Università degli Studi di Padova

Department of Physics and Astronomy "Galileo Galilei"

Master's Degree in Astrophysics and Cosmology

Master's thesis

Examining Non-Gaussianities in 1+1 Dimensions

Supervisor

Dr. Michele Liguori

Università degli Studi di Padova

Candidate

Lauren Halstead Willett

Student ID: 2071474

Academic year

2023/2024

"It's OK to be one-dimensional, by the way."
— Doja Cat

Abstract

This thesis focuses broadly on how to measure Non-Gaussianity, the deviation of perturbations from statistical Gaussianity. Observationally constraining Primordial Non-Gaussianity (PNG) in galaxy surveys will help elucidate the physics of the early universe, but it is unknown precisely which combination of summary statistics at low redshift (power spectrum, bispectrum, etc.) carries the most information about PNG. Instead of the usual 3 dimensions, this thesis considers a scenario in a single dimension, with gravitationally interacting particles simulated via Lagrangian Perturbation Theory. However, the 'perturbation theory' can be solved exactly in one dimension. In this one-dimensional toy model, it is possible to calculate the Fisher Matrix for cosmological parameters at field-level and compare the result to various summary statistics. This thesis first calculates the power spectrum and the bispectrum from an ensemble of 1000 1-dimensional simulations. Then the field-level Fisher Matrix is calculated for Ω_m and σ_8, omitting different levels of matter streams to determine how much information is contained in the multi-stream regions. Finally, the field-level Fisher matrix is compared with that of the power spectrum at the smallest and largest scales. This exercise demonstrates how working in one dimension has the potential to bring more intuition to the three-dimensional case regarding where cosmological information is contained.

Contents

List of figures

List of abbreviations

1D ..... One-Dimensional

3D ..... Three-Dimensional

ΛCDM ..... Lambda Cold Dark Matter

CDM ..... Cold Dark Matter

LSS ..... Large Scale Structure

LPT ..... Lagrangian Perturbation Theory

SPT ..... Standard Perturbation Theory

CAMB ..... Code for Anisotropies in the Microwave Background*

*https://camb.info/

1 Introduction

Our universe today is a complicated 'cosmic web'. Observations have shown that galaxies are clustered together in dense filaments and beads, interspersed by low-density voids. This was not always the case; in the early Universe shortly after the Big Bang, there were only tiny deviations from complete homogeneity. This is evident when looking at a map of the Cosmic Microwave Background for example; there are hot and cold spots, but nothing more. But these tiny overdensities, subject to gravity in the roughly 13.7 billion years since then, have gradually grown and merged upon themselves into larger structures. Our universe as we now see it – the stars, galaxies, clusters, and so on, are thanks to these initial tiny fluctuations whose original nature we still do not fully understand.

A main goal in cosmology is determining how matter is statistically distributed in our universe, and developing the physical explanation behind it. The manner in which matter is distributed critically informs us about the universe's underlying cosmological parameters and the universe's past history – down to the 'inflationary epoch' a mere $\sim 10^{-36}$ seconds after the Big Bang.

Characterizing the matter distribution is not a trivial task, however. Observational data in cosmology, such as galaxy surveys or cosmic microwave background (CMB) maps, are incredibly high-dimensional. For example, a galaxy survey can provide the 3D positions of millions of galaxies. Directly working with such high-dimensional data is often computationally infeasible. Instead, summary statistics provide a way to extract meaningful information from complex fields in an efficient and interpretable manner. Summary statistics condense the information into lower-dimensional forms, making analysis manageable and allowing for direct comparisons with theoretical predictions from cosmological models.

A difficulty with summary statistics, however, is knowing which ones actually contain the most information about the field. This is simple enough in the case of a Gaussian field, as it is completely described by its power spectrum alone. But for non-Gaussian fields, one needs to turn to other

higher-order statistics to describe the field. In many cases, the best combination of summary statistics is still unknown. This is a particularly pressing issue when searching for weak signals of primordial non-Gaussianity (PNG), possibly imprinted during the inflationary epoch. While PNG has been theorized to possibly exist, it has so far not been detected. Besides improving the precision of observations, an important aspect in the search for PNG is also simply knowing where to look – whether it's the bispectrum, the trispectrum, marked power spectrum, or some other statistic.

This thesis focuses on simulations in 1+1 dimensions: one spatial dimension and one time dimension. While obviously unrepresentative of the real 3D universe, this 1+1 toy model gives the opportunity to study the information content of various summary statistics without the computational overhead of a 3D analysis. The findings from this 1+1 universe can then be used to inform real 3D analyses in the future.

This thesis is organized as follows: In Chapter 1 Sections 1.1 and 1.2, the kinematics of the homogeneous universe are described, and the deviations from this homogeneity are then introduced. Section 1.3 then discusses non-Gaussianities, both primordial and those induced by gravity. In Chapter 2, Perturbation Theory is briefly introduced as a method for approximating large-scale structure formation, and the equations for Lagrangian Perturbation Theory are specified in Section 2.1.1. In Section 2.2 the 1+1 model is then introduced, along with the equations for the exact dynamics in one dimension. Chapters 3 and 4 contain the main analysis and results of this thesis. Chapter 3 focuses on measuring the ensemble power spectrum and bispectrum from the 1+1 simulations. Chapter 4 then focuses on Fisher analyses of the simulations for two cosmological parameters: Ω_m and σ_8. Section 4.1 describes the Fisher analysis at field-level and quantifies the contributions from the multistreams. Section 4.2 then explains how the Fisher matrix is computed for the power spectrum instead, and demonstrates the difference in information content between the power spectrum and the full field. Finally, Chapter 5 contains remarks about directions this 1+1 dimensional work could lead to, such as assessing more summary statistics and more cosmological parameters in the future.

1.1 Background: Cosmology

The universe was once much denser and hotter than it is today. It is expanding (and in fact, accelerating in its expansion, as initially found by [1]). The theory of general relativity developed by Albert Einstein provides the principles upon which our understanding of this expanding universe now lies. This standard cosmological model, often referred to as the ΛCDM model (Lambda Cold Dark Matter), is the prevailing framework for understanding the large-scale structure and evolution of the universe. In ΛCDM, the universe is assumed to be homogeneous and isotropic on large scales, meaning that its properties are similar in all directions and locations when viewed from a sufficiently far distance.

1.1.1 The Components of the Universe

The physical contents of the universe can be broken down into several main components. The first is ordinary 'baryonic' matter, comprised of protons, neutrons, and electrons. This matter can emit and absorb light and makes up the familiar stars, planets, gas, dust, and so on. Baryonic matter is thought to make up only ∼5% of the universe's total energy density. Another ∼27% of the energy density is comprised of 'dark matter', a mysterious form of matter that does not interact with electromagnetic forces [2]. However, it is thought to play a crucial role in the formation of structure in the universe, helping galaxies and clusters hold together. Then there is dark energy; it accounts for around 68% of the universe's total energy density, making it the dominant component in the current era. It is an enigmatic component that appears to be driving the accelerated expansion of the universe but the details of which are still largely unknown. Finally, there is radiation. While photons dominated the universe in its earliest moments, they now contribute only a small fraction of the total energy density. Relativistic particles are often also grouped with radiation as well. For example neutrinos, which are nearly massless and move at nearly the speed of light, behaved similar to photons for most of the universe's history.

1.1.2 The Friedmann Equations

At the core of the standard cosmological model are the Friedmann equations, which describe how the universe's scale factor a, which represents its overall size, changes over time. These equations arise from Einstein's field equations under the assumption of a universe that is both homogeneous and isotropic, a simplification known as the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. This metric describes a universe with a time-dependent scale factor $a(t)$ that governs the expansion or contraction of space. By convention, the scale factor at the time t_0 of present day is $a(t_0) = 1$. Besides its scale factor, a homogeneous and isotropic universe is also characterized by its geometric curvature k. The three possibilities are an open ($k = -1$), closed ($k = 1$), and flat 'Euclidean' ($k = 0$) universe. Current data (from combined Planck CMB observations and LSS observations) suggests that the universe is nearly flat, with $k = 0$ [3]. Matter on large cosmic scales is typically taken to be a perfect fluid, where dissipative processes between particles like viscosity and heat conduction are negligible. The fluid can be characterized by its density ρ and pressure P. The Friedmann equations are two differential equations that connect the rate of expansion of the universe with its contents and curvature. They can be written as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{1.1}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P), \tag{1.2}$$

where $\dot{a}$ is the time derivative of the scale factor $a(t)$ and $\ddot{a}$ is the second derivative. ρ is the total average energy density of the universe, summing all components (both dark and baryonic matter,

radiation, and dark energy) as $\rho = \rho_m + \rho_r + \rho_\Lambda$. The quantity P is the total pressure from all the components. This thesis will consider the case of pressureless CDM, which has $P = 0$.

The Hubble parameter $H(t)$ quantifies how the scale factor changes over time, and is defined as

$$H(t) = \frac{\dot{a}}{a}. \tag{1.3}$$

Often the Hubble parameter is defined in a dimensionless way, as $h = H_0/100$ km s^{-1} Mpc^{-1}. The measured value is $h \approx 0.7$ (though with some currently unreconcilable disagreements; see the discussion on the Hubble Tension in [4]). For the purposes of this thesis, equations will typically use comoving coordinates $\mathbf{r} = a(\tau)\mathbf{x}$ which factor out the underlying universal expansion. The corresponding 'conformal time' τ is defined with respect to the cosmic time t by $d\tau = dt/a(t)$. Using conformal time, one can define a 'conformal Hubble parameter' as $\mathcal{H} = d\ln a/d\tau = aH$.

The critical density ρ_{crit} is the theoretical value marking the boundary between an open ($k = -1$) universe with $\rho < \rho_{\text{crit}}$, and a closed ($k = 1$) universe with $\rho > \rho_{\text{crit}}$. Because our universe has been found to be nearly flat, its average density ρ_0 at present day is thought to be $\rho_0 \approx \rho_{\text{crit},0}$. The density parameter Ω is a dimensionless quantity describing the fraction of ρ_{crit} contributed by each component (eg. matter, radiation, and dark energy) to the total density, such that $\rho/\rho_{crit} = \Omega_{tot} = \Omega_m + \Omega_r + \Omega_\Lambda + ...$, where the total density parameter $\Omega_{tot} \approx 1$ in the case of a flat universe. For example, Ω_m represents the density parameter for matter, both baryonic and dark. It can be defined alongside ρ_{crit} as

$$\rho_m = \Omega_m \rho_{\text{crit}} = \Omega_m \frac{3H^2}{8\pi G} = \Omega_m \frac{3\mathcal{H}^2}{8\pi G a^2}, \tag{1.4}$$

where ρ_m is the average matter density, H is the Hubble parameter, and $\mathcal{H}$ the conformal Hubble parameter. With baryons and dark matter considered together, it is thought that the matter density parameter at present day is $\Omega_{m,0} \approx 0.30$ [2].

1.1.3 The Evolution of the Universe

The universe is thought to have gone through several main phases in its history following the Big Bang. The first was inflation, a brief early period of accelerated expansion driven by an extraordinarily high energy (thought to be $\sim 10^{15} - 10^{16}$ GeV) which is still largely not understood. This mysterious phase provides solutions to a number of problems associated with the Hot Big Bang model, such as the horizon problem (the question of why the universe is homogeneous on the largest scales) and the flatness problem (the conundrum of explaining why our universe appears so nearly flat). In its simplest form, inflation was driven by the potential energy of a scalar field φ called the 'inflaton'. The potential energy was nearly constant, allowing it to play the role of a cosmological constant driving the exponential expansion of the universe. There are numerous alternatives to this model which will be briefly discussed in Section 1.3. Regardless, inflation flattened and smoothed out the universe, but also was characterized by tiny quantum fluctuations in the inflaton field. These tiny fluctuations are then thought to have been the seeds

which gradually grew into large-scale structures of the universe today [5].

After inflation ended, the universe expansion slowed down and transitioned into the radiation-dominated era, where the energy density was mainly comprised of radiation ($\rho_r \propto a^{-4}$). Crucially near the end of this phase was the release of the Cosmic Microwave Background Radiation (CMB). At this time, the universe had cooled enough (to about 3000 K) for protons and electrons to combine into neutral hydrogen atoms in a process called *recombination*. This led to the emittance of light in a near-perfect blackbody, but including the small temperature fluctuations first seeded by the inflaton field. The CMB serves as a footprint of the early universe, which we have now observed and studied through missions such as WMAP and Planck [6, 7, 2].

As the universe expanded and cooled further, the energy density of matter ($\rho_m \propto a^{-3}$) became more significant than radiation, leading to the matter-dominated era. Mainly during this phase the gravitational pull of dark matter and baryonic matter drove the formation of galaxies and other cosmic structures from the small perturbations. This process (elaborated further in Section 1.2) resulted in a highly irregular distribution of matter, which is studied today through surveys of galaxies and other tracers. Only in fairly recent times did the universe enter the dark energy-dominated era, where an enigmatic constant energy component, called dark energy ($\rho_\Lambda \propto a^0$), is now driving the accelerated expansion of the universe and slowing down structure formation. These phases highlight the dynamic nature of the universe's expansion over time.

1.2 Large Scale Structure

The standard Friedmann-Robertson-Walker model correctly depicts the overall evolution and expansion of the universe from a radiation-dominated initial state, to matter dominated, and then finally through a dark-energy dominated phase. However, this spatially homogeneous and isotropic model is not sufficient to mimic the true distribution of matter and energy today, which is only effectively homogeneous on the largest scales. In the early post-inflation universe, it is thought that the density was nearly uniform, except for some low-amplitude fluctuations on top of an otherwise homogeneous and isotropic background. Present surveys of the large-scale distribution of matter – traced by galaxies for example – have instead painted a picture of a complex 'cosmic web'. This cosmic web is thanks to gravity, which caused those initially low-amplitude density fluctuations left over from the inflationary era to gradually coalesce and form the large-scale structures we observe today.

