ONE-POINT MATTER PDF'S BEYOND TOPHAT FILTERS

ONE-POINT MATTER PDF'S BEYOND TOPHAT FILTERS

By ALEXANDER M. KAYSSI, BS

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2024)

Hamilton, Ontario, Canada (Physics & Astronomy)

TITLE:	One-point matter PDF's beyond TopHat filters
AUTHOR:	Alexander M. Kayssi BS (Math & Physics), McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. Sergey Sibiryakov
NUMBER OF PAGES:	xiv, 77

Lay Abstract

In this thesis, I study the statistics of matter distributions in our universe. Their statistics are expected to be dependent on how we choose to average over the matter content in a particular patch in the sky. For this work, I wrote `Python` code which calculates the probabilities of these matter densities given an arbitrary averaging scheme. After examining several cases, it was found that the statistics were qualitatively similar to one another. This suggests an overall universality in the statistics of matter densities in our universe.

Abstract

In this thesis, I studied the one-point probability distribution function (PDF) for averaged matter densities over spherical cells, which can be used to non-perturbatively probe the large-scale structure of our universe. The PDF depends on a function, known as the filter/window function, which takes some weighted average over the observed matter density within each cell. This averaging allows one to study the density field as some smoothed function rather than discrete points. In order to consider filters of different kinds, the PDF's are constructed numerically using `Python` code. The PDF is analytically modeled using a path integral framework. By considering a family of radial window functions interpolating between the TopHat and Gaussian filters in coordinate space, I investigated the sensitivity of the PDF to the shape of the window function. It was found that the sensitivity is rather mild suggesting that the PDF is robust against the precise choice of the filter. Effective field theory (EFT) corrections were included and used to examine how sensitive different filters are to short-scale physics. Similar to the PDF, the effects coming from short-scale physics appeared weakly dependent on the choice of filter, regardless of how smooth the filter's boundary was.The contribution coming from aspherical fluctuations to the collapse dynamics of the cell were computed by comparing the numerical PDF to high-resolution N-body simulations. It was found that this contribution factorizes as

a prefactor to the PDF, which is redshift independent, with the exception of smaller sized cells which display some mild redshift dependent shifting. These discrepancies are thought to be associated with two-loop corrections to the PDF. We expect this model to be flexible enough to study beyond the ΛCDM model and act as a probe for new fundamental physics.

To my family, my mentors, and Sammi

Acknowledgements

I would like to thank my supervisor, Dr. Sergey Sibiryakov, for his patience, mentoring, and support throughout my master's experience. You have taught me to view and think about physics in an entirely new light and provided me with an invaluable set of skills which I will continue.

I would also like to thank Dr. Anton Chudaykin, Dr. Cliff Burgess, Dr. James Wadsley, Dr. Cora Uhlemann, and Dr. Elena Pinetti for the many instructive and guiding discussions about the research and its applications.

To my mom, dad, and brother, thank you for acting as my anchor throughout this process. It is thanks to your continual and constant support which helped me get to where I am today.

Thank you to all my friends for not only being by my side, but also providing me with an excuse to enjoy a coffee/snack break throughout random parts of the day.

I would further like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support throughout this work and the Digital Research Alliance of Canada for providing me access to their computational resources.

Table of Contents

List of Figures

List of Tables

Notation, Definitions, and Abbreviations

Notation

$P(k)$	Linear power spectrum
$g(z)$	Linear growth factor
$\delta(\mathbf{x})$	Non-linear density contrast
$\delta_L(\mathbf{x})$	Linear density contrast
$\mathcal{P}(\delta_*)$	Probability distribution function
$\mathcal{P}_{\mathbf{sp}}(\delta_*)$	Spherical probability distribution function
$\mathcal{A}_0$	Monopole prefactor
$\mathcal{A}_{asp}$	Aspherical prefactor

Abbreviations

LSS	Large-scale structure
EFT	Effective field theory
SPT	Standard cosmological perturbation theory
PDF	Probability distribution function
CiC	Counts-in-cells
DST	Discrete sine transform
SLSQP	Sequential least squares programming
ΛCDM	Lambda cold dark matter

Declaration of Academic Achievement

I declare that the following research presented in this thesis is my own work with recognized contributions from Dr. Sergey Sibiryakov and Dr. Anton Chudaykin.

Chapter 1

Introduction

Our universe contains everything we can see; all of the space, matter, and energy that we can observe. When observed at large enough scales, the distribution of matter in our universe appears to be homogeneous. This homogeneity, however, is not exact. Dense distributions of matter, such as galaxies and galaxy clusters, are often surrounded by vast regions of empty space. This large-scale structure (LSS), also known as the cosmic web, is thought to be sourced by small matter fluctuations during the inflationary era; a period in which the universe underwent a rapid accelerative expansion. Over time, these fluctuations interacted gravitationally, forming the current LSS we see today. Recently, there has been an abundance of observational data coming from the growing number of cosmological surveys mapping LSS of the Universe. This data does carry vital information on cosmological parameters, dark matter properties, and fundamental physics occurring at the time of the early universe. The importance of these data becomes even more apparent as we continue through the era of 'high-precision cosmology' where the uncertainties in the measurements are sub-percent [1].

Remarkably, despite its complexity, we can model the history and dynamics of our universe using only six free parameters[1]; Ω_b, Ω_m, h, n_s, σ_8, and τ. Here, Ω_b (Ω_m) represent the baryonic (matter) density parameter. This variable describes the ratio of the baryonic (matter) density, ρ_b (ρ_m), to the total density ρ_{tot} of the universe today[2]. Note that baryonic density includes all matter made from protons, neutrons, and electrons which we can observe and interact with. The matter density includes all baryons and dark matter, the latter of which is a type of matter which is not visible and does not interact with any known force or material, including itself, except through gravity. The parameter h is related to the Hubble constant, H_0, characterizing the rate of expansion of the universe, and is conventionally represented in the form[3] $H_0 = h \cdot 100 \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}$. The parameter n_s represents the 'scalar spectral index' and characterizes initial matter density fluctuations. The parameter σ_8 describes the variance of matter on scales of 8 h^{-1}Mpc. In other words, by examining the distribution of matter over scales of 8 h^{-1}Mpc, σ_8 then tells us how spread the recorded values are from the average. Finally, τ represents the 'optical depth', which describes how opaque the universe is as light travels through it over time. Larger values of τ imply a more opaque medium while smaller values imply a more transparent medium. This is closely related to the mean-free path of light, defined as the average distance traveled by a particle before experiencing some interaction. All together, this model of the universe is referred to as the 'lambda cold dark matter (ΛCDM)

[1]This holds true when we consider a universe with zero spatial curvature. If we have non-zero curvature, then we can characterize the universe with more parameters.

[2]ρ_{tot} only describes the total density in our universe assuming we have zero spatial curvature. This assumption is consistent with current data and observations about our universe.

[3]Here, Mpc stands for 'megaparsec'- a standard unit in cosmology defined as 1 Mpc=3.086×10^{22} meters.

cosmology'. These parameters have been well studied and measured for our own universe using data collected from satellites [2]. We may, however, more generally define a given 'cosmology' as a universe with any specified values for these parameters.

Numerous models have since come out which extend and modify the above ΛCDM one. Such models are known as 'beyond ΛCDM' and consider additional effects on cosmology such as non-Gaussian matter distributions, new gravitational affects, and others [3]. One type of model considered in beyond ΛCDM is parametric dark energy. This model describes dark energy, a type of energy expected to source the expansion of our universe, using a non-standard 'equation of state', w. The equation of state relates the pressure, P and energy density, ρ_Λ, of dark energy through the relation $P = w\rho_\Lambda$. The current ΛCDM model assumes $w = -1$ while parametric dark energy may consider non-constant and dynamical forms for w.

In order to extract useful information from cosmological data, one needs a complete understanding of the non-linear collapse dynamics of matter within the specified cosmology. Although numerical methods and simulations have commonly been used to study these regimes, progress toward analytic methods has been developing at a rapid rate in an attempt to reduce computational strain and provide a fundamental understanding of the physics at play. Common analytic methods often relied on the use of perturbation theory to study cosmic fluctuations on top of a uniform and homogeneous background. These fluctuations were known to be small in the early universe, allowing perturbation theory to act as a successful and powerful tool. Its application to the cosmic microwave background (CMB) in the early universe is a well-known example of this. The CMB describes the background radiation of microwaves in the universe today. This background was sourced by high-energy photons in the early

universe which became less energetic as the universe expanded, eventually turning into the low-energy microwave radiation we see today. The CMB allows us to see the distributions of temperature in the early universe. Remarkably, it was found that the temperature was uniform, with an average value of $T = 2.7255 \pm 0.0006$ K [4] and fluctuations from point to point of $\sim 10^{-5}$ [5]. These small temperature fluctuations explain why linear perturbation theory has been so successful in studying the CMB. As these thermal fluctuations are sourced by the same quantum fluctuations as in LSS, the CMB has been able provide us with a plethora of data on the history of our universe.

