^{(2)} - ifk_z\mathcal{F}\left[\left(\delta^{(1)} + f\nabla_z u_z^{(1)}\right)u_z^{(2)}\right]\right) \quad + \\ + \ ifk_z\mathcal{F}\left[f u_z^{(1)}\nabla_z u_z^{(2)}\right]\Big\}\Big] .\end{aligned} \tag{B.28}$$

For ε_f the result is a 5th order polynomial in $\bar{\varepsilon}_f$, and is given by:

$$\sum_{k\leq k_{\max}}\left[\sum_{i,j=0}^{3}\sum_{l,m=0}^{3}\overline{A}_{ij}B_{lm}\bar{\varepsilon}_{f,\Lambda}^{j+m}+\text{conj.}\right]=0\,, \tag{B.29}$$

where the A and B are lower-triangular matrices with the following nonzero elements:

$$\begin{aligned}
A_{00} &= if_{\Lambda\text{CDM}}k_z\left(u_z^{(1)}+u_z^{(2)}+u_z^{(3)}+\mathcal{F}\left[\delta^{(1)}u_z^{(1)}+\delta^{(1)}u_z^{(2)}+\delta^{(2)}u_z^{(1)}\right]\right)\,,\\
A_{10} = A_{11} &= 2if_{\Lambda\text{CDM}}^2k_z\mathcal{F}\left[u_z^{(1)}\nabla_z u_z^{(1)}+u_z^{(1)}\nabla_z u_z^{(2)}+u_z^{(2)}\nabla_z u_z^{(1)}\right]-f_{\Lambda\text{CDM}}^2k_z^2\mathcal{F}\left[\delta^{(1)}\left(u_z^{(1)}\right)^2\right]\,,\\
2A_{20} = A_{21} &= 2A_{22} = -2f_{\Lambda\text{CDM}}^3k_z^2\mathcal{F}\left[\left(u_z^{(1)}\right)^2\nabla_z u_z^{(1)}\right]\,,\\
B_{00} &= \delta-\delta^{(1)}-\delta^{(2)}-\delta^{(3)}-c_\Lambda^{[3]}k^2\delta^{(1)}\\
B_{10} &= B_{11} = -A_{00}\,,\\
2B_{20} &= B_{21} = 2B_{22} = -A_{10}\,,\\
3B_{30} &= B_{31} = B_{32} = 3B_{33} = -\frac{3}{2}A_{20}\,.
\end{aligned}$$

The estimator for the effective sound speed counterterm $c_\Lambda^{[3]}$ retains the same form as in equation (5.77), except that the model is now the third order redshift space expansion $\delta_S^{[3]}$ introduced by Taruya, Nishimichi, and Jeong [161] and augmented according to section 6.1. Since second order real space quantities mix within redshift space quantities at third order, the above estimators are actually coupled, and need to be solved as a system of equations.

B.5 Primer on automatic differentiation

Consider a complicated, but analytical function $f(x)$. Such a function can always be decomposed into a set of n basic functions $\{f_i\}_{i=1}^n$, each function f_i taking the output of the $(i-1)^{\text{th}}$ as input. In other words, $f(x)$ can be rewritten as a composition

$$f \;=\; f_n(f_{n-1}(\cdots(f_1(x))\cdots)) \;\equiv\; f_n\circ f_{n-1}\circ\cdots\circ f_1\,. \tag{B.30}$$

In general, this is how a computer program works: starting from some input x, it produces an output through a series of simple operations (sums, products, trigonometric functions, matrix multiplications, etc.). Computing the derivative of $f(x)$ is then a matter of applying the chain rule:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}f_n}{\mathrm{d}f_{n-1}} \frac{\mathrm{d}f_{n-1}}{\mathrm{d}f_{n-2}} \cdots \frac{\mathrm{d}f_1}{\mathrm{d}x} \,. \tag{B.31}$$

Since the f_i are simple, their derivatives are straightforward; by chaining these local rules one can differentiate very complex programs.