Gravitational growth is opposed by two major phenomena. The first is the background expansion of the universe. As the universe expands, the physical distance between matter particle grows, opposing the force of gravity attempting to bring the particles together. When dark energy comes to dominate in the late universe, and the universe expansion accelerates, this effect becomes even more potent. The universe therefore experiences an ongoing competition between the underlying expansion of space pulling matter apart, and the force of gravity pulling matter together. The second phenomenon is the pressure on baryons and photons. In the particular case of baryons and photons, their non-zero pressure inhibits gravitational growth. This is because the pressure

is higher in the overdense regions, which counteracts further growth of the overdensity.

This thesis focuses on simulations of collisionless cold dark matter (CDM) density perturbations. Theoretically, CDM is pressureless and does not interact with the other components of the universe through any avenue besides gravity. Therefore, the evolution of the CDM overdensities follows directly from the evolution in the gravitational potential and the background expansion of the universe. Besides this simplicity, the main reason for focusing on CDM is that dark matter makes up most of the matter density in the universe. In fact, at late times, the baryon distribution closely follows that of the dark matter distribution. Therefore, characterizing the distribution of CDM today is a powerful first step in predicting the power spectrum of visible baryonic matter tracers, such as galaxies.

The parameter σ_8 is an important cosmological quantity that characterizes the clustering of matter on large scales, in particular on the scale of 8 megaparsecs (Mpc). Specifically, σ_8 measures the root-mean-square (RMS) fluctuation of the matter density field within spheres of radius 8 Mpc in the present-day universe. This essentially quantifies how clumped the distribution of matter is on large scales. It is formally defined as

$$\sigma_8 = \sqrt{\left\langle \left(\frac{\delta M}{M} \right)^2 \right\rangle_{R=8h^{-1}\mathrm{Mpc}}}, \tag{1.5}$$

where $\delta M/M$ represents the fractional mass fluctuation within the 8 Mpc spheres, and the brackets $\langle ... \rangle$ represent the ensemble average over many such spheres. The value for σ_8 has been measured to be approximately 0.80 [2].

1.3 Non-Gaussianity

A Gaussian field is a type of random field where the statistical distribution of values at different points is entirely characterized by a Gaussian distribution. Gaussian fields often emerge as the result of the central limit theorem, which states that the sum of a large number of independent random variables will tend to be Gaussian, regardless of the individual distributions [8].

Under the standard cosmological model, the initial condition immediately after inflation was a Gaussian, adiabatic, and nearly scale-invariant density field borne from quantum fluctuations. This Gaussian field, subsequently processed by gravitational evolution, became non-Gaussian. This is due to the nonlinear nature of gravitational interactions, which introduces coupling between initially independent modes. Eventually, structures such as galaxies, clusters, and filaments began to form, leading to strong deviations from Gaussianity. The matter density field today is highly non-Gaussian, marked by prominent features like voids and clusters that cannot be adequately described by a Gaussian distribution.

1.3.1 Primordial Non-Gaussianity

Under the current standard model of cosmology, the *primordial* adiabatic perturbations are nearly perfectly Gaussian. This is because the most basic models of inflation assume a single scalar free field φ, in which the action is quadratic and there is no mechanism to mix perturbation modes (such as through interactions). As a result, each mode is an independent stochastic Gaussian process [5]. Moreover, the scalar perturbations are thought to be nearly scale-invariant, with a spectral index $n_s \approx 1$ (It has in fact been measured as $n_s = 0.9649 \pm 0.0042$ at 68% CL by [3]).

However, there exists great interest among cosmologists in looking for *initial* deviations from Gaussianity, before the effects of gravitational evolution. Such a scenario is called Primordial Non-Gaussianity (PNG). A multitude of inflationary models have been investigated for the PNG signal they would produce (eg. [9, 10, 11, 12]). However, PNG has not been observed as of yet. Strongly non-Gaussian initial conditions have already been ruled out by CMB and LSS observations (see, for example, the upper limits set by the Planck experiment in [13], which place the upper bound of PNG at $\lesssim 0.1\%$ the amplitude of the Gaussian field component). Some small amount of non-Gaussianity could nonetheless be possible below currently-established upper limits [14].

Any such deviation from Gaussianity can supply knowledge about the early universe, and in particular on the mechanism driving inflation. Different inflationary models produce different shapes and amplitudes of PNG. Measuring the characteristics of PNG would therefore allow the possibility of constraining the interactions, symmetries, and degrees of freedom governing inflation. The simplest 'slow roll' inflation models, involving a single scalar field φ, predict a tiny PNG signature too small to presently observe. For this reason, any observed PNG signal would be what is often described as the 'smoking gun' of non-standard inflation models [5].

We have entered a new era of 'precision cosmology' where measurements have needed to get more and more precise to refine our cosmological model. The latest Planck observations of the CMB have already put tight constraints on PNG, and the CMB temperature is reaching its limitation in being observed any more precisely (due to, for example, Silk damping of smaller-scale modes). Estimates have suggested that it is likely that further CMB observations may improve constraints on PNG by at most a factor of ~ 2 [15, 16]. Therefore in the quest to further constrain non-Gaussianity, LSS is an enticing alternative. Unlike the CMB which is a 2D projection on the sky, galaxy clustering can be mapped out in three dimensions. According to [17], for the case of weak PNG, the majority of non-Gaussian information is initially contained within the bispectrum of primordial cosmological perturbations. A 3D map of structure, such as a galaxy survey, contains more bispectrum modes than a two-dimensional CMB map.

A difficulty with using LSS, however, is that it is observed at low redshifts. This makes PNG harder to parse out, especially at smaller scales. At low redshifts, the PNG signal is expected to have an extraordinarily small amplitude (by a factor of ~ 1000) compared to the non-Gaussianity introduced by the nonlinear gravitational evolution of the perturbations [17]. To extract the dilute PNG signal, one must therefore differentiate it from the much larger non-primordial signal; this is not a trivial task [18].

Another problem is how to best capture the information about PNG within summary statistics. It is computationally demanding to examine large 3-dimensional surveys (eg. galaxy surveys) at field-level; instead, the information typically needs to be condensed into summary statistics, such as the bispectrum, that can then be compared to predictions from cosmological models. This comes with the advantage that summary statistics can be easier to physically interpret. That said, it is still unclear at low redshifts what is the optimal combination of summary statistics to best describe various types of PNG. The goal is to find the summary statistics which are informative enough compared to field-level, without the computational demand of field-level inference. [18].

In this thesis, two key summary statistics, the power spectrum and bispectrum, are calculated in Chapter 3 from 1-dimensional simulations. Then in Chapter 4, the information content from the power spectrum is compared to field level for the cosmological parameters Ω_m and σ_8 using their Fisher matrices. The main goal of this exercise is to examine how much information is lost between a field-level inference and compression into a power spectrum alone. There are many promising avenues for further exploration beyond this thesis; the most obvious being a computation of the Fisher matrix for the bispectrum as well. Possible future directions will be discussed further in Section 5.1.

2

Perturbation Theory and the Universe in 1+1 Dimensions

2.1 Perturbation Theory

The most straightforward way to simulate gravitational evolution of fluctuations is to make an N-body simulation. However, this can become very computationally expensive when dealing with a large number of particles or when exploring a large cosmological parameter space. Therefore, to attempt to describe observations, cosmologists often use 'perturbation theory'. This thesis does not review all perturbation theory in great detail, but the interested reader can consult reviews such as [19]. Perturbation theory, as its name suggests, assumes that any deviations from homogeneity can be treated as small perturbations. This allows their gravitational evolution to be treated systematically as a series expansion around the homogeneous background solution.

This thesis will focus on perturbation theory in the Newtonian limit, rather than the full general relativistic perturbation theory. This is because CDM particles are non-relativistic, so at scales below the Hubble radius the equations of motion are approximately the same as Newtonian gravity [19].

The Newtonian gravitational potential Φ at some position $\mathbf{r}$ is created by the local matter density $\rho_m(\mathbf{r})$,

$$\Phi(\mathbf{r}) = G \int d^3\mathbf{r}' \, \frac{\rho_m(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|}, \tag{2.1}$$

where G is the gravitational constant. For a large number of particles, one can assume to have a smooth gravitational potential which causes the particles (of a velocity $\mathbf{v}$) to accelerate:

$$\frac{d\mathbf{v}}{dt} = -\frac{\partial \Phi}{\partial \mathbf{r}}. \tag{2.2}$$

One can also define the *cosmological* gravitational potential ϕ, which factors out the underlying expansion of the universe such that ϕ depends only on the density fluctuations [19]. This equation (and the ones hereafter) are expressed with the comoving coordinate $\mathbf{x}$ and conformal time τ. This ϕ relates to Φ as

$$\Phi(\mathbf{x}, \tau) \equiv -\frac{1}{2}\frac{\partial \mathcal{H}}{\partial \tau} x^2 + \phi(\mathbf{x}, \tau). \tag{2.3}$$

The potential ϕ obeys the Poisson equation:

$$\nabla^2 \phi(\mathbf{x}, \tau) = 4\pi G \bar{\rho}_m(\tau) a^2(\tau) \delta(\mathbf{x}, \tau) = \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(\mathbf{x}, \tau). \tag{2.4}$$

Perturbations are typically represented by the density contrast $\delta(\mathbf{x}, \tau)$, which can be written from the local matter density $\rho_m(\mathbf{x}, \tau)$ and mean matter density $\bar{\rho}_m(\tau)$ as

$$\rho_m(\mathbf{x}, \tau) \equiv \bar{\rho}_m(\tau)\left[1 + \delta(\mathbf{x}, \tau)\right], \tag{2.5}$$

where here the time dependence τ is explicitly included, to emphasize that the perturbations evolve with time. A perturbation is considered 'small' as long as $|\delta| \ll 1$. This tends to generally hold true at early times and on large scales, when the growth of structures has not yet entered the 'nonlinear' regime.

The peculiar velocity $\mathbf{u}(\mathbf{x}, \tau)$ of a particle is defined as its deviation from the Hubble flow:

$$\mathbf{v}(\mathbf{x}, \tau) \equiv \mathcal{H}\mathbf{x} + \mathbf{u}(\mathbf{x}, \tau). \tag{2.6}$$

The main cornerstones of Perturbation Theory are the continuity and Euler equations. The continuity equation represents the conservation of mass, and the Euler equation the conservation of momentum. They each respectively can be written in terms of δ and $\mathbf{u}$ as

$$\frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} + \nabla \cdot \left\{\left[1 + \delta(\mathbf{x}, \tau)\right]\mathbf{u}(\mathbf{x}, \tau)\right\} = 0, \tag{2.7}$$

$$\frac{\partial \mathbf{u}(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H}(\tau)\,\mathbf{u}(\mathbf{x}, \tau) + \mathbf{u}(\mathbf{x}, \tau) \cdot \nabla \mathbf{u}(\mathbf{x}, \tau) = -\nabla \phi(\mathbf{x}, \tau) - \frac{1}{\rho_m}\nabla_j(\rho_m\, \sigma_{ij}). \tag{2.8}$$

The $\sigma_{ij}(\mathbf{x}, \tau)$ in Equation 2.8 is the stress tensor of the cosmological fluid. The continuity and Euler equations are actually only the first two equations of the full Boltzmann hierarchy, but the hierarchy can be closed if an assumption is made for $\sigma_{ij}(\mathbf{x}, \tau)$. The stress tensor describes how much the motion of the particles deviates from a singular stream. If one assumes there is a single stream, then $\sigma_{ij} \approx 0$. Assuming a single stream is a decent approximation in the early phases of gravitational evolution and on large scales. However, $\sigma_{ij} \approx 0$ ceases to be true when there are

multiple streams; this phenomenon of 'shell crossing' is discussed further in Section 2.2.2. For many applications of perturbation theory though, it is assumed that the cosmological fluid is a pressureless, perfect fluid with $\sigma_{ij} \approx 0$, such as for CDM. Moreover, it is often assumed the velocity field has no vorticity, and can be completely described by its divergence $\theta(\mathbf{x}, \tau) = \nabla \cdot \mathbf{u}(\mathbf{x}, \tau)$. This is because if $\sigma_{ij} \approx 0$ and there is no primordial vorticity (as is such for a pressureless perfect fluid), then the vorticity will continue to be zero at later times [19].

At the largest scales, the density fluctuations in the universe are relatively small enough ($\delta \ll 1$, $\theta \ll 1$) that Equations 2.7 and 2.8 can be truncated to only the terms which depend *linearly* on fluctuations. This is the basis of Linear Perturbation Theory, and results in a density contrast that is proportional with the initial field: $\delta(\mathbf{x}, \tau) = D_1(\tau)\delta(\mathbf{x}, 0)$. One can also expand δ and θ about their linear solutions into higher orders, as is done in Eulerian Non-linear Perturbation Theory.

There are a number of advanced approaches aiming to extend the validity of perturbation theories beyond shell crossing and into the multistreaming regime. For example, the Effective Field Theory of Large Scale Structure uses smoothed density and velocity fields for which perturbation theory is still valid. The theory then accounts for the small-scale, nonlinear physics that cannot be directly modeled, by simply enclosing it into effective terms. These new terms absorb the contributions from small scales and their values must be informed by outside information such as separate N-body simulations [20].

2.1.1 Lagrangian Perturbation Theory

The Lagrangian framework is essentially a reformulation of non-linear perturbation theory, where rather than solving for the density and velocity fields in a fixed coordinate system, one instead follows the trajectories of the particles or fluid elements [19]. This can be a particularly useful approach for understanding the growth of cosmic structure into the mildly nonlinear regime, because it can describe large-scale flows and bulk motions without requiring high-order corrections. In Lagrangian Perturbation Theory (LPT), one solves for the displacement field $\Psi(\mathbf{q}, \tau)$, which is a mapping from the initial position $\mathbf{q}$ to the final position $\mathbf{x}(\tau)$ of each particle (in Eulerian comoving coordinates at some conformal time τ). The particle position can thus be expressed as

$$\mathbf{x}(\tau) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, \tau), \tag{2.9}$$

where at an initial time τ_0, $\mathbf{\Psi}(q, \tau_0) = 0$ [19].