In the case of LSS, one treats the matter density contrast, $\delta(\mathbf{x}) = \rho(\mathbf{x})/\rho_{\text{univ}} - 1$, where $\rho(\mathbf{x})$ is the matter density at a given position and ρ_{univ} is the spatial averaged density, as being small. With LSS, however, the small matter fluctuations formed in the early universe have now grown well past the regime where perturbative approaches can be used. This has called for new analytic methods which can probe into the non-linear regime of matter perturbations [6].

One of the earliest attempts to study non-linear effects of $\delta(\mathbf{x})$ was done by Zel'dovich [7]. He found an approximate solution to $\delta(\mathbf{x})$ by applying Lagrangian perturbation theory, which traces particle trajectories or fluid elements over position and time [8]. By restricting the system to small displacements around the initial position, corresponding to small density fluctuations around the homogeneous background, Zel'dovich found an analytic expression describing the growth of perturbations. While quantitatively different, the approximation can qualitatively be used to describe large, non-linear perturbations as well. More recent approaches, such as standard cosmological perturbation theory [8] (SPT), have improved this approach by

modeling dark matter dynamics as a pressureless fluid, which interacts only through gravity. Using SPT, one can then compute various statistical quantities, such as the 2-point correlator, $\xi(\mathbf{x}_1, \mathbf{x}_2)$, is defined by

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

Here the angled brackets, $\langle . \rangle$, denote an ensemble average, ie. an averaging over all possible configurations of the system. This correlator represents the probability of finding two overdensity regions, ie. matter densities, galaxies, etc., separated by a distance $r = |\mathbf{x}_2 - \mathbf{x}_1|$. Note that we can also examine the Fourier transform of $\xi(\mathbf{x}_1, \mathbf{x}_2)$. This quantity, denoted by $P(k)$, is known as the 'power spectrum'. The 2-point correlator can also be generalized as an N-point correlators, $\xi(\mathbf{x}_1, ..., \mathbf{x}_N)$, defined by

$$\xi(\mathbf{x}_1, ..., \mathbf{x}_N) = \langle \delta(\mathbf{x}_1)...\delta(\mathbf{x}_N) \rangle \, ,$$

which are also studied throughout cosmology. Once implemented, SPT introduces higher-order corrections to these correlators, with the lowest-order correction often referred to as the 'one-loop correction'. This one-loop correction captures the non-linear behavior of the fluctuations and can extend the theory past the linear regime.

While SPT has a number of advantages, there are several drawbacks preventing it from fully being able to describe LSS. As discussed in [9], the large-scale physics of most physical systems is expected to be independent of how its short-scale physics behaves. However, this is not true in SPT as the short-scale contributions to the system are large. As a result, SPT alone cannot be expected to correctly describe LSS. One successful method which is capable of properly capturing the effects of short-scale

physics is 'effective field theory of LSS' (EFTofLSS) [10, 11]. EFTofLSS addresses the issue of small-scale physics by introducing a window function (or filter) that smooths observable quantities, such as densities and velocities, over a large length scale. By first smoothing our quantities over a large length scale, then applying our perturbative expansion, the effects coming from small scales can be systematically accounted for using only large-scale variables. This process, often described as 'integrating out' the short-scale physics, now treats the previously described pressureless fluid as an effective fluid, with some effective viscosity. This is to say that the effects of the microscopic, short-scale physics of the fluid are now captured as an effective viscosity in the system. Once EFTofLSS is applied to the system, it introduces additional correction terms to the correlators, known as 'counterterms'. Inclusion of the counterterms accounting for short-scale physics is referred to as 'renormalization'. This allows us to probe mildly non-linear scales $k_{\rm NL} \lesssim 0.3\,h{\rm Mpc}^{-1}$, where k is the wavenumber defined in the Fourier transform. Since then, developments in its renormalization procedures [12, 13], effects on the two-point correlator (power spectrum), three-point correlator (bispectrum) [14, 15], and higher order corrections to the power spectrum [16] have been studied.

While perturbative methods are commonly used to study non-linear matter clustering, several non-perturbative approaches have been developed in the past. One well-known example which does this is the Press-Schechter formalism [17]. In this formalism, one can study present-day non-linear structure using the small, Gaussian distributed density contrast in the early universe, $\delta_0(\mathbf{x})$. As this density fluctuation is initially small, one can use perturbation theory to examine the growth of structure in the linear regime. Doing this, one finds that the linearly extrapolated density contrast,

$\delta_L(\mathbf{x}, t)$, known as the 'linear density contrast', evolves as $\delta_L(\mathbf{x}, t) = g(t)\delta_0(\mathbf{x})$. Here, $g(t)$ is defined as the 'linear growth factor' and carries all of the time dependence for how the initial density perturbation evolves. We emphasize here that $\delta_L(\mathbf{x}, t)$ cannot describe density fluctuations in present time, as non-linear effects become relevant. Now, as matter collapses under its gravitational attraction, its density will continue to increase until it has reached a state of virial equilibrium, in which the collapsing matter forms a stable bounded object. Such bounded objects are said to be 'virialized'. In order to classify $\delta_L(\mathbf{x}, t)$ as virialized in the present-day, its smoothed density contrast, $\delta_L(\mathbf{x}; R)$ is first calculated. Similar to EFTofLSS, this smoothed density comes from convolving the density field with some window function over some effective length scale, R. Note that the motivation to smooth over the density fields is to allow us to conveniently work with smooth fields. For example, if one was to examine the distribution of galaxies in our sky, one may describe their locations using discrete points. However, galaxies are not discrete point-like objects as their matter content is spread over some finite size. So to overcome this issue, they are often smoothed over some scale and made into smooth distributions. According to linear theory, once the smoothed linear density contrast is above some critical value, δ_c, ie. $\delta_L(\mathbf{x}; R) \geq \delta_c$, it is then assumed to be virialized, representing a non-linear bounded object in the current-day. Regions where $\delta_L(\mathbf{x}; R) \geq \delta_c$ are often referred to as 'peaks'. Through the Press-Schechter formalism, the number density of bounded objects with a mass M is related to the probability of finding a peak. This process involves constructing a probability distribution function (PDF) for density contrasts at a fixed smoothing scale, then integrating over all probabilities in which $\delta_L \geq \delta_c$.

While this approach generally agrees well with simulations, it requires manually inserting a factor of 2 into the equations in order to get the correct fraction of virialized objects in the universe. Without this, Press-Schechter would be underestimating the fraction of virialized objects by half. However, this factor could not be rigorously argued or justified in their original work.

An alternate formalism which non-perturbative examines non-linear density regimes is the BBKS formalism [18]. This formalism analytically examines the number density, N-point correlators, and other statistical properties of non-linear overdense structures in a Gaussian random field. In order to construct the number density of bounded structures, the BBKS formalism assumes that this must be equal to the number of peaks in the system. This approach, however, suffers from an issue known as the 'cloud-in-cloud' problem. This occurs when larger sized overdensities contain within them smaller sized overdensities, each of which are virialized on their respective length scale. With cases of virialized objects being contained within one another, it is possible that the statistics of bounded structures are not accurately be studied. The cloud-in-cloud problem was also found to exist in the Press-Schechter approach, explaining the missing factor of two. It wasn't until the excursion set formalism [19], also known as the 'extended Press-Schechter formalism', which could avoiding the cloud-in-cloud problem. The issue faced in previous formalisms was that it does not consider the scenario in which $\delta(\mathbf{x}; R_1) < \delta_c$ but $\delta(\mathbf{x}; R_2) \geq \delta_c$ for $R_1 < R_2$ [20]. In other words, previous methods did not take into account that a small underdense region may be enclosed by a larger overdense region. In order to account for this, the excursion set formalism first finds the largest smoothing scale corresponding to a bound object. It then moves down the length scale to avoid miscounting any overdense

regions. This process was also found to reproduce the factor of two in Press-Schechter without imposing it by hand.

In this thesis, we compute the one-point probability distribution function (PDF) of the matter density field. This observable is closely related to another non-perturbative formalism known as 'counts-in-cells (CiC) statistics' [21]. In this formalism, one divides up the density field into spherical cells of radius r_* then computes the averages of the field in each cell using some window function. Given a collection of cell averaged densities, a PDF for measuring a given density can be constructed. Note that the deviation of the averaged density from the mean density of the universe does not need to be small, hence the PDF carries information about non-perturbative dynamics.