There are two standard ways to carry out this chaining automatically. Let $f_0 \equiv x$; then (B.31) can be rewritten as

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}f_n}{\mathrm{d}f_n} \frac{\mathrm{d}f_n}{\mathrm{d}f_{n-1}} \frac{\mathrm{d}f_{n-1}}{\mathrm{d}f_{n-2}} \cdots \frac{\mathrm{d}f_1}{\mathrm{d}f_0} \frac{\mathrm{d}f_0}{\mathrm{d}f_0} \,. \tag{B.32}$$

Evaluating from $i = 0$ to $i = n$ yields *forward-mode* automatic differentiation, while evaluating from $i = n$ down to $i = 0$ yields *reverse-mode* automatic differentiation.

Forward mode. Consider the intermediate results of the individual f_i as $v_i \equiv f_i(v_{i-1})$, with $v_0 \equiv f_0 = x$. Then the derivatives of the intermediate results are

$$\dot{v}_0 = 1, \qquad \dot{v}_i = \left.\frac{\mathrm{d}f_i}{\mathrm{d}v_{i-1}}\right|_{v_{i-1}} \dot{v}_{i-1} \,,$$

with $\mathrm{d}f/\mathrm{d}x = \dot{v}_n$. $\dot{v}_0 = 1$ is propagated forward through the chain rule above, where each individual derivative is simple and is evaluated at the intermediate value propagated from the input x. For programs involving multiple inputs and outputs, x can be replaced by a vector of inputs $\vec{x}$ and the chain rule gets propagated in terms of Jacobians $J_{f_i} = \partial f_i / \partial v_{i-1}$ and a seed direction u (generally taken to be one for input parameter of interest and 0 otherwise), then one computes the so called *Jacobian-vector products* (JVP):

$$\dot{\vec{v}}_0 = \vec{u}, \qquad \dot{\vec{v}}_i = J_{f_i}(\vec{v}_{i-1}) \dot{\vec{v}}_{i-1} \quad \Rightarrow \quad \dot{\vec{v}}_n = \frac{\partial f}{\partial x} \vec{u} \,.$$

Reverse mode. In reverse mode automatic differentiation one propagates *adjoints*, the sensitivities of the final output with respect to each intermediate. Similarly to before, the adjoints are

$$\bar{v}_n = 1, \qquad \bar{v}_i \equiv \frac{\mathrm{d}f}{\mathrm{d}v_i} \,,$$

and the desired derivative is $\mathrm{d}f/\mathrm{d}x = \bar{v}_0$. Considering the multidimensional case, the

adjoint of a linear map $A : \mathbb{R}^p \to \mathbb{R}^q$ is defined as $A^* : \mathbb{R}^q \to \mathbb{R}^p$ such that, given two vectors $v \in \mathbb{R}^p$ and $u \in \mathbb{R}^q$, and a bilinear form $\langle \cdot, \cdot \rangle$, the following holds:

$$\langle A^* u,\, v \rangle \;=\; \langle u,\; Av \rangle \,.$$

When $\langle \cdot, \cdot \rangle$ is the dot product, $A^* = A^T$. The Jacobians $J_{f_i}(\vec{v}_{i-1})$ are then treated as linear maps, and the covectors $\bar{v}_i$ are "pulled back" into $\bar{v}_{i-1}$ through their adjoints $J^*_{f_i}(\vec{v}_{i-1}) \equiv J^T_{f_i}(\vec{v}_{i-1})$,

$$\bar{v}_{i-1} = \bar{v}_i J^*_{f_i}(\vec{v}_{i-1}) = \left(J_{f_i}(\vec{v}_{i-1})\, \bar{v}_i^T \right)^T \,.$$

The evaluation of these *vector-Jacobian products* (VJP) in terms of covectors and adjoints flows the derivative backwards along the chain.

Reverse mode needs the forward intermediates v_i to evaluate the Jacobian factors, which means the program has to be evaluated forwards first, and each intermediate is recorded in memory through a "tape"; alternatively checkpoints are used when memory is limited, and intermediates are recomputed in this case. The computational cost is usually a small multiple of the forward run. This extra cost is greatly justified by programs that compute a scalar quantity from many inputs, like functions to evaluate a likelihood or a loss function. In this case the gradient of the output with respect to the n inputs would require n forwards sweeps, or only one reverse sweep. Reverse mode automatic differentiation is thus the method of choice in many applications such as high-dimensional inference or machine learning.