Mass is conserved as particles move from their initial positions $\mathbf{q}$ to the final Eulerian positions $\mathbf{x}$; the mass contained within a volume element in Eulerian space is thus equal to the mass contained in the corresponding volume element in Lagrangian space. The density field must therefore obey the continuity relation

$$\bar{\rho}_m \left(1 + \delta(\mathbf{x})\right) d^3\mathbf{x} = \bar{\rho}_m \, d^3\mathbf{q}. \tag{2.10}$$

The Euler equation is linearized and then expressed in Eulerian coordinates as

$$\frac{d^2\mathbf{x}}{d\tau^2} + \mathcal{H}(\tau)\frac{d\mathbf{x}}{d\tau} = -\nabla_x \phi, \tag{2.11}$$

where the time-dependence of $\mathbf{x}$ has been left implicit for readability. The equation describes the motion of particles in a cosmological gravitational potential ϕ, where ∇_x is the gradient operator in Eulerian comoving coordinates $\mathbf{x}$.

Given the relationship between $\mathbf{x}$ and $\mathbf{\Psi}$ in Equation 2.9, Equation 2.11 can alternatively be written in terms of $\mathbf{\Psi}$ as

$$\frac{d^2\mathbf{\Psi}(\mathbf{q})}{d\tau^2} + \mathcal{H}\frac{d\mathbf{\Psi}(\mathbf{q})}{d\tau} = -\nabla_x \phi(\mathbf{q} + \mathbf{\Psi}(\mathbf{q})). \tag{2.12}$$

If one takes the divergence of Equation 2.12 and substitutes in Equation 2.4 for $\nabla_x^2\phi$, Equation 2.11 can then be expressed as

$$J(\mathbf{q},\tau)\,\nabla_x \cdot \left[\frac{d^2\mathbf{\Psi}}{d\tau^2} + \mathcal{H}(\tau)\frac{d\mathbf{\Psi}}{d\tau}\right] = \frac{3}{2}\Omega_m \mathcal{H}^2 (J-1), \tag{2.13}$$

where $J(\mathbf{q},\tau)$ is the Jacobian that encodes the transformation between Eulerian and Lagrangian space. The right hand side has made use of the continuity relation in Equation 2.10, which allows the Jacobian $J(\mathbf{q},\tau)$ to be expressed with

$$1 + \delta(\mathbf{x}) = \frac{1}{\mathrm{Det}\left(\delta_{ij} + \Psi_{i,j}\right)} \equiv \frac{1}{J(\mathbf{q},\tau)}, \tag{2.14}$$

where $\Psi_{i,j} = \partial\Psi_i/\partial\mathbf{q}_j$ [19]. The gradient ∇_x can also be converted into Lagrangian coordinates $\nabla_q = \partial/\partial\mathbf{q}$ by noting that $\nabla_{x_i} = \left(\delta_{ij} + \Psi_{i,j}\right)^{-1}\nabla_{q_j}$.

LPT aims to perturbatively solve the differential equation, Equation 2.12, for the displacement field $\mathbf{\Psi}$. LPT treats the displacement field as a perturbative series. The zeroth-order solution represents an unperturbed universe, the first-order solution corresponds to the Zel'dovich approximation (see Section 2.2.1), and higher-order terms provide additional corrections to account for nonlinearities in the matter distribution [21]:

$$\mathbf{\Psi}(\mathbf{q},\tau) = \mathbf{\Psi}^{(1)}(\mathbf{q},\tau) + \mathbf{\Psi}^{(2)}(\mathbf{q},\tau) + \mathbf{\Psi}^{(3)}(\mathbf{q},\tau) + \cdots \tag{2.15}$$

2.2 The 1+1 Model

Cosmological perturbation theory is typically solved in three dimensions, but this thesis considers instead a toy model with only one spatial dimension and one time dimension (hence the nickname '1+1 model'). This approach has been used in the past by [22] to compare standard Eulerian perturbation theory with Lagrangian perturbation theories and also to test 'effective field' theories of Large Scale Structure. The 1+1 model was also used in [23] to explore the breakdown

of perturbation theory at shell crossing, and to consider the presence of an asymptotic attractor solution deep in the multistreaming regime.

Physically, the 1+1 model requires all modes to be along a single direction. Rather than picturing particles, the masses can best be envisioned as thin, infinite sheets of matter which are restricted to travel along one axis. These sheets gravitationally interact with one another, but the gravitational force is independent of the distance from each sheet. Therefore, how much gravitational force a single sheet experiences depends only on the number of other sheets to the left and right of it. Just as in real-world cosmology, the sheets travel within a Hubble flow that works against the gravitational force. Perpendicular to the axis of movement, the Hubble flow also stretches the sheets, causing their surface density to decrease as a^{-2}. There are some similarities between this 1D toy model and 3D cosmology. In the limit of two modes with wavenumbers of very different magnitudes (ie. one very large and one very small), their coupling is the same in 1D and in 3D, up to a geometric factor. This 1D situation does go against the usual statistical isotropy assumed in cosmology. However, it is nonetheless still subject to the same dynamical equations as in 3D, such as the continuity and Euler equations with matter treated as a pressureless fluid. The one assumption lacking in 1D is that of the non-linear velocity being curl-free [22].

There are a number of advantages in using the 1+1 model. The first is that in 1D, the Zel'dovich approximation turns out to be exact up to shell crossing, unlike in 3D. This can make it easier to accurately isolate multistreams and consider their effects. Moreover, even post-shell crossing, the exact dynamics can be recovered without doing a full N-body simulation [23]. This aspect will be discussed more in Section 2.2.1. Simulating in one dimension also simply reduces the amount of required computational power compared to 3D. All of the integrals d^3k within calculations become dk integrals, which allows for a larger range in wavenumber and higher orders of perturbation theory to be calculated with the same computational resources [22].

Finally, it is simpler to assess 1D simulations at field-level compared to their 3D counterpart. This is because it is much more computationally feasible in 1D to calculate the full field-level likelihood than it is in 3D. In Chapter 4, the field-level Fisher matrix is computed, something which would be difficult to accomplish in 3D. The field-level result can then be compared to Fisher matrices from various summary statistics, such as for the power spectrum and the bispectrum. This exercise thus allows one to determine just how much information is lost when a field is compressed into summary statistics, and also experiment with new combinations of summary statistics to find one which preserves the most information possible. This potentially useful finding in 1D can subsequently be applied in future analyses in 3D, where the assessment at field-level is typically not possible.

2.2.1 Exact Dynamics in 1+1 Dimensions

As was already alluded in the previous section, a main benefit of working in one dimension is that the *exact* dynamics of the particles can be solved starting from the LPT framework, rather than doing an entire N-body simulation. This section describes how the equation of motion for

the particles is written and solved, both for cases with a single matter stream and for multiple streams.

One starts with the same differential equation, Equation 2.12, used for LPT in general. But because the calculation here is being done in one dimension, we now drop the vector notation $\mathbf{\Psi}(\mathbf{q})$ and use $\Psi(q)$ instead:

$$\frac{d^2\Psi(q)}{d\tau^2} + \mathcal{H}\frac{d\Psi(q)}{d\tau} = -\nabla_x\phi(q + \Psi(q)). \tag{2.16}$$

The one-dimensional displacement field $\Psi(q, \tau)$ can be used to express the density contrast $\delta(x)$ at some time τ as

$$1 + \delta(x) = \int dq\delta^D[x - q - \Psi(q)]. \tag{2.17}$$

The gravitational potential obeys the Poisson equation:

$$\begin{aligned}\nabla_x^2\phi(x,\tau) &= \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(x,\tau), \\ &= \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\int dq\left(\delta_D(x - q - \Psi(q,\tau)) - \delta_D(x - q)\right),\end{aligned} \tag{2.18}$$

where in the second line, $\delta(x, \tau)$ has been expressed using Equation 2.17.

The force is then determined by taking the integral, such that

$$-\nabla_x\phi(x,\tau) = -\frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\int dq\left(\Theta(x - q - \Psi(q,\tau)) - \Theta(x - q) + c(\tau)\right), \tag{2.19}$$

where $\Theta(x)$ is the Heaviside theta function, and $c(\tau)$ is a possibly time-dependent integration constant. This constant can be set to zero in the CMB rest frame, where the force vanishes so long as the particles are unperturbed (ie. having $\Psi(q, \tau) = 0$). From hereafter we assume to be in this CMB rest frame, so $c(\tau) = 0$.

Equation 2.19 can be thought of as simply counting how many sheets of matter are to the left and right at each position x. This is because in one spatial dimension, the force is completely independent of the distance. Consider for example the force at some position x, made by some segment of matter of length dq, which was initially at some Lagrangian coordinate q. If the segment was initially to the left of x for instance, and it has since moved to the right, then it exerts an infinitesimal rightward force on x given by $\frac{3}{2}\Omega_m\mathcal{H}^2dq$. On the contrary, if the segment has remained to the left of x, then it will have not induced any change to the force on x. In Equation 2.19, the Heaviside functions Θ simply keep track of the amount of matter (the number of sheets) which have crossed over x, and in particular the sign of the crossing.

The Single-Stream Zel'dovich Approximation

The first-order solution to LPT is called the Zel'dovich Approximation (ZA). In this approximation, one assumes that the displacement field $\mathbf{\Psi}(\mathbf{q}) \approx \mathbf{\Psi}^{(1)}(\mathbf{q})$ evolves linearly with the initial density perturbations. This is called an 'approximation' because in 3 dimensions, there are expected to

be higher orders than just the linear one. But in 1D, it was shown by [22] that in the case of a single matter stream, all the orders beyond linear are in fact zero. In one dimension, it can be seen that Equation 2.14 reduces to

$$1 + \delta = \mathrm{Det}[1 + \nabla_q \Psi(q)]^{-1} \tag{2.20}$$

and that the conversion between the Eulerian and Lagrangian gradients is now simply

$$\nabla_x = (1 + \nabla_q \Psi(q))^{-1} \nabla_q. \tag{2.21}$$

If one then takes the Eulerian gradient ∇_x of Equation 2.16 and plugs in the Poisson equation (the first line of Equation 2.18), it can be found that

$$(1 + \nabla_q \Psi(q))^{-1} \nabla_q \left[\frac{d^2\Psi}{d\tau^2} + \mathcal{H} \frac{d\Psi}{d\tau} \right] = \frac{3}{2} \Omega_m \mathcal{H}^2 \left(1 - \mathrm{Det}[1 + \nabla_q \Psi(q)]^{-1} \right), \tag{2.22}$$

and since in one dimension, $\mathrm{Det}[1 + \nabla_q \Psi(q)] = 1 + \nabla_q \Psi(q)$, it further reduces to

$$\nabla_q \left[\frac{d^2\Psi}{d\tau^2} + \mathcal{H} \frac{d\Psi}{d\tau} \right] = \frac{3}{2} \Omega_m \mathcal{H}^2 \nabla_q \Psi(q). \tag{2.23}$$

This is a linear equation, so the linear-order solution is the exact solution. Therefore in 1D for a single matter stream, the Zel'dovich approximation is actually the exact form of LPT, and $\Psi(q) = \Psi^{(1)}(q)$ [22]. Moreover, [22] showed that in 1D, standard n^{th} Eulerian perturbation theory converges to the Zel'dovich approximation for $n \to \infty$. Going to higher and higher orders becomes more and more computationally heavy. The Zel'dovoch approximation thus gives one the privilege in 1+1 dimensions to outright avoid using such high-order SPT in the single-stream case.

The solution for $\Psi(q)$ from the Zel'dovich approximation can be expressed in the form

$$\nabla_q \Psi = -D_1(\tau)\, \delta(q, 0), \tag{2.24}$$

where $\delta(q, 0)$ is the density field specified by initial conditions at $\tau = 0$, and $D_1(\tau)$ is a linear growth factor [19].

The Multi-stream Dynamics

Although the Zel'dovich approximation solves LPT in one dimension, this does not imply that the Zel'dovich approximation in itself provides an exact representation of the *true* gravitational evolution. In the Zel'dovich approximation the fluid elements evolve locally; they do not depend on the evolution of the other fluid elements. However, this assumption eventually breaks down at 'shell crossing'. As mentioned in Section 2.1, the 'single stream approximation' is a common assumption made in perturbation theories. By setting $\sigma_{ij} \approx 0$, one is setting the velocity dispersion to zero. Each spacetime point thus has only a single velocity value (a single stream). Eventually however, gravitational evolution proceeds to a state where there can be multiple matter streams

in the same spacetime point. This phenomenon is called 'shell crossing'. Shell crossing is the point at which fluid elements can attain the same Eulerian position x despite having had different initial positions q. At shell crossing, the Jacobian encoding the transformation between Eulerian and Lagrangian coordinates vanishes, and Equation 2.20 becomes infinite [22].

The mathematics of shell-crossing in 1D is explained in [23], but will also be outlined here using their equations. To identify shell-crossing points, one must first identify (for some position $\bar{x}$) the elements with $q = q_i(\bar{x})$ that make the following equation zero,

$$x(q,\tau) - \bar{x} = q + \Psi(q,\tau) - \bar{x} = 0, \tag{2.25}$$

so the $q_i(\bar{x})$ are thus the roots of the equation. The index i identifies each root $i = 1, 2, 3...$ ordered so that they are monotonically increasing. The position $x(q,\tau)$ is required to be a continuous function of q, taking values $(-\infty, \infty)$. This implies that for any $\bar{x}$, there will always be an odd number of roots $q_i(\bar{x})$. Each root solves the equation

$$q_i + \Psi(q_i,\tau) = q + \Psi(q,\tau) = x(q,\tau), \tag{2.26}$$

where it can be seen that there will always exist at least one solution, $q_i(x) = q(x)$. If for some position x the *only* root is this $q(x)$ itself, then there has been no shell crossing at that point; there is a single stream. But when two new real roots $q_j(x,\tau)$ and $q_{j+1}(x,\tau)$ appear at some time τ, shell crossing has occurred. It is also possible to have 5 roots or more at some position x. Any place where there exists more than one root is a 'multistreaming region', with the number of streams corresponding to the number of roots.