Up to now, most studies have relied on the TopHat filter, with some exceptions [22]. This filter provides a uniform weight to matter inside the cell, and zero weight outside. For example, if one wanted to apply the TopHat filter to a galaxy cluster, then the radius of the filter would correspond to the radius of the cluster. This would ensure that the TopHat contains the entire structure within it, and that averaging over the cell corresponds to averaging the matter content of only the cluster. One can, however, also consider the statistics of the PDF for non-TopHat filters. For example, applying a Gaussian weight would prioritize the matter distribution closer to the center of the cell, and less near its boundary. Note that in the case of a general filter, the average density may not only depends on the field content within the cell, but outside it as well, as seen in the Gaussian case. We note here that, while we work with spherical cells to construct statistics for LSS, the actual matter distribution is not spherically symmetric. Rather, it is composed of filaments and sheet-like structures. We will examine the effects of aspherical fluctuation using both analytic techniques

from perturbation theory as well as N-body simulations. While not discussed in this work, smoothing over cylindrical cells is also possible and was investigated in [23].

Here we apply a path integral formalism for the CiC method following [24, 25], generalizing it to the case of an arbitrary filter. In this approach, the PDF is described as a path integral over all possible density configurations, which is evaluated in the saddle-point approximation. This approximation perturbatively expands the integral around the solution that extremizes the action of our system, known as the saddle-point configuration, using a small expansion parameter. In this work, the linear matter variance of the matter density field within a cell plays the role of the expansion parameter. While perturbatively expanding the integral, the saddle-point solution dictates the exponential behavior of the PDF while next-to-leading order corrections were found to contribute to the prefactor of the PDF, corresponding to the one-loop EFT corrections in the background of the saddle-point configuration.

The work done in [24, 25] limits itself only to the use of TopHat filters. While this case greatly simplifies the calculations, resulting in analytic expressions for the saddle-point configuration and prefactor, there are a number of limitations in its uses. One of them is that the spherical PDF becomes independent of the statistical properties of the density field inside or out the cell. Instead, it only depends on the power spectrum and the spherical collapse dynamics in a given cosmology. While this simplicity may be advantageous in some cases, it removes any sensitivity of the matter distribution in the neighborhood of the cell. This can become problematic when studying objects such as voids, where the matter distribution along the walls of the underdensity can play an important role in identifying and classifying them. Furthermore, the TopHat filter introduces a sharp boundary of the cell in position

space. This transition leads to momentum modes of all scales being included when examined into Fourier space, making it inconvenient to work with [20]. Furthermore, certain theories, such as excursion set theory [19], rely heavily on Sharp-k filters over the TopHat, while some halo clustering models [26] prefer Gaussian's. Recent work studying the effects of different filters on the mass function of primordial black holes [27] and galaxy parameters [28] have also been conducted, suggesting an interest in using filters outside of TopHat.

Another motivation to study other filters has to do with how they limit effective field theory (EFT) corrections of small scale physics. As discussed in [24], counterterm corrections using a TopHat filter can change the PDF up to $\sim 30\%$ near the tails of the distribution. Thus, it would be informative to study the influence that different filter have on the counterterm and whether some choices can reduce this dependence on EFT corrections.

The goal of this thesis is to develop a numerical pipeline to construct the PDF given an arbitrary filter. To do this, we discretize our system over a lattice space and apply a numerical minimization procedure to solve for the saddle-point configuration. Contributions coming from leading order corrections are also numerically implemented. We find that our code is able to recreate the TopHat PDF within 3% error in the extreme overdense and underdense regions. Comparing the TopHats results with PDF's from non-TopHat filters, we find that qualitatively there is no difference between any of the cases. A similar conclusion is reached when examining different filters sensitivity to the short-scale physics of EFT corrections.

This thesis is organized as follows. In chapter 2 we review the path integral formulation developed in [25, 24] to construct the 1-point PDF. In chapter 3 we discuss the

discretization scheme needed to find the PDF numerically. In chapter 4 we validate our pipeline and construct PDF's of non-TopHat filters. In chapter 5 we investigate perturbatively the effects coming from leading-order corrections in our saddle-point approximation and their sensitivity to short-scale physics. In chapter 6 we examine the higher order, aspherical fluctuations coming from our saddle-point expansion by comparing our results with N-body simulations. Chapter 7 is dedicated to the discussion of our results and future directions.

Several appendices contain supplementary information. Appendix A lists the conventions used in the thesis. Appendix B provides derivations of results used in chapter 2 and 3. Finally, Appendix C gives details on our convergence tests of the code.

Chapter 2

Path integral for the PDF

Consider a density contrast averaged with a spherically symmetric window function, with a characteristic radius of r_*,

$$\bar{\delta}_W = \int \frac{d^3x}{r_*^3} \tilde{W}\left(r/r_*\right) \delta(\mathbf{x}) = \int_{\mathbf{k}} W(kr_*)\delta(\mathbf{k}), \tag{2.0.1}$$

where $\delta(\mathbf{x}) = \rho(\mathbf{x})/\rho_{\rm univ} - 1$ where $\rho(\mathbf{x})$ is the density field and $\rho_{\rm univ}$ is the spatially averaged density of the universe. $\tilde{W}(r/r_*)$ is the position space window function obeying the normalization condition $\int \frac{d^3x}{r_*^3}\tilde{W}(r/r_*) = 1$, and $W(kr_*)$ is its Fourier transform. Conventions about the above notation, Fourier transform, and more can be found in Appendix A. We emphasize here that $\delta(\mathbf{x})$ is the non-linear density contrast seen today, which is sourced by the gravitational dynamics of matter throughout the evolution of the universe. Here we also define the TopHat window function as

$$\tilde{W}_{\rm th}(r/r_*) = \frac{3}{4\pi}\Theta_{\rm H}\left(1 - \frac{r}{r_*}\right) \Longleftrightarrow W_{\rm th}(kr_*) = \frac{3j_1(kr_*)}{kr_*}, \tag{2.0.2}$$

where $\Theta_{\rm H}(x)$ is the Heaviside function and $j_1(x)$ is the spherical Bessel function. Here we assume the initial conditions in the far past for the density perturbations to be adiabatic and Gaussian such that their two-point correlator can be written as

$$\langle \delta_i(\mathbf{k})\delta_i(\mathbf{k}')\rangle = (2\pi)^3\delta_D(\mathbf{k}+\mathbf{k}')g^2(z_i)P(k), \tag{2.0.3}$$

where δ_D is the Dirac delta-function, $P(k)$ is the linear power spectrum at $z = 0$, and $g(z)$ is the linear growth factor which evolves the state from an initial redshift z_i. We emphasize here that g^2 plays the role of a small expansion parameter which will be used when carrying out our saddle-point approximation. In reality, the linear matter variance (introduced in chapter 4) is the true expansion parameter with g^2 acting as a book-keeping term to keep track of our perturbation order.

We will however find it more convenient to rescale the density field to redshift z using the linear growth factor,

$$\delta_L(\mathbf{k}, z) = \frac{g(z)}{g(z_i)}\delta_i(\mathbf{k}). \tag{2.0.4}$$

This will be referred to as the 'linear density contrast' and will omit the explicit redshift-dependence to simplify notation.

Given that we are assuming a Gaussian random field for $\delta_i(\mathbf{k})$, then we know the linear density contrast also forms a Gaussian random field. To construct a PDF, we integrate the Gaussian weight over all initial configurations that produce a given final density contrast, δ_*. This condition is implemented by inserting into the integral a

Dirac delta function;

$$\mathcal{P}(\delta_*) = \mathcal{N}^{-1} \int \mathcal{D}\delta_L \exp\left\{ -\frac{1}{g^2} \left[\int_{\mathbf{k}} \frac{|\delta_L(\mathbf{k})|^2}{2P(k)} \right] \right\} \delta_D(\delta_* - \bar{\delta}_W[\delta_L]), \tag{2.0.5}$$

where $\mathcal{N}$ is a normalization constant. Expanding the Dirac delta function using its integral representations, one can rewrite the PDF as [25, 24]

$$\mathcal{P}(\delta_*) = \mathcal{N}^{-1} \int_{-i\infty}^{i\infty} \frac{d\lambda}{2\pi i g^2} \int \mathcal{D}\delta_L \exp\left\{ -\frac{1}{g^2} \left[\int_{\mathbf{k}} \frac{|\delta_L(\mathbf{k})|^2}{2P(k)} - \lambda(\delta_* - \bar{\delta}_W[\delta_L]) \right] \right\}, \tag{2.0.6}$$

where λ is a free variable which, when integrated over, reintroduced the Dirac delta function enforcing the constraint $\bar{\delta}_W = \delta_*$. This constraint ensures that the averaged density contrast of the cell, $\bar{\delta}_W$, is equal to the density contrast we are interested in δ_*. Note that λ can also be seen as the Lagrange multiplier enforcing this non-linear constraint in our equations. Here we also want to emphasize that $\bar{\delta}_W$ carries all information about the evolution of the density contrast from its initial value. This evolution is carried out by the non-linear collapse dynamics of the cell, which we use to relate the non-linear density contrast, $\delta(\mathbf{x})$, to the linear density contrast, $\delta_L(\mathbf{x})$. Solving the integral using the saddle-point approximation, we expect the form of the PDF to follow

$$\mathcal{P}(\delta_*) = \exp\left\{ -\frac{1}{g^2} \left(\alpha_0 + \alpha_1 g^2 + \alpha_2 g^4 + ... \right) \right\}. \tag{2.0.7}$$

The leading order term α_0 corresponds to the saddle-point solution which minimizes the exponential term in (2.0.6). The next-to-leading term $\alpha_1 g^2$ comes from the Gaussian integral around the saddle-point configuration. This term corresponds to the one-loop correction of the saddle-point and introduces a $1/g$ factor to the PDF through a logarithmic term appearing in α_1. Similarly, $\alpha_2 g^4$ represents two-loop corrections.