[23] demonstrated that the integral on the right-hand side of Equation 2.19 can be expressed in terms of an arbitrary number of roots q_i as

$$\lim_{L\to\infty} \int_{-L/2}^{L/2} dq \left(\Theta(x - q - \Psi(q,\tau)) - \Theta(x-q)\right) = -\sum_{i=1}^{N_s(x,\tau)} (-1)^{i+1} \Psi(q_i(x,\tau),\tau), \tag{2.27}$$

where $N_s(x(q,\tau),\tau)$ is the number of roots (the number of streams) in x. The equation of motion (Equation 2.16) can then be written as

$$\frac{d^2\Psi(q,\tau)}{d\tau^2} + \mathcal{H}\frac{d\Psi(q,\tau)}{d\tau} = \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2 \sum_{i=1}^{N_s(x(q,\tau),\tau)} (-1)^{i+1}\Psi(q_i,\tau). \tag{2.28}$$

If at a given position x there is no shell crossing, then $N_s(x(q,\tau),\tau) = 1$ at that point. In this single-stream case, the Lagrangian Perturbation Theory reduces *exactly* to the first-order solution (the Zel'dovich approximation). But in the multistream case, the force term on the right hand side becomes nonlocal in Lagrangian space and can no longer be represented perturbatively like in the Zel'dovich approximation. When particles cross one another, the force on each particle changes sign; but under the ZA the particles have no choice but to continue along their current ballistic paths. Instead, using the full form of Equation 2.28 allows one to recover the exact dynamics

post-shell-crossing.

The practical difficulty in solving Equation 2.28 is the right-hand side force term. [23] describes how one can solve it numerically. First, the logarithm of the scale factor is taken to be a "time" variable, here called η, defined as

$$\eta = \log \frac{a}{a_0} = -\log(1+z). \tag{2.29}$$

The equation of motion, Equation 2.28, can then be written as a system of equations in terms of η:

$$\begin{aligned} \partial_\eta \Psi(q,\eta) &= \chi(q,\eta), \\ \partial_\eta \chi(q,\eta) &= -\frac{1}{2}\chi(q,\eta) + \frac{3}{2}\Omega_m(\eta) \sum_{i=1}^{N_s(x,\eta)} (-1)^{i+1} \Psi(q_i(x,\eta),\eta). \end{aligned} \tag{2.30}$$

The system of equations is given an initial condition, at some early η, where one can assume the growing mode is described by linear perturbation theory such that

$$\Psi(q,\eta_{\rm in}) = \chi(q,\eta_{\rm in}) = \frac{v(q,\eta_{\rm in})}{\mathcal{H}(\eta_{\rm in})}, \tag{2.31}$$

where $v(q,\eta_{\rm in})$ is the peculiar velocity. The system of equations can then be solved numerically, in the following way. At each time step, for each position x, the set of roots $\{q_i\}$ are identified for the equation $x - q - \Psi(q,\eta) = 0$. Then for each q, the corresponding $x = q + \Psi(q,\eta)$ is calculated. Both $\Psi(q,\eta)$ and $\chi(q,\eta)$ are evolved forward by one time step, and their calculation depends on the previous set of roots $\{q_i\}$ via the summation in Equation 2.30.

This algorithm is then tested by [23] by applying some simple initial conditions and seeing how they evolve. They start with a single Gaussian overdensity and find that before shell crossing, the solution found by the algorithm is exactly the same as the Zel'dovich Approximation. After shell crossing, as expected, the solution diverges from the ZA. They also consider an additional initial condition, with another smaller Gaussian feature added on top, and find the same phenomenon.

2.2.2 The 1+1 Simulations

This thesis uses an ensemble of 1000 simulations, generated using the exact one-dimensional LPT framework described above. The 1-dimensional 'box' has a length of $L = 3000$ Mpc, and there are $N_p = 20{,}000$ particles in each simulation. In Fourier space this results in a maximum wavenumber $k_{\rm Nyquist} = 20.944$ and fundamental frequency $k_f = 0.00209$. The main fiducial cosmological parameters used for the simulations are $h = 0.6781$, $\Omega_m = 0.3$, $\Omega_b = 0.04867$, and $\sigma_8 = 0.8052$. The spectral index of the primordial scalar power spectrum is taken to be $n_s = 0.9660$. Later in Chapter 4, simulations with two non-fiducial values each for Ω_m and σ_8 are used in calculating the Fisher matrix. These non-fiducial values are presented in Table 4.1.

Each of the 1000 simulations is seeded from a random Gaussian initial field δ_{in}. A linear theory power spectrum P_{CAMB} is first generated using the cosmological transfer function code CAMB at $z = 0$ [24]. This power spectrum is then scaled back to a redshift $z = 100$, using the linear growth factor computed from the cosmological parameters. The power spectrum is translated from 3 dimensions into 1 dimension using Equation 3.16. The resulting power spectrum P_L, at $z = 100$, then dictates the Gaussian initial conditions for the simulations. The simulations are calculated for 300 time steps between $z = 100$ and $z = 0$.

Identifying Multistreaming Regions

For this thesis, there will be a particular interest in the multiple-stream regions in these 1+1 simulations. This is because in these regions, one expects to have collapsing structures that have mostly lost their cosmological information. In Chapter 4 Section 4.1, the field-level Fisher-matrix will be calculated, both including and excluding the identified multistreams. Comparing the Fisher matrices, it can be verified that the multistreaming regions *do not* contain much in the way of cosmological information, and can generally be removed before inferring cosmological parameters.

Using the algorithm described above, the multistreams are identified at each time step in each simulation. A visual example of multistreaming is shown in Figure 2.1 for one simulation, zoomed in on a 70 Mpc segment. This plot shows the Eulerian x vs. Lagrangian q coordinate for each particle. The particles highlighted in red are ones for which streams have already crossed, whilst the single-stream regions are plotted in green. Visually, the multistream regions appear as segments which are either horizontal or have a negative slope in x vs. q.

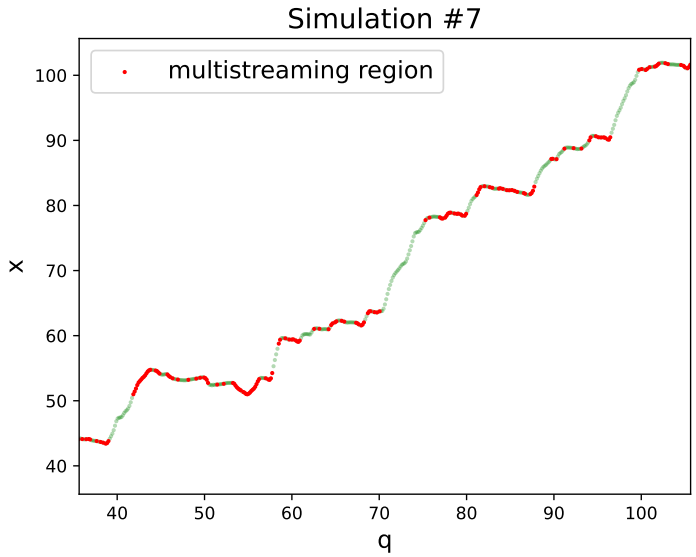

Figure 2.1: Example of multistreaming at $z = 0$, in a plot of x vs. q

3

Summary Statistics

How does one statistically describe a non-Gaussian field? This is a rather difficult question. For a Gaussian field, all the information about the field is contained solely in its two-point correlation function (or equivalently its Fourier transform, the power spectrum). However, for a strongly non-Gaussian field, the power spectrum alone is insufficient. Moreover, there is usually not one statistic that entirely encodes non-Gaussianity on its own. The simplest statistic that measures non-Gaussianity is the 3-point function, or the bispectrum. Other statistics such as the 4-point function (trispectrum) and 'marked' power spectrum have also been employed ([18]). This thesis focuses on the power spectrum and bispectrum, but there are a number of other statistics which can contain information from non-Gaussianity.

3.1 The Power Spectrum

Let $\rho(\mathbf{r})$ be the cosmic density field in physical space, and $\bar{\rho}$ be the mean density. The density contrast $\delta(\mathbf{r})$ is generally defined as

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \bar{\rho}}{\bar{\rho}}. \tag{3.1}$$

To calculate the power spectrum, one must first transform this $\delta(\mathbf{r})$ into Fourier space. If one assumes a large cosmological volume V_μ in which $\rho(\mathbf{r})$ adequately represents the entire cosmic density field, and that $\rho(\mathbf{r})$ can be assumed to be periodic, the Fourier transform of $\delta(\mathbf{r})$ can be defined as

$$\delta(\mathbf{k}) = \frac{1}{V_\mu} \int_{V_\mu} \delta(\mathbf{r}) e^{i\mathbf{r}\cdot\mathbf{k}}\, d\mathbf{r}. \tag{3.2}$$

Assuming $\delta(\mathbf{k})$ has isotropic statistics, the power spectrum $P(k)$ is then expressed as

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \delta_D(\mathbf{k} + \mathbf{k}')P(k), \tag{3.3}$$

where $\mathbf{k}$ and $\mathbf{k}'$ are two wavenumbers [19]. This can also be expressed more compactly as

$$P(k) \equiv \left\langle |\delta(\mathbf{k})|^2 \right\rangle, \tag{3.4}$$

where $\langle ... \rangle$ represents the ensemble average [25]. This thesis assumes a one-dimensional version of isotropy, 'reflection invariance' in which $P(k) = P(-k)$ [22].

Oftentimes cosmologists do not work directly with a continuous field, but rather with a distribution of discrete tracers (eg. the dark matter particles in this thesis). The density field $\rho(\mathbf{r})$ is then represented by a discrete number density distribution, and the discrete density contrast $\delta^d(\mathbf{k})$ is written as

$$\delta^d(\mathbf{k}) = \frac{1}{V_\mu \bar{n}} \int_{V_\mu} n(\mathbf{r}) e^{i\mathbf{r}\cdot\mathbf{k}} d\mathbf{r} - \delta^K_{\mathbf{k},0}, \tag{3.5}$$

where $n(\mathbf{r}) = \sum_j \delta^D(\mathbf{r} - \mathbf{r}_j)$ is the number density comprised of a summation of Dirac delta functions $\delta^D(\mathbf{r})$, $\mathbf{r}_j$ is the coordinate of object j, and $\bar{n}$ is the mean number density. Furthermore, the volume V_μ can be broken up into infinitesimal elements dV_i which each contain only 1 or 0 particles. In this case, $\delta^D(\mathbf{r})$ can be expressed with

$$\delta^d(\mathbf{k}) = \frac{1}{N} \sum_i n_i e^{i\mathbf{r}_i\cdot\mathbf{k}} - \delta^K_{\mathbf{k},0}, \tag{3.6}$$

where n_i can only be either 0 or 1, and N is the total number of objects in V_μ. In this case, the power spectrum can be expressed from $\delta^d(\mathbf{k})$ as

$$P(k) \equiv \langle |\delta(\mathbf{k})|^2 \rangle = \langle |\delta^d(\mathbf{k})|^2 \rangle - \frac{1}{N}, \tag{3.7}$$

where $1/N$ is a 'shot noise' term accounting for the discreteness of the particles [25].

Due to computational limitations, oftentimes the power spectrum is grouped into wavenumber bins of width Δk to make a 'binned' power spectrum $\hat{P}(k)$ that is averaged within each bin. In 3 dimensions this is expressed by [26] as

$$\hat{P}(k) = \frac{1}{V_{\text{bin}}} \int_{\text{bin}} d^3\mathbf{q}\, |\delta(\mathbf{q})|^2, \tag{3.8}$$

where $\mathbf{q}$ are the individual wavevectors in the bin, and $V_{\text{bin}} = \int_{\text{bin}} d^3\mathbf{q}$ is the volume of the bin, which collapses to $V_{\text{bin}} = \int_{\text{bin}} dq$ in the one-dimensional case.

3.1.1 The Count-in-Cell Scheme

When calculating the Fourier transform of $\delta(\mathbf{r})$ for a discrete particle distribution, it can be useful to use a Fast Fourier Transform (FFT) algorithm. This is because for a large number of particles, calculating the discrete Fourier Transform directly from the summation in Equation 3.6 can become rather computationally prohibitive. However, the FFT requires that the number density $n(\mathbf{r})$ is sampled on a uniform, evenly-spaced grid. The Count-in-Cell (CIC) scheme interpolates particles onto a grid, creating a density field that can be efficiently Fourier-transformed with FFT algorithms. Using the FFT, the resulting density contrast $\delta^f(\mathbf{k})$ can then be expressed as

$$\delta^f(\mathbf{k}) = \frac{1}{N}\sum_g n^f(\mathbf{r}_g)e^{i\mathbf{r}_g\cdot\mathbf{k}} - \delta^K_{\mathbf{k},0}, \tag{3.9}$$

where there are g grid points each with a position r_g, and $n^f(\mathbf{r}_g)$ is the convolved number density at each grid point. According to [27] and [25], the convolution itself is written as

$$n^f(\mathbf{r}_g) = \int n(\mathbf{r})W(\mathbf{r}-\mathbf{r}_g)d\mathbf{r}, \tag{3.10}$$

where $W(\mathbf{r})$ is the 'mass assignment function' which determines the mass at each grid point. There are many choices for mass assignment functions, with CIC being one of the most common. For the CIC scheme in particular, the mass assignment can be expressed in real space with

$$W(x_i) = \begin{cases} 1-|x_i|, & |x_i|<1 \\ 0, & \text{else} \end{cases} \tag{3.11}$$

where x_i is the i-th component of x. In three dimensions there would be three components (i=1,2,3), but for one dimension, x_1 is the only component. Moreover, the total mass assignment function $W(\mathbf{r})$, which is typically a product of the three components as $W(\mathbf{x}) = \prod_i W(x_i)$, reduces to $W(x_1)$ in 1D. The CIC mass assignment function can be described in Fourier space as

$$W(k_i) = \left[\frac{\sin\left(\frac{\pi k_i}{2k_N}\right)}{\left(\frac{\pi k_i}{2k_N}\right)}\right]^2, \tag{3.12}$$

where $k_N = \pi/H$ is the Nyquist frequency and H is the spacing between grid points [28].