In this paper, we consider only up to one-loop corrections.

At this point, it will be useful to impose a symmetry in our system which will help us look for saddle-point configurations. By enforcing that the filter is spherically symmetric, and thus depends only on the magnitude of the momentum[1], k, one finds that the most probable configuration, $\hat{\bar{\delta}}_L(\mathbf{k})$, is also spherically symmetric. Given such a symmetry, we may construct a one-to-one mapping relating the linear and non-linear density fields by

$$\bar{\delta}_L(R) = F(\bar{\delta}(r)) \Longleftrightarrow \bar{\delta}(r) = f(\bar{\delta}_L(R)), \tag{2.0.8}$$

where the functions f and F come from equations governing the spherical collapse dynamics of a spherical density perturbation and

$$\bar{\delta}(r) = \frac{3}{r^3}\int_0^r dr_1 r_1^2 \delta(r_1), \qquad \bar{\delta}_L(R) = \frac{3}{R^3}\int_0^R dR_1 R_1^2 \delta_L(R_1). \tag{2.0.9}$$

Here, R is the comoving radius of cell at some early time in the universe while r is the comoving radius of cell as it undergoes the collapse process. Note that as we go back in time to the early universe, then $r \to R$. We may also interpret R and r as the Lagrangian and Eulerian radii of the cell respectively. In the case of an Einstein-de Sitter (EdS) cosmology, defined by $\Omega_m = 1$ and $w = 0$, the collapse dynamics for overdense regions are given by

$$\mathcal{F}_+(\theta) = \frac{9(\theta - \sin\theta)^2}{2(1-\cos\theta)^3} - 1, \tag{2.0.10a}$$

[1]Note that the vector $\mathbf{k}$ is a wavenumber coming from the Fourier transform and does not refer to any physical momenta coming from velocities of the system. However, following the EFTofLSS literature, we will refer to it as a momenta.

$$\mathcal{G}_+(\theta) = \frac{3}{20}\left[6(\theta - \sin\theta)\right]^{2/3}, \tag{2.0.10b}$$

where $0 \leq \theta \leq 2\pi$ is a parametric variable used to track time throughout the collapse. Here, $\mathcal{F}_+$ ($\mathcal{G}_+$) represent the averaged non-linear (linear) density contrasts, $\bar{\delta}$ ($\bar{\delta}_L$). From this, we see that the non-linear density go to infinity in finite time ($\theta = 2\pi$), implying that the overdensity collapses to zero size. In reality, matter overdensities do not do this, but rather virialize and form objects of a finite size. This virialization process is characterized by the value of linear density contrast at $\theta = 2\pi$, with $\mathcal{G}_+(2\pi) \approx 1.686$. In other words, by examining the value of $\bar{\delta}_L$, we can determine if an overdensity is virialized or not by checking if $\delta_L(R) \geq 1.686$ (assuming an EdS cosmology). Using the above functions, we may then define the functions $F(x), f(x)$ as

$$F(x) = \mathcal{G}_+(\mathcal{F}_+^{-1}(x)), \qquad f(x) = \mathcal{F}_+(\mathcal{G}_+^{-1}(x)). \tag{2.0.11}$$

For underdense regions, the collapse dynamics are given by

$$\mathcal{F}_-(\theta) = \frac{9(\sinh\theta - \theta)^2}{2(\cosh\theta - 1)^3} - 1, \tag{2.0.12a}$$

$$\mathcal{G}_-(\theta) = -\frac{3}{20}\left[6(\sinh\theta - \theta)\right]^{2/3}. \tag{2.0.12b}$$

with the functions $F(x), f(x)$ defined in the same way.

Using this mapping function, we can write (2.0.1) as

$$\bar{\delta}_W = \frac{4\pi}{r_*^3}\int dR\; R^2\tilde{W}\left[R\left(1 + f\left(\bar{\delta}_L(R)\right)\right)^{-1/3}/r_*\right] - 1. \tag{2.0.13}$$

When solving for a saddle-point solution of (2.0.6), we look for functions $\delta_L(\mathbf{k})$ and λ which minimize the exponent of the integrand. Thus, the problem of looking for saddle-point solutions turns to that of a constrained-minimization problem. To approach this problem, we may define the exponent as,

$$\mathcal{S}_{con} = \int_{\mathbf{k}} \frac{|\delta_L(k)|^2}{2P(k)} - \lambda(\delta_* - \bar{\delta}_W[\delta_L]), \tag{2.0.14}$$

which we will refer to as the *constrained action* in what follows. By taking variations of (2.0.14) with respect to $\delta_L(\mathbf{k})$ and λ, we get the following equations of motion,

$$\frac{\hat{\delta}_L(k)}{P(k)} + \hat{\lambda}\, \frac{\partial \bar{\delta}_W}{\partial \delta_L(k)}\bigg|_{\hat{\delta}_L(k)} = 0, \tag{2.0.15a}$$

$$\delta_* = \bar{\delta}_W[\hat{\delta}_L]. \tag{2.0.15b}$$

The solutions $\hat{\delta}_L(\mathbf{k})$ and $\hat{\lambda}$ which satisfy the above equations then describe our saddle-point configuration. Note that, by this point on, any hatted quantities will be used to describe saddle-point configurations of our system.

To study the prefactor, we first consider small perturbations around the saddle-point solution up to order g^2. This involves expanding our variables into $\delta_L(\mathbf{k}) = \hat{\delta}_L(\mathbf{k}) + \delta_L^{(1)}(\mathbf{k})$ and $\lambda = \hat{\lambda} + \lambda^{(1)}$. Expanding $\delta_L^{(1)}(\mathbf{k})$ into spherical harmonics and substituting into (2.0.6), we find that the PDF factorizes into the contributions with different multipole numbers, ℓ,

$$\mathcal{P}(\delta_*) = \exp\left\{-\frac{1}{g^2}\int_{\mathbf{k}} \frac{|\hat{\delta}_L(\mathbf{k})|^2}{2P(k)}\right\} \cdot \prod_{\ell=0} \mathcal{A}_\ell(\delta_*). \tag{2.0.16}$$

Singling out the contributions of the monopole fluctuations, we define the 'spherical part' of the PDF as

$$\mathcal{P}_{SP}(\delta_*) = \mathcal{A}_0(\delta_*) \cdot \exp\left\{-\frac{1}{g(z)^2}\int_{\mathbf{k}} \frac{|\hat{\delta}_L(\mathbf{k})|^2}{2P(k)}\right\}. \tag{2.0.17}$$

We emphasize here that the function, which we will refer to as the *spherical PDF* is not the true PDF. It does, however, make up a large fraction of the true PDF as it has the strongest dependence on δ_*.

The monopole prefactor $\mathcal{A}_0$ is given by (see Appendix B for derivation)

$$\mathcal{A}_0 = \frac{1}{\sqrt{2\pi g^2}}\sqrt{\frac{\det\left(\frac{\mathbb{I}(k_1,k_2)}{4\pi P(k)}\right)}{-\det\mathcal{H}}}. \tag{2.0.18}$$

Here, $\mathbb{I}(k_1, k_2)$ is the symmetric unit operator in radial k-space

$$\mathbb{I}(k_1, k_2) = \frac{(2\pi)^3}{k_1^2}\delta_D(k_1 - k_2). \tag{2.0.19}$$

and $\mathcal{H}$ is the block matrix

$$\mathcal{H} = \begin{pmatrix} 0 & S(k) \\ S^{\intercal}(k) & \mathcal{O}(k_1, k_2) \end{pmatrix}, \tag{2.0.20}$$

where

$$\mathcal{O}(k_1, k_2) = \frac{\mathbb{I}(k_1, k_2)}{4\pi P(k)} + \frac{\hat{\lambda}}{2\pi}Q_0(k_1, k_2) \tag{2.0.21}$$

and $S(k)$ and $Q_0(k_1, k_2)$ are some known functions (refer to appendix B). Note that the determinants appearing in the monopole prefactor are functional determinants,

and so cannot be evaluated in the usual way. These come from evaluating the leading order, Gaussian path integrals in the saddle-point approximation (see Appendix B).

Before continuing to the aspherical contributions to the prefactor, we will first discuss the redshift dependence of the spherical PDF. Looking at (2.0.17) and (2.0.18), we see that $g(z)$ carries all dependence about the redshift. We also see that $g(z)$ factors out of the prefactor and exponent as $1/g$ and $1/g^2$ respectively. This tells us that in order to evaluate our spherical PDF at any given redshift, we can simply rescale the prefactor and exponent accordingly.

The remaining aspherical contributions to the prefactor can now be defined as

$$\mathcal{A}_{asp}(\delta_*) = \prod_{\ell \geq 1} \mathcal{A}_\ell(\delta_*), \tag{2.0.22}$$

which we will refer to as the *aspherical prefactor*. We do not attempt to calculate it in this work (see [24, 25] for the results for TopHat filters). Rather, we will estimate it by comparing our spherical PDF with N-body simulations (see chapter 6).

Chapter 3

Numerical implementation

As summarized in chapter 1, Ref. [24], derived Euler-Lagrange equation for the saddle-point configuration of (2.0.6). In general, this is a complicated non-linear integral equation, which gets drastically simplified only in the case of the TopHat filter. Solving it for a general filter is impractical. We follow a different route and develop an approach to find the saddle-point configuration, as well as monopole prefactor in (2.0.18) based on numerical minimization. To do this, we have constructed two separate codes which compute the saddle-point profiles and monopole prefactor separately.

3.0.1 Numerical saddle-point

The saddle-point solution $\hat{\delta}_L(k)$ minimizes the Gaussian weight

$$\mathcal{S}[\delta_L] = \int_{\mathbf{k}} \frac{|\delta_L(k)|^2}{2P(k)}, \tag{3.0.1}$$

in (2.0.6), subject to the constraint $\bar{\delta}_W[\delta_L] = \delta_*$. We will refer to this as the *action*.

For a given δ_*, we use the `Scipy` and `Numpy` libraries in `Python` to discretize our position and momentum space over an evenly-spaced lattice using the following parameters,

$$\begin{aligned} &R_{\max} = DR_*, \quad R_{\min} = R_{\max}/N, \quad R_* = r_*(1+\delta_*)^{1/3}, \\ &k_{\max} = N\frac{\pi}{DR_*}, \quad k_{\min} = \frac{\pi}{DR_*}. \end{aligned} \tag{3.0.2}$$

Here N is the number of lattice points, D is some integer, and R_* interpreted as being the initial radius of the cell in the early universe. We have found that different choices of δ_* require different values of N and D to reach a given level of precision. This is discussed in more detail in section 4.

As one applies this discretization procedure, the question of finding the saddle-point configuration reduces to a multidimensional minimization problem of the action subject to non-linear constraints. In order to see this, recall that the saddle-point configuration and constraint condition in chapter 2 were found by extremizing the constrained action. However, as we discretize our functions into arrays, the act of performing functional extremization translates to that of multidimensional minimization.

Once the system has been discretized, we can then study PDF's of any filter given that we are able to implement the window function numerically through the non-linear constraint. Note that due to the constraints non-trivial structure, it sourced many of the challenges and difficulties during the discretization procedure.

In order to solve this constraint minimization problem, we apply the following algorithm:

1. Define the functions $\Delta_L(R) = R\delta_L(R)$ and $\Delta_L(k) = k\delta_L(k)$ which are related to one another through the sine transformation (A.0.5).

2. After discretizing onto the lattice (3.0.2), we construct a function which computes $\mathtt{Y} = (\Delta_L(k_1), ..., \Delta_L(k_N))$ given $\mathtt{X} = (\Delta_L(R_1), ..., \Delta_L(R_N))$. This is done by discretizing (A.0.5) as

$$\mathtt{Y}[\mathtt{X}] = 2\pi\Delta R\, \mathtt{DST}[\mathtt{X}], \tag{3.0.3}$$

   where $\Delta R = (R_{\max} - R_{\min})/N$ and `DST` is the `scipy` discrete sine transform (DST) function. Note that the scipy `DST` function is scaled by a factor of two compared to its standard definition.

3. Discretize the action over the lattice as

$$\mathcal{S}[\mathtt{X}] = \sum_{\text{int.}} \frac{1}{(2\pi)^2} \frac{\mathtt{Y}_i^2[\mathtt{X}]}{P(k_i)}, \tag{3.0.4}$$

   where $\sum_{\text{int.}}$ denotes the Simpson numerical integration scheme.

4. Discretize the constraint function as

$$\bar{\delta}_W[\mathtt{X}] = -1 + \frac{4\pi}{r_*^3} \sum_{\text{int.}} R_i^2 \mathtt{W}\left[\frac{R_i}{r_*}\left(1 + \mathtt{f}[\mathtt{Avg}[\mathtt{X}]]\right)^{-1/3}\right] \tag{3.0.5}$$

   where `W` and `f` are numerical implementations of the window function and collapse mapping ($f(x)$ in (2.0.11)) respectively. Note that in order to solve for $f(x)$ numerically, we used a root finding algorithm to construct the inverse function $\mathcal{G}_{\pm}^{-1}(x)$. The function `Avg[X]` numerically computes (2.0.9) given the

array `X`. Note that `Avg` takes in the array `X` and outputting an array of the form $(\bar{\delta}_L(R_1), ..., \bar{\delta}_L(R_N))$.

5. Select a value for δ_* and construct an initial array, $\texttt{X}_i$ representing the initial guess for optimization scheme.

6. Use the `minimize` function under the `scipy.optimize` library, minimize (3.0.4) subject to the non-linear constraint $\bar{\delta}_W = \delta_*$. For the case of non-linear constraints, the sequential least squares programming (SLSQP) solver is used as our optimizer.

This algorithm then returns the optimized array, $\hat{\texttt{X}}$, representing the discretized function for $R\hat{\delta}_L(R)$, at a given δ_* for any specified window function. Using values of N and D ranging from $300 \leq N \leq 450$ and $20 \leq D \leq 30$, this algorithm currently takes on average ~ 18 hours to complete.

Given the saddle-point configuration, $\hat{\delta}_L(R)$, we are able to compute the Lagrange multiplier which we need for the prefactor (see Eq. (2.0.21)). To see this, we note the following equality,

$$\frac{d \ln \mathcal{P}}{d\delta_*} = \frac{\hat{\lambda}}{g^2}, \tag{3.0.6}$$

which comes from differentiating (2.0.6). Substituting in (2.0.16), we find that the Lagrange multiplier, up to leading order in our saddle-point expansion, is given by

$$\hat{\lambda} = -\frac{d\mathcal{S}[\hat{\delta}_L]}{d\delta_*}. \tag{3.0.7}$$

Thus, (3.0.7) tells us that given a sufficiently fine grid of points over δ_*, the Lagrange multiplier can be computed via numerical differentiation of the minimized action,

$\mathcal{S}[\hat{\delta}_L]$. In our case, we sample 100 values of $1+\delta_*$ within the range $[0.1, 10]$ which are evenly-spaced on a log-scale. By running the minimization procedure for every $\delta_{*,i}$, the points $(\delta_{*,i}, \mathcal{S}[\hat{\delta}_{L,i}])$ can be interpolated and differentiated over, giving us a way to calculate $\hat{\lambda}$.

3.0.2 Monopole prefactor

Given the optimized profile, $\hat{\mathtt{X}} = (\hat{\Delta}_L(R_1), ..., \hat{\Delta}_L(R_N))$, where $\hat{\Delta}(R_i) = R_i\hat{\delta}_L(R_i)$, we want to construct (2.0.18) numerically. To do this, we apply the following algorithm,

1. Interpolate $\hat{\mathtt{X}}$ over the lattice space defined in (3.0.2) using a cubic spline function.

2. Construct a denser lattice space by fixing $R_{\max}$ but changing N to some[1] $\tilde{N} > N$. Define the array, $\hat{\mathtt{X}}'$, as the interpolated function from step 1 evaluated over this new lattice.

3. Fourier transform $\hat{\delta}_L(R)$ to $\hat{\delta}_L(k)$ through a discrete sine transform (DST) on $\hat{\mathtt{X}}'$.

4. Discretize $\mathbb{I}(k_1, k_2)$ over the lattice into a diagonal $\tilde{N} \times \tilde{N}$ matrix given by

$$\mathbb{I}_{ij} = \frac{(2\pi)^3}{k_i k_j}\frac{\delta_{ij}}{\Delta k}, \tag{3.0.8}$$

   where $i, j = 1, 2, ..., \tilde{N}$ with $k_1 = k'_{\min}$, $k_{\tilde{N}} = k'_{\max}$, and $\Delta k = (k'_{\max} - k'_{\min})/\tilde{N}$. Note that $k'_{\min}$ and $k'_{\max}$ are the minimum and maximum momentum modes in

[1] At this point, the choice to create a denser lattice space may not seem obvious. However, it is required to reach a high enough numerically accuracy. This will be discussed more in later sections.

the new, denser lattice defined in step 2.

5. Discretize $S(k)$ and $Q_0(k_1, k_2)$ by numerically implementing (B.0.8) over the points k_i, k_j in $[k'_{\rm min}, k'_{\rm max}]$.