Confining particles to a discrete grid results in augmentations to $\delta(\mathbf{k})$ which therefore has an effect on the measured power spectrum and bispectrum. According to [27], the FFT power spectrum estimator in 3 dimensions, taking into account the window function, is

$$\begin{aligned} \langle|\delta^f(\mathbf{k})|^2\rangle &= \sum_n |W(\mathbf{k}+2k_N\mathbf{n})|^2 P(\mathbf{k}+2k_N\mathbf{n}) + \frac{1}{N}\sum_n |W(\mathbf{k}+2k_N\mathbf{n})|^2, \\ &= \sum_n |W(\mathbf{k}+2k_N\mathbf{n})|^2 P(\mathbf{k}+2k_N\mathbf{n}) + D^2(\mathbf{k}), \end{aligned} \tag{3.13}$$

where the summation is over all 3D integer vectors $\mathbf{n}$.

The first term in Equation 3.13 is difficult to extract $P(\mathbf{k})$ from. The summation is over $\mathbf{n}$, but the contribution where $\mathbf{n} = 0$ is the only one intended to be measured. The $\mathbf{n} \neq 0$ terms are aliases which prevents one from easily recovering the true power spectrum $P(\mathbf{k})$. The window function in Equation 3.12 is equal to 1 at $k = 0$, but decreases as $k > 0$. To account for this 'smearing' or 'smoothing' effect, for $k \leq k_N$, the density contrast in Fourier space can simply be divided by the window function as $\delta^f(\mathbf{k})/W(\mathbf{k})$ [25]. This still leaves the 'aliasing' effect unaccounted for; however, this effect is expected to be an issue only at the largest wavenumbers $k \approx k_N$ and has to be handled in a less straightforward way (such as with the iterative scheme presented in [27]).

On the other hand, according to [27] the shot noise term $D^2(\mathbf{k})$ for the CIC case can simply be expressed (here specialized to one dimension) as

$$D^2(k) \approx \frac{1}{N}\left[1 - \frac{2}{3}\sin^2\left(\frac{\pi k}{2k_N}\right)\right]. \tag{3.14}$$

3.2 The Power Spectrum in 1D

Using the equations from Section 3.1, the power spectrum is calculated from the 1+1 simulations at $z = 0$. Because there is only one spatial dimension, all the modes $k_\parallel$ lie along a single axis and there are no perpendicular modes $\mathbf{k}_\perp$ that contribute. According to [22], the power spectrum $P(\mathbf{k})$ at a given redshift z can therefore be written as

$$P(k_\parallel, \mathbf{k}_\perp; z) = P_{1d;z}(k_\parallel)(2\pi)^2\delta^D(\mathbf{k}_\perp). \tag{3.15}$$

Working in one dimension gives the advantage of being able to calculate the Fourier transform using the 'brute force' summation approach in Equation 3.6, because there are few enough points for this Fourier Transform to be computationally tractable. Using this approach, one can avoid the step of gridding the particles with the CIC scheme in Section 3.1.1, as was required for the FFT. For this particular work, it was found that the two methods ('brute force' Fourier transform vs. the CIC+FFT) yielded virtually the same power spectrum except at very small scales ($k \gtrsim 10$) at which point aliasing effects from the CIC scheme start to become relevant. The power spectrum is individually calculated for all 1000 simulations, and then averaged to produce the final measured power spectrum. In presenting the power spectrum here it is kept unbinned, but the binned power spectrum will become useful later in Chapter 4 when calculating the Fisher matrix.

The measured power spectrum can be compared to the linear CDM power spectrum from CAMB, P_L at $z = 0$, which had been used to initialize the simulations in Section 2.2.2. CAMB calculates the power spectrum for three dimensions though, so the outputted power spectrum P_{CAMB} needs to be converted into a one-dimensional 'CDM-like' form. According to [22] this conversion can be written at a given redshift z as

$$\pi^{-1}kP_L(k; z) = (2\pi^2)^{-1}k^3P_{\mathrm{CAMB}}(k; z). \tag{3.16}$$

Moreover, the power spectra are plotted here in their dimensionless form, which in 1D is defined with

$$\Delta^2 \equiv k_\parallel P(k_\parallel, \mathbf{k}_\perp; z)/\pi. \tag{3.17}$$

Figure 3.1 shows the measured power spectrum from the 1+1 simulations, both for the 'brute force' Fourier transform and the CIC+FFT approach. The linear power spectrum P_L is also overplotted for comparison.

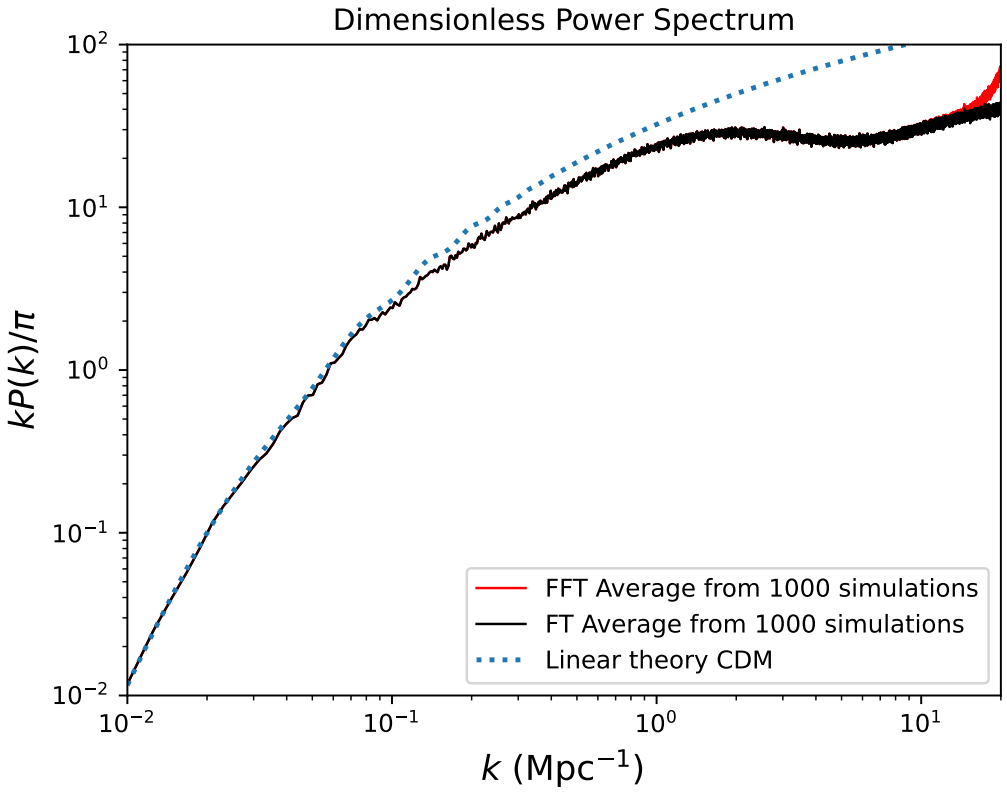

Figure 3.1: The measured power spectrum from the 1+1 simulations, compared with linear theory

3.3 The Bispectrum in 1D

Assuming isotropic statistics, the bispectrum $B(k_1, k_2)$ is defined as

$$\langle \delta(\mathbf{k_1})\delta(\mathbf{k_2})\delta(\mathbf{k_3})\rangle = \delta_D(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})B(k_1,k_2), \tag{3.18}$$

where $\delta_D(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})$ is the condition constraining the three wavevectors to form a triangle, such that there are really only two free parameters $\mathbf{k_1}$ and $\mathbf{k_2}$. In one spatial dimension, the bispectrum

is calculated entirely from flattened triangles. This is a natural byproduct of the triangle condition for the wavevectors $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$. Since all the wavevectors are confined along one axis, the magnitudes strictly add as $k_1 = k_2 + k_3$ rather than the usual triangle inequality $k_1 \leq k_2 + k_3$. According to [29], the bispectrum in one dimension can be written as

$$B(k_1, k_2) = L^{-1} \langle \delta(k_1)\delta(k_2)\delta(-k_1 - k_2) \rangle \,, \tag{3.19}$$

where L is the length of the box (the 1D volume).

In this work, the bispectrum is calculated using 1000 bins each for k_1, k_2, and k_3. Within the wavenumber bins, one must average over all values of $\delta(k_1)\delta(k_2)\delta(k_3)$ that form a triangle. The binned bispectrum $\hat{B}(k_1, k_2)$ (for binned wavenumbers k_1, k_2, k_3) is written in 3 dimensions by [26] as

$$\begin{aligned} \hat{B}(k_1, k_2, k_3) &= \frac{1}{V_\Delta} \int_{\text{bin 1}} d^3\mathbf{q}_1 \int_{\text{bin 2}} d^3\mathbf{q}_2 \int_{\text{bin 3}} d^3\mathbf{q}_3 \, \delta(\mathbf{q}_1)\delta(\mathbf{q}_2)\delta(\mathbf{q}_3)\delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3), \\ \text{where} \quad V_\Delta &\equiv \int_{\text{bin 1}} d^3\mathbf{q}_1 \int_{\text{bin 2}} d^3\mathbf{q}_2 \int_{\text{bin 3}} d^3\mathbf{q}_3 \, \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3), \end{aligned} \tag{3.20}$$

and where $\mathbf{q}_i$ represents each individual wavevector contained in the bin centered on $\mathbf{k}_i$. For this thesis, the expression is collapsed into one dimension by simply integrating over $\int_{\text{bin i}} dq_i$ instead of $\int_{\text{bin i}} d^3\mathbf{q}_i$.

The bispectrum is calculated for each of the 1000 simulations, and the final measured bispectrum is taken to be the ensemble average of these. Figure 3.2 shows the measured bispectrum, plotted on the k_1 vs. k_2 plane. It can be seen that the bispectrum peaks at configurations where both k_1 and k_2 are small. For larger wavenumbers, the bispectrum magnitude is higher when $k_3 \ll k_2 \approx k_1$ (squeezed triangles). Conversely, the magnitude of the bispectrum is lowest at large wavenumbers with $k_2 = k_3 = k_1/2$ (flattened triangles).

3.3.1 The Theoretical Bispectrum from Tree-level SPT

In three dimensions, it is possible for the bispectrum to peak for equilateral triangles with $k_1 = k_2 = k_3$. But in one dimension, this is not a feasible configuration because all the triangle configurations are flat. Since the one dimensional bispectrum can thus be difficult to intuitively justify, it is prudent to calculate a theoretical bispectrum to compare the result with. Here the theoretical bispectrum is calculated from tree-level Standard Perturbation Theory in one dimension.

In Section 4.1.3 of [19], they show that the tree-level bispectrum in Eulerian non-linear perturbation theory can be written at a given redshift z as

$$\begin{aligned} B^{(0)}(\mathbf{k_1}, \mathbf{k_2}, z) = 2F_2(\mathbf{k_1}, \mathbf{k_2})P_L(\mathbf{k_1}, z)P_L(\mathbf{k_2}, z) + 2F_2(\mathbf{k_2}, \mathbf{k_3})P_L(\mathbf{k_2}, z)P_L(\mathbf{k_3}, z) \\ + 2F_2(\mathbf{k_3}, \mathbf{k_1})P_L(\mathbf{k_3}, z)P_L(\mathbf{k_1}, z). \end{aligned} \tag{3.21}$$

The bispectrum is constructed from the linear power spectrum $P_L(\mathbf{k}, z)$ at redshift z, and the second-order kernel $F_2(\mathbf{k_1}, \mathbf{k_2})$. This kernel encodes the coupling between modes, and is obtained from the equations of motion for an Einstein de-Sitter universe in Eulerian non-linear perturbation theory. However, instead of using the 3-dimensional kernel provided in [19], this thesis instead uses the 1-dimensional version provided in [22], which has been symmetrized (averaged over permutations in the arguments k_1 and k_2):

$$F_2^{\text{sym}}(k_1, k_2) = 1 + \frac{1}{2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right). \tag{3.22}$$

Plugging this in for F_2^{sym} in Equation 3.21 yields the bispectrum plotted in Figure 3.3. One can see that just like the measured bispectrum in Figure 3.2, it demonstrates the same qualitative behavior, peaking for small wavenumbers and demonstrating the lowest magnitude for flattened configurations.

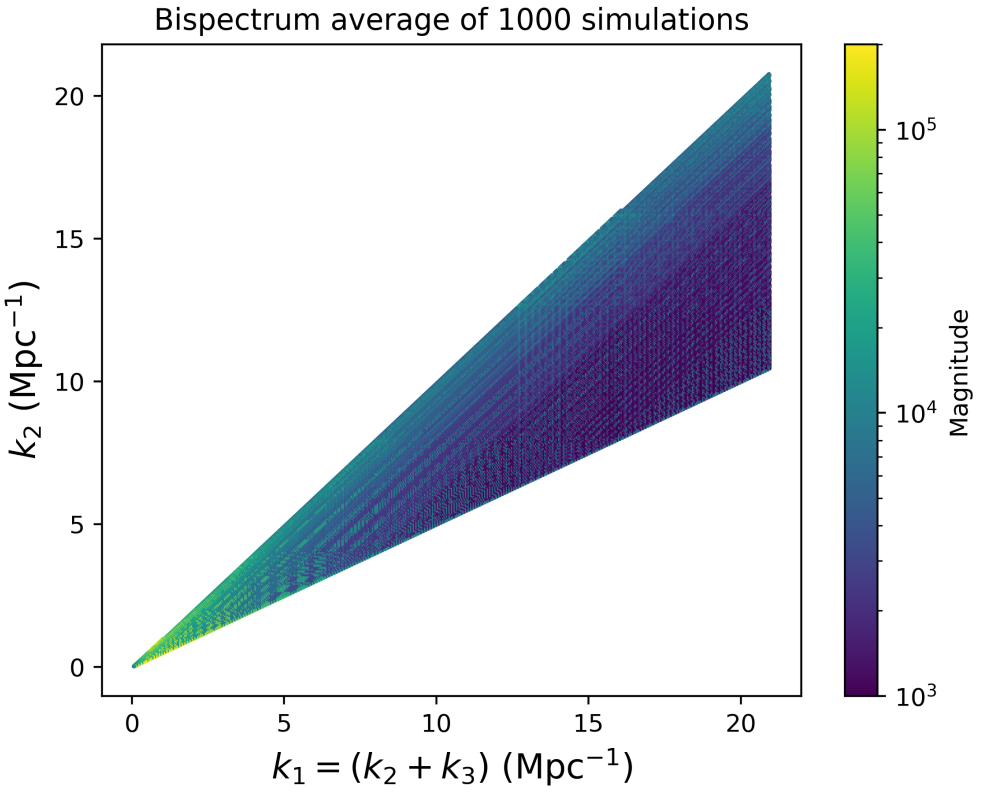

Figure 3.2: The measured bispectrum from the 1+1 simulations

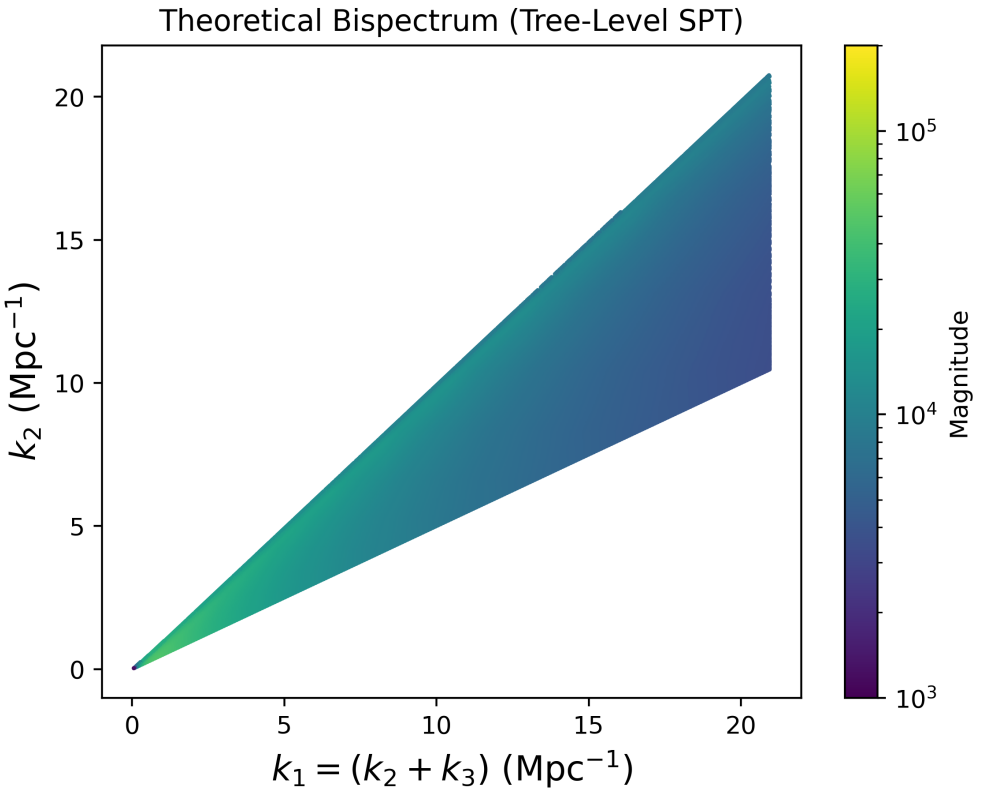

Figure 3.3: The 1-dimensional theoretical bispectrum derived from tree-level SPT

4

The Fisher Matrix

In this thesis, the Fisher matrix is used to estimate the amount information contained either at field-level or by a summary statistic. The Fisher matrix is useful because its inverse essentially quantifies the lowest-possible error bars that can be achieved for a model parameter, from a given observable (or set of observables) before having even gathered data.

Considering some model parameters θ_α and θ_β, the Fisher matrix $F_{\alpha\beta}$ is defined as

$$F_{\alpha\beta} \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}}. \tag{4.1}$$

In this equation, $\mathbf{d}$ is the vector of data points, and $\boldsymbol{\theta}$ is the vector of the parameters for the model. $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})$ is the likelihood of data, given the model. The $\langle ... \rangle$ brackets indicate the ensemble average over all possible realizations of the dataset. The subscript $_{\boldsymbol{\theta}=\bar{\theta}}$ indicates that the quantity is evaluated at the parameter values θ_α, θ_β which maximize the likelihood; $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_{\max \mathcal{L}}$. These are called the *fiducial* parameter values. The Fisher matrix is thus the ensemble average of the negative curvature of the log-likelihood, evaluated for the maximum likelihood.

The estimate for some parameter θ_α from the data will be called $\hat{\theta}_\alpha$. If the estimator is *unbiased*, therefore $\langle \hat{\theta}_\alpha \rangle = \theta_\alpha$, then the standard deviation of $\hat{\theta}_\alpha$ has a lower bound:

$$\sigma_{\hat{\theta}_\alpha} \geq \sqrt{\langle (F^{-1})_{\alpha\alpha} \rangle}. \tag{4.2}$$

This is the Cramer-Rao bound for the parameter θ_α. It does not imply that we can practically find an estimator $\hat{\theta}_\alpha$ to actually saturate the bound; it is simply the theoretical limit for an unbiased estimator. Another caveat is that computing the Fisher matrix requires already knowing in advance what the fiducial parameter value(s) should be to maximize the likelihood. These

fiducial values $\bar{\boldsymbol{\theta}}$ are generally informed by the results of previous experiments.

For this thesis, the Fisher matrix will be calculated for the two cosmological parameters $\theta_\alpha = \Omega_m$ and $\theta_\beta = \sigma_8$.

4.1 The Field-Level Fisher Matrix

First, the Fisher matrix will be calculated at the field level, to assess the information contained. The field-level Fisher matrix will be compared with itself, for various levels of multistreaming excluded. Then in Section 4.2, the field-level results will be compared with the Fisher matrix from the power spectrum. The equations for this section were kindly provided by Dr. Massimo Pietroni (private communication). The likelihood of some field $\delta(\mathbf{r}, \tau)$ with cosmological parameters $\boldsymbol{\theta}$ given an initial field $\delta_{\rm in}$ can be expressed in the 3-dimensional case as

$$\begin{aligned} P[\delta; \boldsymbol{\theta}|\delta_{\rm in}] &= \mathcal{N}(\det\sigma)^{-1/2} \\ &\times \exp\left[-\frac{1}{2}\int d^3r\, d^3r' \left(\delta(\mathbf{r}) - \hat{\delta}[\delta_{\rm in};\boldsymbol{\theta}](\mathbf{r})\right)\sigma^{-1}(\mathbf{r},\mathbf{r}')\left(\delta(\mathbf{r}') - \hat{\delta}[\delta_{\rm in};\boldsymbol{\theta}](\mathbf{r}')\right)\right], \end{aligned} \tag{4.3}$$

where $\mathcal{N}$ is a normalization factor such that

$$\int \mathcal{D}\delta P[\delta; \boldsymbol{\theta}|\delta_{\rm in}] = 1. \tag{4.4}$$

$\hat{\delta}[\delta_{\rm in};\boldsymbol{\theta}]$ is the forward model predicting the observed field $\bar{\delta}(\mathbf{r})$ from the initial field $\delta_{\rm in}$, which encodes all of the gravitational evolution. $\sigma(\mathbf{r},\mathbf{r}')$ is the noise covariance matrix, which can also be expressed in Fourier space as

$$\sigma^{-1}(\mathbf{r},\mathbf{r}') = \int \frac{d^3k}{(2\pi)^3} e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\sigma_{\mathbf{k}}^{-1}. \tag{4.5}$$

If one assumes white noise, then $\sigma_{\mathbf{k}}$ is a constant. This thesis uses $\sigma_{\mathbf{k}} = \sigma_0 = L/N_p = 1/\bar{n}$ (Poisson shot noise). The field-level Fisher matrix (with the likelihood conditioned on δ_{in}) is then expressed for some cosmological parameters θ_α and θ_β as

$$F_{\alpha\beta}[\delta_{\rm in}] = -\left\langle \frac{\partial^2 \ln P[\delta;\boldsymbol{\theta}|\delta_{\rm in}]}{\partial\theta_a \partial\theta_b}\right\rangle_{\theta=\bar{\theta},\delta_{\rm in}}. \tag{4.6}$$

In the three-dimensional case, plugging in $P[\delta;\boldsymbol{\theta}|\delta_{\rm in}]$, this becomes

$$\begin{aligned} F_{\alpha\beta}[\delta_{\rm in}] &= \int d^3r d^3r'\, \frac{\partial\hat{\delta}[\delta_{\rm in};\boldsymbol{\theta}](\mathbf{r})}{\partial\theta_a}\sigma^{-1}(\mathbf{r},\mathbf{r}')\frac{\partial\hat{\delta}[\delta_{\rm in};\boldsymbol{\theta}](\mathbf{r}')}{\partial\theta_b}, \\ &= \sigma_0^{-1}\int d^3r\, \frac{\partial\hat{\delta}[\delta_{\rm in};\boldsymbol{\theta}](\mathbf{r})}{\partial\theta_a}\frac{\partial\hat{\delta}[\delta_{\rm in};\boldsymbol{\theta}](\mathbf{r})}{\partial\theta_b}. \end{aligned} \tag{4.7}$$

	N_{sims}	Ω_m	σ_8	Ω_b	n_s	h
Fiducial	1000	0.30	0.80521	0.04868	0.96605	0.67810
Displaced +	2000	0.31	0.82021	”	”	”
Displaced −	2000	0.29	0.79021	”	”	”

Table 4.1: The fiducial and non-fiducial cosmological parameter values used for computing the Fisher matrices.

where in the second line the white noise σ_0 has been brought outside the integral.

Finally, in one dimension the $\int d^3r$ collapses to an integral $\int dr$ over the box length L. The conditioned version of the Fisher matrix can thus be expressed as

$$F_{\alpha\beta}[\delta_{\text{in}}] = \sigma_0^{-1} \int_{-L/2}^{L/2} dr \left(\frac{\partial \hat{\delta}[\delta_{\text{in}}; \boldsymbol{\theta}](r)}{\partial \theta_\alpha} \frac{\partial \hat{\delta}[\delta_{\text{in}}; \boldsymbol{\theta}](r)}{\partial \theta_\beta} \right). \tag{4.8}$$

Here the parameters of interest are $\theta_\alpha = \Omega_m$ and $\theta_\beta = \sigma_8$, and the noise covariance is $\sigma_0 = L/N_p$, so the final field-level Fisher matrix is written as

$$F_{\alpha\beta}[\delta_{\text{in}}] = \frac{N_p}{L} \int_{-L/2}^{L/2} dr \left(\frac{\partial \hat{\delta}[\delta_{\text{in}}; \boldsymbol{\theta}](r)}{\partial \Omega_m} \frac{\partial \hat{\delta}[\delta_{\text{in}}; \boldsymbol{\theta}](r)}{\partial \sigma_8} \right). \tag{4.9}$$

The derivatives in the Equation 4.9 are derivatives of the density field with respect to the parameters, evaluated at their fiducial values. To numerically calculate each of these derivatives, one must also generate simulations of the density field δ at *non-fiducial* parameter values, ideally on either side of the fiducial ones. The derivative is then computed numerically using the finite central difference. Therefore in addition to the fiducial values, 1000 of the 1+1 simulations were each generated for $\Omega_m = 0.30 \pm 0.1$, and $\sigma_8 = 0.805 \pm 0.015$ as listed in Table 4.1. To take the difference in δ between the fiducial and non-fiducial simulations, the field needs to be gridded through a mass-assignment scheme. The CIC scheme described in Section 3.1.1 is used to accomplish this.