6. Construct the $(\tilde{N}+1) \times (\tilde{N}+1)$ discretized $\mathcal{H}$ matrix as,

$$\mathcal{H}_{ij} = \begin{pmatrix} 0 & S(k_j) \\ S^{\intercal}(k_i) & \mathcal{O}_{ij} \end{pmatrix}, \tag{3.0.9}$$

   where we have defined

$$\mathcal{O}_{ij} = \frac{2\pi^2}{\Delta k} \frac{\delta_{ij}}{k_i k_j \sqrt{P(k_i)P(k_j)}} + \frac{\hat{\lambda}}{2\pi} Q_0(k_i, k_j). \tag{3.0.10}$$

7. Compute the determinant of $\mathcal{H}_{ij}$ and $\mathbb{I}_{ij}$ using the `scipy.linalg` library in `Python`.

The average runtime of this algorithm is 10 minutes following the same range of values for $300 \leq N \leq 450$ and $20 \leq D \leq 30$.

Chapter 4

PDF's of filters

In this section, we display the results of our numerical procedure described in the previous section for several different filters. Here, the linear power spectrum is generated using the linear Boltzmann solver `CLASS` [29] with cosmological parameters matching that of the `Quijote` N-body simulations [30] (see chapter 6 for the precise values of the cosmological parameters in `Quijote` ΛCDM simulations). While the power spectrum and N-body simulations used in this paper are based on the ΛCDM cosmology, the gravitational collapse mapping (2.0.8) is known to be robust against different cosmologies [31]. Thus, we choose to use the Einstein-de Sitter (EdS) collapse mapping (2.0.10)-(2.0.12) as they have fully analytic expressions. The EdS cosmology is defined as having $\Omega_m = 1$ and $w = 0$, ie. a matter filled universe with no dark energy. Looking first at the TopHat case, we systematically examine the error in the code by comparing the numerical results with the analytical solutions obtained in [24]. This allows us to identify suitable values of the lattice parameters (N, D) from (3.0.2). Afterwards, we apply the code to several window functions and compare with the TopHat.

4.0.1 Validation of numerical procedure

In order to test the accuracy of the numerical solutions, we first examine the case of the TopHat filter and compare with the analytic solutions derived in [24]. For the saddle-point profile, $\hat{\delta}_L(k)$, this solution is given by

$$\hat{\delta}_L(k) = \frac{F(\delta_*)}{\sigma^2_{L,R_*}} P(k) W_{\rm th}(kR_*), \tag{4.0.1}$$

where σ^2_{L,R_*} is the linear density variance defined in (4.0.3). After implementing (2.0.2) into our code[1], we are able to construct the numerical solutions for our saddle-point configuration. We choose the radius of the cell to be r_* =10 Mpc/h and our lattice parameters $(N, D) = (450, 30)$ for $\delta_* \leq -0.4$ and (300,20) for $\delta_* > -0.4$. The choice of using different parameter spaces for different δ_* comes from the need to ensure the numerical convergence. For more details on our convergence tests, see Appendix C.

In Fig. 4.1 we present the results of the saddle-point profiles, $\hat{\delta}_L(R)$, along with residuals comparing the numerical data with the known analytical expressions. Overall, we see that both methods are in good agreement with one another, with the residuals remaining less than 1-2% within the center of the profiles. This implies that our numerical minimization indeed captures the right saddle-point configuration. Note that as the radius grows in the linear profiles, so to does the error. This does not, however, imply that our pipeline fails in this regime, but instead comes from the fact that we are examining the relative error of the two solutions. As the profile asymptote to zero for large R, any small deviations from numerical result may

[1] In order to construct $\Theta_H(x)$ in such a way that it would be compatible with the `minimize` function, it had to be approximated as a smooth curve using the sigmoid function.

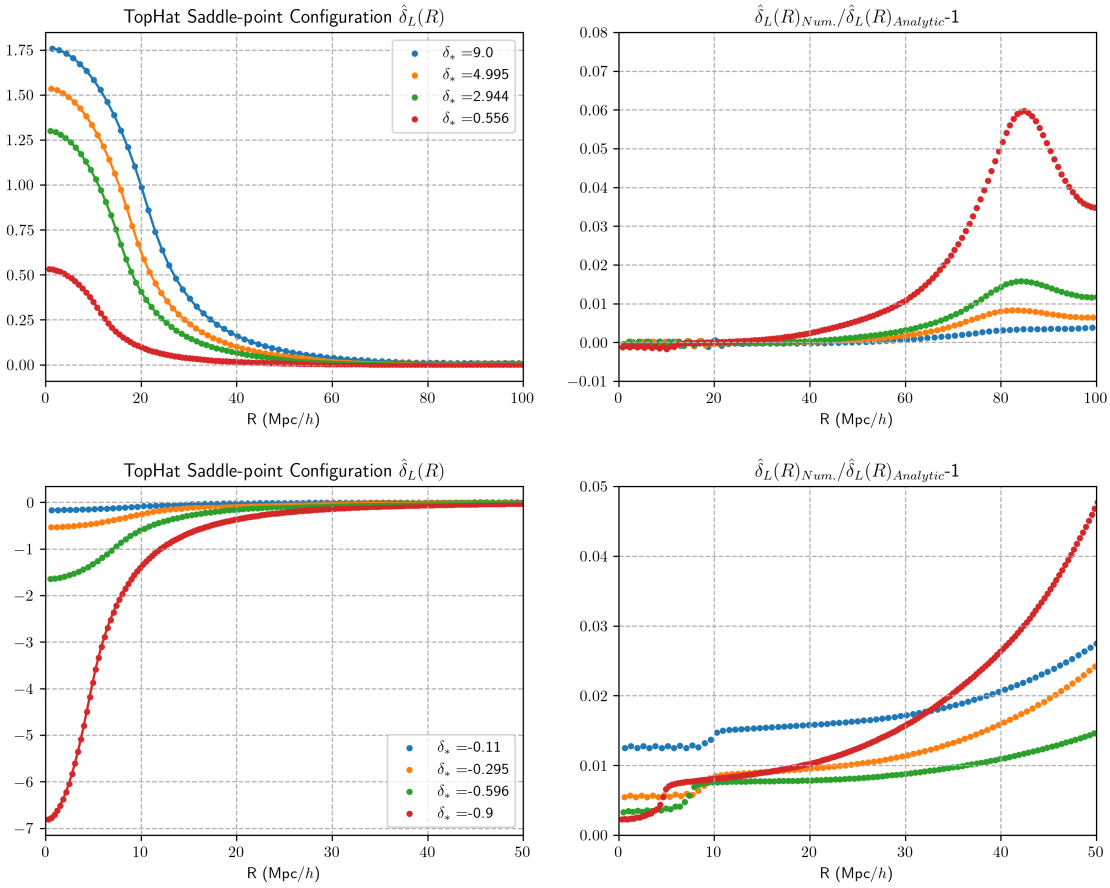

Figure 4.1: Comparison between the numerical (dots) and analytical (solid line) saddle-point linear profiles for several values of δ_*. Left panels show the saddle-point profiles while the right panels display the residuals. Overdense profiles are given in the top panel, with underdense profiles in the bottom.

translate to a large relative error. So, this growing error is not expected to be of concern if it occurs at the tails of the profile.

The virialization of the cell occurs when the linear profile exceeds a critical density $\delta_c = 1.686$ and shell-crossing occurs[2]. In this case, shell-crossing occurs for $\delta_* \gtrsim 7$, which is in agreement with [24]. Surprisingly, we find that the output of the code

[2]The critical density $\delta_c = 1.686$ comes from the assumption that we are in an EdS universe. While its value slightly changes in the ΛCDM cosmology, the change is very small as δ_c is known to be weakly dependent on cosmology.