The Fisher matrix $F_{\alpha\beta}[\delta_{in}]$ is computed for each of the 1000 simulations, and then the ensemble average is taken over all the simulations, $\langle F_{\alpha\beta}[\delta_{in}] \rangle$. Finally, the minimum uncertainties on Ω_m and σ_8 are estimated as

$$\sigma_{\Omega_m} \geq \sqrt{\langle (F^{-1})_{\Omega_m \Omega_m}[\delta_{in}] \rangle} \quad \sigma_{\sigma_8} \geq \sqrt{\langle (F^{-1})_{\sigma_8 \sigma_8}[\delta_{in}] \rangle}. \tag{4.10}$$

4.1.1 The Information Content of Multistreaming Regions

One very interesting application of the field-level Fisher matrix is determining how much information is contained in multistreaming regions. In the multistreaming regions, it is expected that there are collapsed structures which no longer carry much cosmological information. In these

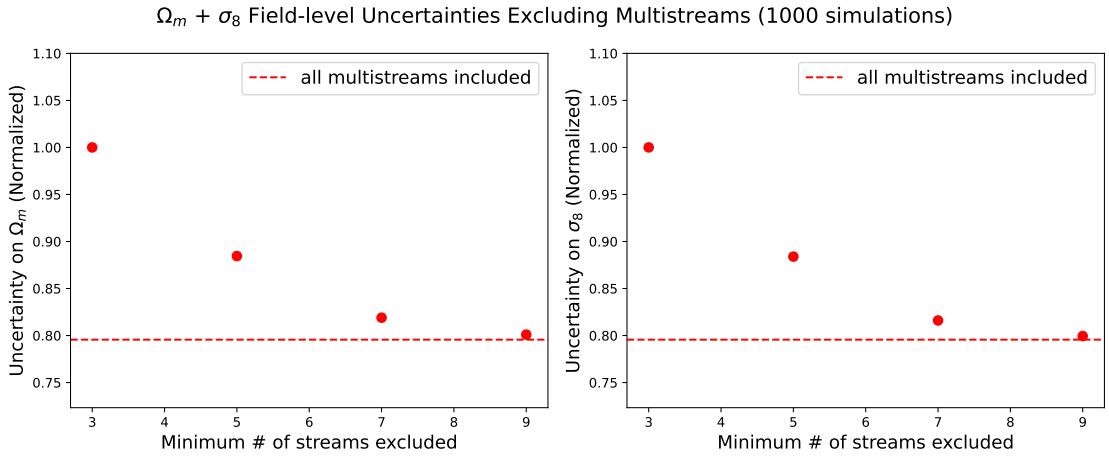

Figure 4.1: Field-level normalized uncertainties for Ω_m and σ_8, excluding multistreams

regions, shell crossing has already taken place (as described in Section 2.2.2) and the Zel'dovich Approximation is no longer exact, even in 1D. But because the Zel'dovich Approximation is exact right up to shell crossing in 1D, working with 1+1 dimensional simulations gives a unique opportunity to identify the multistreams more easily than in 3D [23]. Calculating the field-level Fisher matrix before and after removing multistreams allows one to estimate how much information is contained in the multistreaming regions. Using the procedure described in Section 2.2.2, different levels of multistreams were removed from the simulations: first only regions with 9+ streams, then 7+, then 5+, and finally 3+ (such that all multistreams are removed). The ensemble field-level Fisher matrices were calculated separately for each of these instances. The resulting uncertainties on Ω_m and σ_8 are compared in Figure 4.1 to the result with no multistream removal (the dashed line in the Figure). It can be seen that removing all multistreams (3 streams or more) causes the uncertianties on both Ω_m and σ_8 to increase by $\sim 25\%$. Including more and more streams causes the uncertainties to decrease and converge to the all-stream case, especially once only 9+ streams are excluded. Notably, the uncertainties are plotted relative to no-multistream case. This is because the value of the shot noise term σ_0 is dependent on the number of particles. In circumstances where the matter distribution is traced by a smaller number of particles, for example dark matter halos, the shot noise is higher. The absolute values of σ_{Ω_m} and σ_{σ_8} therefore do not particularly lend themselves to general conclusions. When examining the values of these uncertainties, it is instead sensible to compare them relative to one another rather than focusing on their absolute values.

The main takeaway is that in one dimension at field level, the multistreams do *not* contain the majority of information about the cosmological parameters Ω_m and σ_8. There is therefore a lot of information in single-stream regions that can be successfully described solely by the Zel'dovich approximation. The next task is to determine how information from the field can best be compressed into summary statistics.

4.2 The Power Spectrum Fisher Matrix

One of the most useful aspects of computing the field-level Fisher matrix is then comparing it to the Fisher matrices of summary statistics, to gauge the information content of each statistic. Here the Fisher matrix is computed from the power spectrum, and the resulting uncertainties on Ω_m and σ_8 are compared to the field-level case.

To start, the likelihood of the power spectrum $P(k)$ is assumed to be a multivariate Gaussian with a mean $\mu(\boldsymbol{\theta}) = \langle P(k) \rangle$ and covariance $\mathbf{C}(\boldsymbol{\theta}) = \langle P(k)P(k)^T \rangle - \mu\mu^T$. The covariance matrix for the power spectrum can be practically calculated from the ensemble of simulations as

$$\mathbf{C}(k_n, k_m) = \frac{1}{N-1} \sum_i \left(P_i(k_n) - \langle P(k_n) \rangle\right) \left(P_i(k_m) - \langle P(k_m) \rangle\right), \tag{4.11}$$

where $P_i(k)$ represents the power spectrum from each of the individual simulations, $\langle P(k) \rangle$ is the ensemble average, and $N = 1000$ is the number of simulations.

According to [30], the equation for the Fisher Matrix from a multivariate Gaussian likelihood can be expressed most generally as

$$\mathbf{F}_{\alpha\beta} = \frac{1}{2} \mathrm{tr} \left[\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_\beta} \right] + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_\beta}. \tag{4.12}$$

In the particular case of the power spectrum, the first term in this equation can be dropped, leaving only

$$\mathbf{F}_{\alpha\beta} = \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_\beta} = \frac{\partial \langle P(k) \rangle^T}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial \langle P(k) \rangle}{\partial \theta_\beta}. \tag{4.13}$$

There are two regimes in which the information content in the power spectrum becomes especially interesting. The first is at small wavenumbers k (large scales). In this 'linear' regime, perturbations are small enough that they can still be treated linearly. The field is still mainly Gaussian; unlike smaller scales which have had enough time to collapse into non-Gaussian structures. As a result, the field on large scales can be described almost entirely by its power spectrum alone. To check this, the power spectrum was truncated at various wavenumbers k_{max} between 0.08 and 1.0 Mpc^{-1}, and the Fisher matrix was calculated for each of these cases. To compare with the power spectrum result, the same truncation was then done in Fourier space at field-level, and the field-level Fisher matrices were calculated from these truncated fields.

Figure 4.2 shows the resulting comparison between the power spectrum and field-level for their uncertainties on Ω_m and σ_8. One can see that at larger and larger scales (smaller k_{max}), the uncertainties from the power spectrum and field-level converge until they are nearly the same at 0.08 Mpc^{-1}.

The other regime where it is interesting to examine information content is at the smallest scales (large k_{max}). At these very nonlinear scales, the amount of information saturates and there is little more to be gained by considering larger and larger k_{max}. The challenge in computing the Fisher matrix for large k_{max} is that the covariance matrix of the power spectrum $P(k)$ is correspondingly

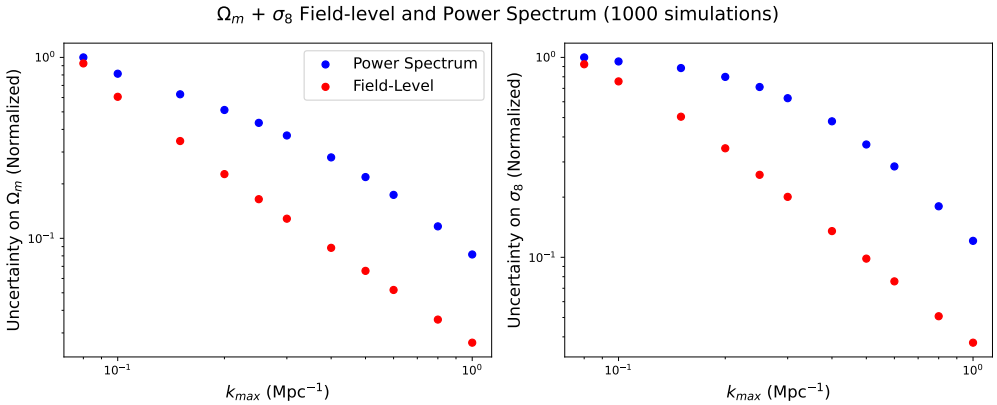

Figure 4.2: Field-level and power spectrum normalized uncertainties for Ω_m and σ_8 at large scales.

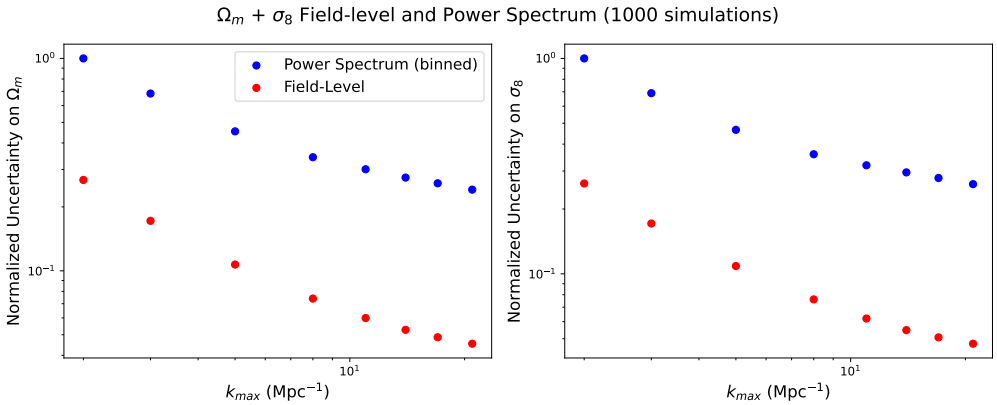

Figure 4.3: Field-level and power spectrum normalized uncertainties for Ω_m and σ_8 at small scales

large and therefore difficult to invert. For example, if the power spectrum is left unbinned and $k_{max} = k_N$, then the covariance matrix in this work would have a size of 10,000x10,000; the inverse $\mathbf{C}^{-1}$ would then be computationally prohibitive to calculate. Therefore for the purpose of calculating Fisher matrices at large k_{max}, the power spectrum is binned into 500 bins.

Figure 4.3 shows the resulting uncertainties on Ω_m and σ_8 for the field-level and binned power spectrum. First off, one can see that the uncertainties from the power spectrum alone are ∼3-4x higher than at field-level. It is indeed expected on such nonlinear scales that the power spectrum is not a sufficient sole descriptor of the field. Second, both the field-level and binned power spectrum produce uncertainties on Ω_m and σ_8 which level off around $k_{max} \sim 10$ Mpc^{-1}. This is quite a large k_{max}; larger than what might be expected in 3 dimensions.

5

Discussion, conclusion, and summary

5.1 Discussion

There are a number of interesting future directions for further analyzing the 1+1 simulations beyond this thesis. One important step would to assess more summary statistics besides the power spectrum alone. An obvious choice would be the bispectrum, to see what improvement it adds on top of the power spectrum in constraining cosmological parameters. Another possible option would be to compute 'marked' power spectra and bispectra. These 'marked' statistics weight the density field based on some chosen property, which can accentuate low/high-density regions, creating an enhanced sensitivity to nonlinear effects and essentially incorporating higher-order correlation information. The marked power spectrum, for example, was explored in 3D simulations in [18] and in [31].

Another useful development would be to trace the CDM particles of the simulation with dark matter halos, rather than examining the CDM matter distribution directly. In real life, we do not see dark matter, and observational surveys instead aim to map out large scale structure using baryonic tracers (such as galaxies) which are biased with respect to the underlying matter distribution. Using the halo model allows one to incorporate this bias. For example, the 3D halo catalogues constructed from the Quijote-PNG N-body simulations [32, 33] have been used to make Fisher forecasts for cosmological parameters and PNG parameters in a number of papers (eg. [17, 34, 18]). From the dark matter halos, one could then use the Halo Mass Function as a possible statistic, as was done in [17] for example.

Another possible statistic could be the wavelet scattering transform (WST). The wavelet transform decomposes the density field into localized, oscillatory components across different scales and positions. These wavelets can capture features like filaments, voids, and halos in the matter

distribution at various scales. The WST is structured like a 3-layer convolutional network, but unlike black-box machine learning models, the WST offers coefficients with clear physical meanings linked to scales and correlations [35, 36].

Instead of Fisher forecasting, one could instead use the 1+1 simulations to train deep neural networks (NNs) to perform a likelihood-free inference of cosmological parameters. The main benefit of using a NN is that training it can require less simulations than what is needed for a fully-converged Fisher forecast, and does not require a specific choice of fiducial cosmology. Instead of computing the Fisher matrix and Cramer-Rao bound from a specified likelihood to determine the merit of a summary statistic, one would instead use the standard deviations outputted by the neural network to compare statistics. Doing so gives an alternative result to compare with the Cramer-Rao bound. This has been carried out in 3D, for example, on the Quijote-PNG simulations [18].

In addition to investigating more summary statistics, future work could focus on adding in more cosmological parameters like Ω_b, n_s and h into the Fisher analysis. This would give a clearer picture about which summary statistics carry the most cosmological information. Moreover, one statistic may also be more suited to constrain a certain parameter more than another.

While a universe simulated in 1D undoubtedly does not describe our 3D reality, it holds some important similarities. As explained in Section 2.2, the 1+1 model uses the same dynamical equations for the Lagrangian Perturbation Theory as in 3D, just specialized such that all modes are along the same direction. Therefore, some qualitative lessons learned in 1D can possibly be translated into 3-dimensional scenarios.

5.2 Conclusion

This thesis examined an ensemble of simulations in 1+1 dimensions, and investigated their information content. In Chapter 1 some main cosmology background was introduced, along with the topic of Non-Gaussianity. Then in Chapter 2, perturbation theory was briefly introduced with a focus on Lagrangian Perturbation Theory in 1+1 dimensions. The details of the 1+1 model were presented and the concept of shell crossing was introduced. Chapter 3 then focused on measuring the mean power spectrum and bispectrum from the simulations. The bispectrum was found to peak for small wavenumbers, a result which was then verified by comparison to the theoretical bispectrum derived from tree-level SPT. Chapter 4 then focused on calculating the Fisher matrix at field-level and for the power spectrum, for the cosmological parameters Ω_m and σ_8. From the Fisher matrix, the minimum uncertainties on these two parameters were then determined from the Cramer-Rao bound. For the field-level analysis, different levels of multistreaming were excluded, and it was found that excluding all multistreams raises the uncertainties on both Ω_m and σ_8 by roughly 25%. Then the field-level result was compared with that of the power spectrum. It was demonstrated that for small k_{max} (ie. only considering large scales), the power spectrum contains most of the field information. On the other hand, for large k_{max}, the field-level uncertainties are considerably lower than that of the power spectrum. For both the field-level and power spectrum,

the information content saturates at $k_{max} \sim 10$ Mpc^{-1}, a scale at which the matter distribution has become highly nonlinear. Finally Chapter 5 Section 5.1 briefly discussed future prospects for using the 1+1 simulations, such as adding more summary statistics into the Fisher analysis and considering more cosmological parameters.

References

[1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry, "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant," , vol. 116, no. 3, pp. 1009–1038, Sep. 1998.

[2] Planck Collaboration, "Planck 2018 results. VI. Cosmological parameters," , vol. 641, p. A6, Sep. 2020.

[3] ——, "Planck 2018 results. X. Constraints on inflation," , vol. 641, p. A10, Sep. 2020.

[4] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, "In the realm of the Hubble tension-a review of solutions," *Classical and Quantum Gravity*, vol. 38, no. 15, p. 153001, Jul. 2021.

[5] M. Celoria and S. Matarrese, "Primordial Non-Gaussianity," *arXiv e-prints*, p. arXiv:1812.08197, Dec. 2018.

[6] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," , vol. 208, no. 2, p. 19, Oct. 2013.

[7] The Planck Collaboration, "The Scientific Programme of Planck," *arXiv e-prints*, pp. astro–ph/0 604 069, Apr. 2006.

[8] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, "The Statistics of Peaks of Gaussian Random Fields," , vol. 304, p. 15, May 1986.

[9] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, "Gauge-invariant second-order perturbations and non-Gaussianity from inflation," *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, Sep. 2003.

[10] J. Maldacena, "Non-gaussian features of primordial fluctuations in single field inflationary models," *Journal of High Energy Physics*, vol. 2003, no. 5, p. 013, May 2003.

[11] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, "Non-Gaussianity from inflation: theory and observations," , vol. 402, no. 3-4, pp. 103–266, Nov. 2004.

[12] R. Holman and A. J. Tolley, "Enhanced non-Gaussianity from excited initial states," , vol. 2008, no. 5, p. 001, May 2008.

[13] Planck Collaboration, "Planck 2018 results. IX. Constraints on primordial non-Gaussianity," , vol. 641, p. A9, Sep. 2020.

[14] ——, "Planck 2018 results. I. Overview and the cosmological legacy of Planck," , vol. 641, p. A1, Sep. 2020.

[15] F. Finelli, M. Bucher, A. Achúcarro, M. Ballardini, N. Bartolo, D. Baumann, S. Clesse, J. Errard, W. Handley, M. Hindmarsh, K. Kiiveri, M. Kunz, A. Lasenby, M. Liguori, D. Paoletti, C. Ringeval, J. Väliviita, B. van Tent, V. Vennin, P. Ade, R. Allison, F. Arroja, M. Ashdown, A. J. Banday, R. Banerji, J. G. Bartlett, S. Basak, P. de Bernardis, M. Bersanelli, A. Bonaldi, J. Borril, F. R. Bouchet, F. Boulanger, T. Brinckmann, C. Burigana, A. Buzzelli, Z. Y. Cai, M. Calvo, C. S. Carvalho, G. Castellano, A. Challinor, J. Chluba, I. Colantoni, A. Coppolecchia, M. Crook, G. D'Alessandro, G. D'Amico, J. Delabrouille, V. Desjacques, G. De Zotti, J. M. Diego, E. Di Valentino, S. Feeney, J. R. Fergusson, R. Fernandez-Cobos, S. Ferraro, F. Forastieri, S. Galli, J. García-Bellido, G. de Gasperis, R. T. Génova-Santos, M. Gerbino, J. González-Nuevo, S. Grandis, J. Greenslade, S. Hagstotz, S. Hanany, D. K. Hazra, C. Hernández-Monteagudo, C. Hervias-Caimapo, M. Hills, E. Hivon, B. Hu, T. Kisner, T. Kitching, E. D. Kovetz, H. Kurki-Suonio, L. Lamagna, M. Lattanzi, J. Lesgourgues, A. Lewis, V. Lindholm, J. Lizarraga, M. López-Caniego, G. Luzzi, B. Maffei, N. Mandolesi, E. Martínez-González, C. J. A. P. Martins, S. Masi, D. McCarthy, S. Matarrese, A. Melchiorri, J. B. Melin, D. Molinari, A. Monfardini, P. Natoli, M. Negrello, A. Notari, F. Oppizzi, A. Paiella, E. Pajer, G. Patanchon, S. P. Patil, M. Piat, G. Pisano, L. Polastri, G. Polenta, A. Pollo, V. Poulin, M. Quartin, A. Ravenni, M. Remazeilles, A. Renzi, D. Roest, M. Roman, J. A. Rubiño-Martin, L. Salvati, A. A. Starobinsky, A. Tartari, G. Tasinato, M. Tomasi, J. Torrado, N. Trappe, T. Trombetti, M. Tucci, C. Tucker, J. Urrestilla, R. van de Weygaert, P. Vielva, N. Vittorio, K. Young, and M. Zannoni, "Exploring cosmic origins with CORE: Inflation," , vol. 2018, no. 4, p. 016, Apr. 2018.

[16] K. Abazajian, G. Addison, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, M. Alvarez, A. Anderson, K. S. Arnold, C. Baccigalupi, K. Bailey, D. Barkats, D. Barron, P. S. Barry, J. G. Bartlett, R. Basu Thakur, N. Battaglia, E. Baxter, R. Bean, C. Bebek, A. N. Bender, B. A. Benson, E. Berger, S. Bhimani, C. A. Bischoff, L. Bleem, S. Bocquet, K. Boddy, M. Bonato, J. R. Bond, J. Borrill, F. R. Bouchet, M. L. Brown, S. Bryan, B. Burkhart, V. Buza, K. Byrum, E. Calabrese, V. Calafut, R. Caldwell, J. E. Carlstrom, J. Carron, T. Cecil, A. Challinor, C. L. Chang, Y. Chinone, H.-M. S. Cho, A. Cooray, T. M. Crawford, A. Crites, A. Cukierman, F.-Y. Cyr-Racine, T. de Haan, G. de Zotti, J. Delabrouille, M. Demarteau, M. Devlin, E. Di Valentino, M. Dobbs, S. Duff, A. Duivenvoorden, C. Dvorkin, W. Edwards, J. Eimer, J. Errard, T. Essinger-Hileman, G. Fabbian, C. Feng, S. Ferraro, J. P. Filippini, R. Flauger, B. Flaugher, A. A. Fraisse, A. Frolov, N. Galitzki, S. Galli, K. Ganga, M. Gerbino, M. Gilchriese, V. Gluscevic, D. Green, D. Grin, E. Grohs, R. Gualtieri, V. Guarino, J. E. Gudmundsson, S. Habib, G. Haller, M. Halpern, N. W. Halverson, S. Hanany, K. Harrington, M. Hasegawa, M. Hasselfield, M. Hazumi, K. Heitmann, S. Henderson, J. W. Henning, J. C. Hill, R. Hlozek, G. Holder, W. Holzapfel, J. Hubmayr, K. M. Huffenberger,

M. Huffer, H. Hui, K. Irwin, B. R. Johnson, D. Johnstone, W. C. Jones, K. Karkare, N. Katayama, J. Kerby, S. Kernovsky, R. Keskitalo, T. Kisner, L. Knox, A. Kosowsky, J. Kovac, E. D. Kovetz, S. Kuhlmann, C.-l. Kuo, N. Kurita, A. Kusaka, A. Lahteenmaki, C. R. Lawrence, A. T. Lee, A. Lewis, D. Li, E. Linder, M. Loverde, A. Lowitz, M. S. Madhavacheril, A. Mantz, F. Matsuda, P. Mauskopf, J. McMahon, M. McQuinn, P. D. Meerburg, J.-B. Melin, J. Meyers, M. Millea, J. Mohr, L. Moncelsi, T. Mroczkowski, S. Mukherjee, M. Münchmeyer, D. Nagai, J. Nagy, T. Namikawa, F. Nati, T. Natoli, M. Negrello, L. Newburgh, M. D. Niemack, H. Nishino, M. Nordby, V. Novosad, P. O'Connor, G. Obied, S. Padin, S. Pandey, B. Partridge, E. Pierpaoli, L. Pogosian, C. Pryke, G. Puglisi, B. Racine, S. Raghunathan, A. Rahlin, S. Rajagopalan, M. Raveri, M. Reichanadter, C. L. Reichardt, M. Remazeilles, G. Rocha, N. A. Roe, A. Roy, J. Ruhl, M. Salatino, B. Saliwanchik, E. Schaan, A. Schillaci, M. M. Schmittfull, D. Scott, N. Sehgal, S. Shandera, C. Sheehy, B. D. Sherwin, E. Shirokoff, S. M. Simon, A. Slosar, R. Somerville, D. Spergel, S. T. Staggs, A. Stark, R. Stompor, K. T. Story, C. Stoughton, A. Suzuki, O. Tajima, G. P. Teply, K. Thompson, P. Timbie, M. Tomasi, J. I. Treu, M. Tristram, G. Tucker, C. Umiltà, A. van Engelen, J. D. Vieira, A. G. Vieregg, M. Vogelsberger, G. Wang, S. Watson, M. White, N. Whitehorn, E. J. Wollack, W. L. Kimmy Wu, Z. Xu, S. Yasini, J. Yeck, K. W. Yoon, E. Young, and A. Zonca, "CMB-S4 Science Case, Reference Design, and Project Plan," *arXiv e-prints*, p. arXiv:1907.04473, Jul. 2019.

[17] G. Jung, A. Ravenni, M. Baldi, W. R. Coulton, D. Jamieson, D. Karagiannis, M. Liguori, H. Shao, L. Verde, F. Villaescusa-Navarro, and B. D. Wandelt, "Quijote-PNG: The Information Content of the Halo Mass Function," , vol. 957, no. 1, p. 50, Nov. 2023.

[18] G. Jung, A. Ravenni, M. Liguori, M. Baldi, W. R. Coulton, F. Villaescusa-Navarro, and B. D. Wandelt, "Quijote-PNG: Optimizing the summary statistics to measure Primordial non-Gaussianity," *arXiv e-prints*, p. arXiv:2403.00490, Mar. 2024.

[19] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, "Large-scale structure of the Universe and cosmological perturbation theory," , vol. 367, no. 1-3, pp. 1–248, Sep. 2002.

[20] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, "The effective field theory of cosmological large scale structures," *Journal of High Energy Physics*, vol. 2012, p. 82, Sep. 2012.

[21] M. White, "The Zel'dovich approximation," , vol. 439, no. 4, pp. 3630–3640, Apr. 2014.

[22] M. McQuinn and M. White, "Cosmological perturbation theory in 1+1 dimensions," , vol. 2016, no. 1, pp. 043–043, Jan. 2016.

[23] M. Pietroni, "Structure formation beyond shell-crossing: nonperturbative expansions and late-time attractors," , vol. 2018, no. 6, p. 028, Jun. 2018.

[24] A. Lewis and A. Challinor, "CAMB: Code for Anisotropies in the Microwave Background," Astrophysics Source Code Library, record ascl:1102.026, Feb. 2011.

[25] W. Cui, L. Liu, X. Yang, Y. Wang, L. Feng, and V. Springel, "An Ideal Mass Assignment Scheme for Measuring the Power Spectrum with Fast Fourier Transforms," , vol. 687, no. 2, pp. 738–744, Nov. 2008.

[26] S. Foreman, W. Coulton, F. Villaescusa-Navarro, and A. Barreira, "Baryonic effects on the matter bispectrum," , vol. 498, no. 2, pp. 2887–2911, Oct. 2020.

[27] Y. P. Jing, "Correcting for the Alias Effect When Measuring the Power Spectrum Using a Fast Fourier Transform," , vol. 620, no. 2, pp. 559–563, Feb. 2005.

[28] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, 1981.

[29] E. Pajer and D. van der Woude, "Divergence of perturbation theory in large scale structures," , vol. 2018, no. 5, p. 039, May 2018.

[30] M. Tegmark, "Measuring Cosmological Parameters with Galaxy Surveys," , vol. 79, no. 20, pp. 3806–3809, Nov. 1997.

[31] M. Marinucci, G. Jung, M. Liguori, A. Ravenni, F. Spezzati, A. Andrews, M. Baldi, W. R. Coulton, D. Karagiannis, F. Villaescusa-Navarro, and B. Wandlet, "The constraining power of the Marked Power Spectrum: an analytical study," *arXiv e-prints*, p. arXiv:2411.14377, Nov. 2024.

[32] F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys, A. Brochard, C. Uhlemann, C.-T. Chiang, S. He, A. Pisani, A. Obuljen, Y. Feng, E. Castorina, G. Contardo, C. D. Kreisch, A. Nicola, J. Alsing, R. Scoccimarro, L. Verde, M. Viel, S. Ho, S. Mallat, B. Wandelt, and D. N. Spergel, "The Quijote Simulations," , vol. 250, no. 1, p. 2, Sep. 2020.

[33] W. R. Coulton, F. Villaescusa-Navarro, D. Jamieson, M. Baldi, G. Jung, D. Karagiannis, M. Liguori, L. Verde, and B. D. Wandelt, "Quijote-PNG: Simulations of Primordial Non-Gaussianity and the Information Content of the Matter Field Power Spectrum and Bispectrum," , vol. 943, no. 1, p. 64, Jan. 2023.

[34] ——, "Quijote-PNG: The Information Content of the Halo Power Spectrum and Bispectrum," , vol. 943, no. 2, p. 178, Feb. 2023.

[35] G. Valogiannis and C. Dvorkin, "Towards an optimal estimation of cosmological parameters with the wavelet scattering transform," , vol. 105, no. 10, p. 103534, May 2022.

[36] M. Peron, G. Jung, M. Liguori, and M. Pietroni, "Constraining primordial non-Gaussianity from large scale structure with the wavelet scattering transform," , vol. 2024, no. 7, p. 021, Jul. 2024.

Acknowledgments

I would like to strongly thank Professor Michele Liguori for closely advising me on this thesis and being so responsive in kindly answering my (many) questions. I would also like to thank Professor Massimo Pietroni and PhD student Matteo Peron at the University of Parma for their crucial work in generating the 1+1 simulations used in this thesis, and for sharing their invaluable knowledge with me. I thank my many friends I met here in Padova, the 'Astronerds' for their camaraderie and endless shenanigans these past 2+ years. I also thank the staff at Bar La Specola for serving me and my friends our ritual breakfast every day while we worked on our theses. Finally, I would also like to thank the staff at INAF for letting me use their microwave. This project made use of the supercomputer 'CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l'Innovazione' funded by the University of Padova Strategic Research Infrastructure Grant 2017.