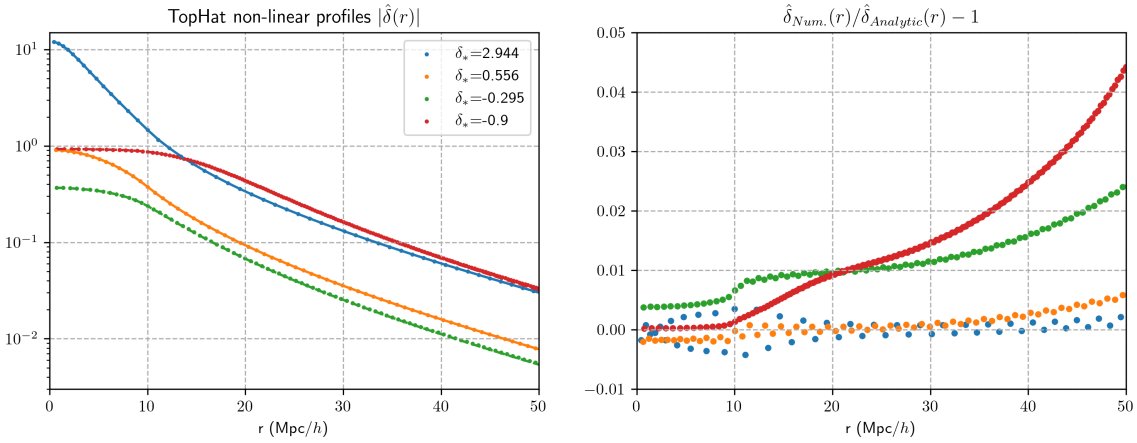

Figure 4.2: *Left panel:* Comparison between the numerical (dots) and analytical (solid/dashed line) non-linear saddle-point profiles $\hat{\delta}(r)$ for several values of δ_*. Note that underdense profiles (green and red) are reflected about the x-axis onto the same quadrant as the overdense profiles (blue and orange). *Right panel:* Residuals of the numerical results to the analytical.

remains robust even after virialization, producing smooth profiles matching the analytic solutions despite the break down of the spherical collapse mapping, $F(\bar{\delta}(r))$, for $\delta_* \geq \delta_c$. This smooth continuation of the analytic solutions may be explained by the due to the slow transfer of information from the center of the cell to its boundary. As the averaged cell density contrast reaches its critical value, δ_c, the matter content at its center would be the first to virialize. However, matter near the boundary may not pick up on this effect until much later due to their slow velocities. As such, information about virialization may take longer to imprint itself across the cell, allowing the analytic solutions to hold even when $\delta_* \geq \delta_c$ [24]. However, it is not entirely clear why the numerics remain robust in this case.

We also display in Fig. 4.2 the numerical and analytical non-linear saddle-point profiles, $\hat{\delta}(r)$, with their residuals. Just like the linear profiles, we find that the numerical result agrees with the analytics to within $1 - 2\%$ across the profiles center.

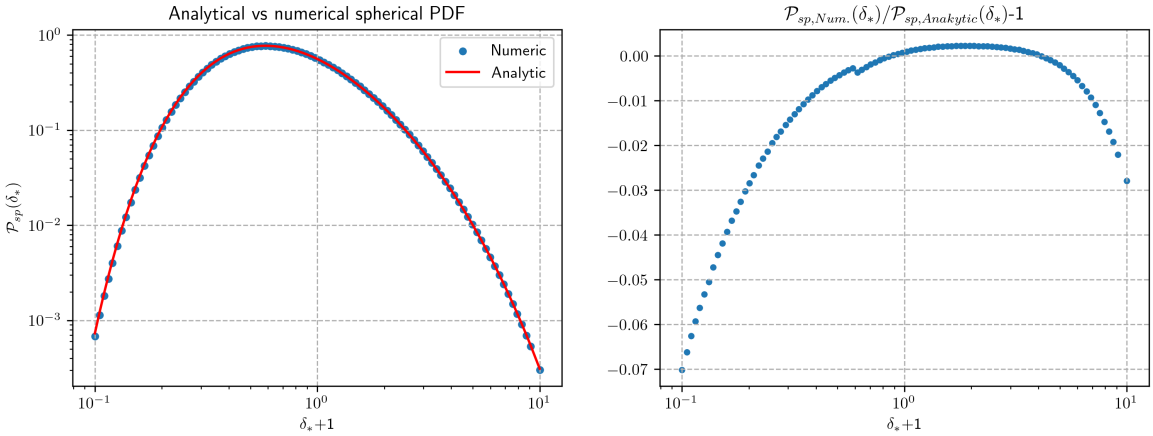

Figure 4.3: *Left panel*: Comparison between the numerical TopHat spherical PDF (dots) and its analytic expression (solid line) at redshift $z = 0$. *Right panel*: Residuals of the numerical results compared to the analytical. Note that the discontinuous jump occurring at $\delta_* = -0.4$ is a result of the different lattice parameters used for smaller and larger δ_*. The relative height of the discontinuity is $\sim 10^{-3}$.

Using the profiles computed by our algorithm, we can examine the saddle-point action, $\mathcal{S}[\hat{\delta}_L]$. By numerically differentiating the action, as per (3.0.7), we can then construct the Lagrange multiplier. The following results, and its residuals, are shown in Appendix C.

Having an estimate of the accuracy of the minimization code, we can now examine the results for the monopole prefactor. In order to reach a high enough accuracy between the numerical and analytic prefactor, the resolution of the lattice had to increase by a factor of 4 to 6 depending on the density. The prefactor code is initialized such that the lattice space is given by $(\tilde{N}, D) = (1800, 30)$ for $\delta_* \leq -0.4$ and $(\tilde{N}, D) = (1800, 20)$ for $\delta_* > -0.4$. Out of all the cases studied, these parameters were found to produce the minimal error for the TopHat filter. More detail about this parametrization will be discussed in the following section. The results of the

prefactor code when using the TopHat filter can be found in Appendix C. There we see that the numerical scheme is able to reproduce the analytic expression to within 1% error towards the extreme underdense and central densities. As we grow to extreme overdense regions, we start to see an climb in the error, growing to $\sim 3\%$ for $1 + \delta_* = 10$.

Combining the profile and prefactor results, we can then construct the numerical solution for the spherical PDF, $\mathcal{P}_{sp}(\delta_*)$, following (2.0.17). Here we note that the redshift dependence of the spherical PDF, captured in $g(z)$, is known. The result and its residuals can be seen in Fig. 4.3. We see that $\mathcal{P}_{sp}(\delta_*)$ can be estimated to within a 3% error for the central part of the PDF and its overdense tails. The underdense regions, however, begin to accumulate a higher error of 7%. By using a larger and denser lattice, we expect that this error can be improved on. However, this increase would become more computationally taxing, requiring a runtime of over 24 hours for each density point. While this runtime is not an issue for most cases, it becomes problematic if one were to use this pipeline for Monte Carlo simulations, where fast processes become necessary. This is discussed in more detail in chapter 7.

Comparing the analytic and numerical results not only validates the minimization and prefactor code, but also presents us with an understanding of the errors for non-TopHat filters, where no analytic solutions are available.

We also note that, unless otherwise stated, any analysis using the TopHat filter in the following sections will be done using its analytical expressions and not with the reconstructed numerical results of our pipeline.

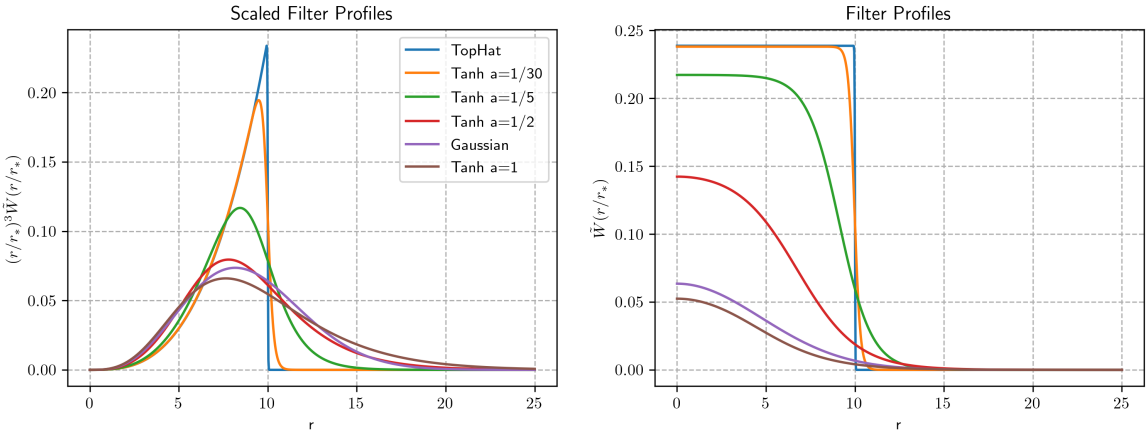

Figure 4.4: Comparison between the radial Gaussian, TopHat, and Tanh filters. The parameters r_* and a are taken from Table 4.1 to ensure that all filters have the same linear variance, σ^2_{L,r_*}=0.506. *Left panel:* Displays the filter profiles scaled by $(r/r_*)^3$. This is a dimensionless value. *Right panel:* Profiles of the filters themselves. Note the normalization in (2.0.1) which factors out the r_*^{-3} from the filters.

4.0.2 PDF's with non-TopHat filters

Now that we have validated the code, we are able to apply it to a variety of different window functions. In this section, we examine the following radial filters,

Gaussian :

$$\tilde{W}_{\rm g}(r/r_*) = \frac{1}{(2\pi)^{3/2}} e^{-\frac{1}{2}\left(\frac{r}{r_*}\right)^2} \Longleftrightarrow W_{\rm g}(kr_*) = e^{-\frac{(kr_*)^2}{2}}, \tag{4.0.2a}$$

Hyperbolic Tangent (Tanh):

$$\tilde{W}_{\rm ht}(r/r_*) = \frac{3}{2\pi(4+a^2\pi^2)} \left[\tanh\left(\frac{1-\frac{r}{r_*}}{a}\right) + \tanh\left(\frac{1+\frac{r}{r_*}}{a}\right)\right]. \tag{4.0.2b}$$

Here the Tanh function represents a family of curves parameterized by a. Note that for the Tanh filter, its Fourier transform does not have a simple analytic expression. Plots of these filters can be seen in Fig. 4.4 which visually compare the Gaussian and

	Smear, a	r_* (Mpc/h)
TopHat	-	10
Gaussian	-	4.7186
Tanh	1	4.2267
Tanh	1/2	6.7547
Tanh	1/5	9.1121
Tanh	1/30	9.9695

Table 4.1: Parameters for (4.0.2) needed to ensure that all filters share the same linear density variance as the TopHat filter with $r_* = 10$ Mpc/h. In this case, we have that $\sigma^2_{L,r_*} = 0.506$.

Tanh filters to the TopHat. Filters given by (4.0.2b) are intended to describe a family of functions which approximately interpolate between the Gaussian and TopHat filter. In the limit where $a \to 0$ we recover the case for the TopHat filter exactly, while for $a \sim 1$, we approach a curve which is qualitatively similar the Gaussian. We emphasize however that the function does not exactly reproduce the Gaussian in this limit.

With the goal to compare the statistics of PDF's with different filters, we enforce the condition that they all share the same linear density variance, σ^2_{L,r_*}, defined by

$$\sigma^2_{L,r_*} = \int_{\mathbf{k}} P(k)|W(kr_*)|^2. \tag{4.0.3}$$

This condition is further supported by the fact that the linear density variance plays the role of the expansion parameter in our perturbative solution of the PDF. Table 4.1 summarizes the parameters needed for all filters which will be considered in this paper to have the same variance compared to the TopHat.

Applying our code to these new filters with the above parameters, the results for the saddle-point density profiles are shown in Fig. 4.5. Here we had also set our lattice parameters to be $(N, D) = (450, 30)$ for $\delta_* \leq -0.4$ and (300,20) for $\delta_* > -0.4$.

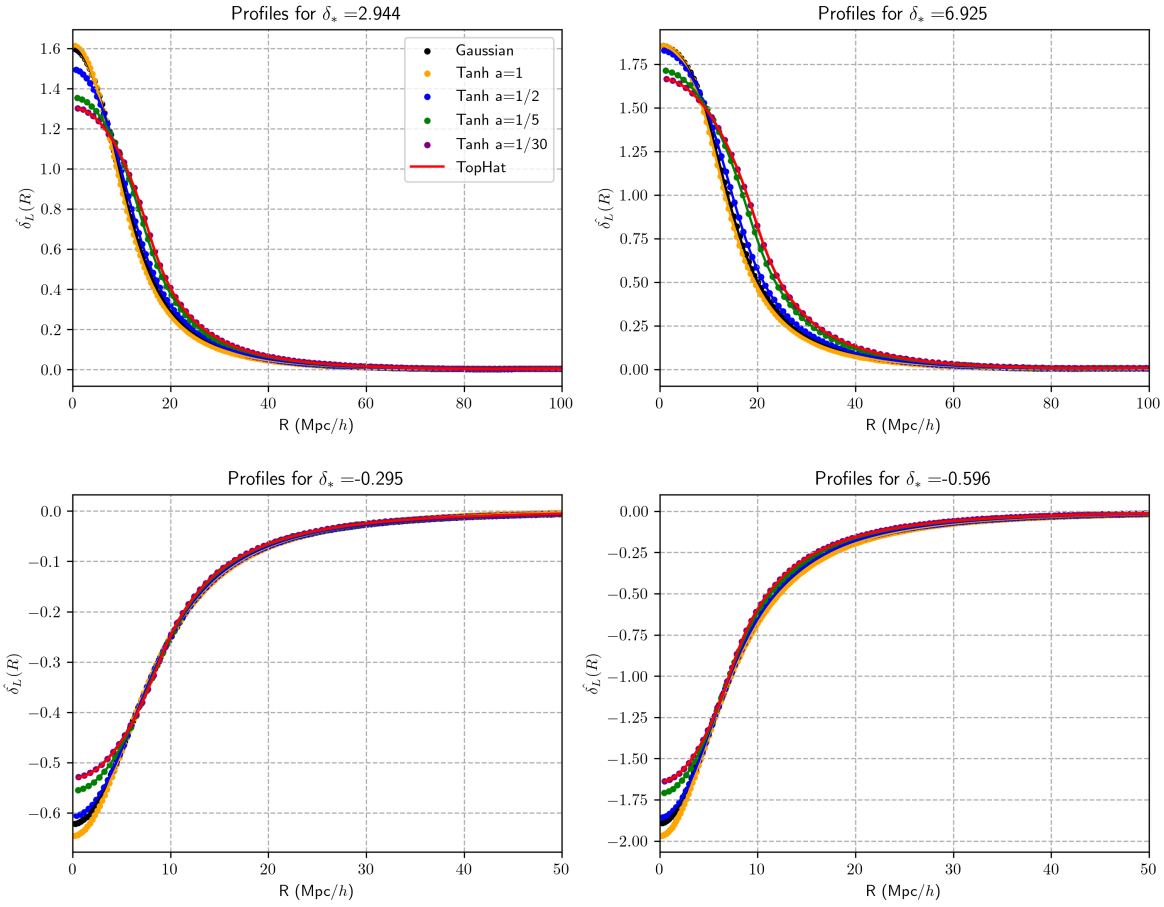

Figure 4.5: Profiles for several density contrasts using the Gaussian and Tanh filters (dots). For comparison, we show the profile for the TopHat filter (solid line). Top panels show the overdense profiles, with underdense profiles in the bottom.

As previously mentioned, this choice is made to ensure the convergence of the saddle-point configuration in the minimization code. This was done by gradually varying the N and D parameters in the lattice space and producing the corresponding action and prefactor curves. Once the difference between consecutive trials was less than 2%, the values for N and D would be selected. This tolerance of 2% was held only for $\delta_* \geq -0.8$ as attempting to improving the accuracy for small underdensities would not only be more computationally taxing, but may not be required due to the high

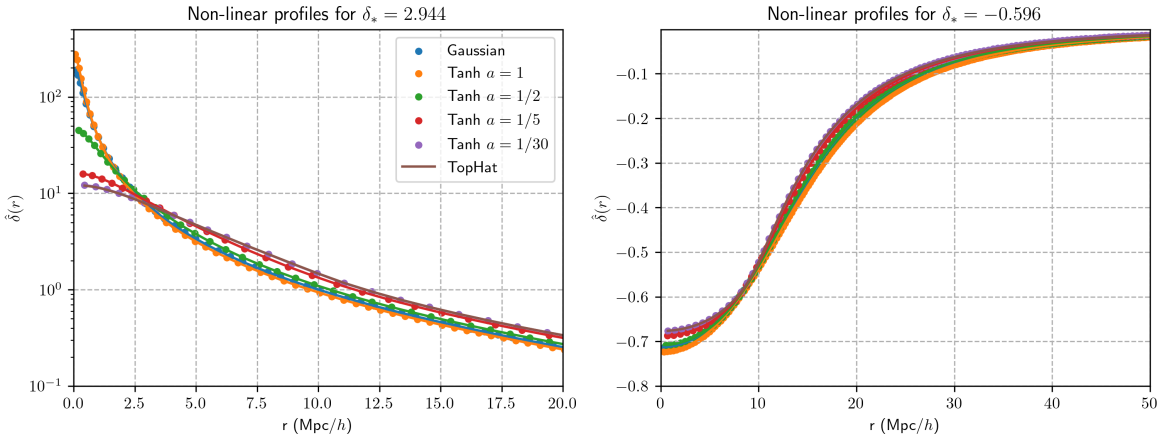

Figure 4.6: Comparison of the non-linear saddle-point profiles for the TopHat, Gaussian, and Tanh filters. Left panel displays an overdense profile while the right panel shows an underdense profile.

uncertainties associated with extreme underdense regions. We emphasize here that using convergence testing is a crucial step when examining non-TopHat filters. We have found that with the Gaussian and Tanh filters, the numerical results are more sensitive to the lattice than for the TopHat filter. Thus, using parameters for the TopHat filter do not always ensure the correct parameters for arbitrary filters in our minimization code.

Examining the profiles in Fig. 4.5, we see that in both overdense and underdense cases, the Tanh profiles for $a = 1/30$ appear to reproduce the TopHat results with high accuracy while larger values of a interpolate towards the Gaussian results. From the plots, we can also get an idea of the virialization process for a given filter. Recall that when $\delta_* \gtrsim 7$, shell-crossing begins to occur for a TopHat filter. Given the Gaussian and Tanh $a = 1, 1/2, 1/5, 1/30$ filters, we now find that the critical density occurs at $\delta_c \approx 3, 3, 4, 6, 7$ respectively. Again, note how the Tanh filters agree with the TopHat and Gaussian for the appropriate value of a. We see that Gaussian filters

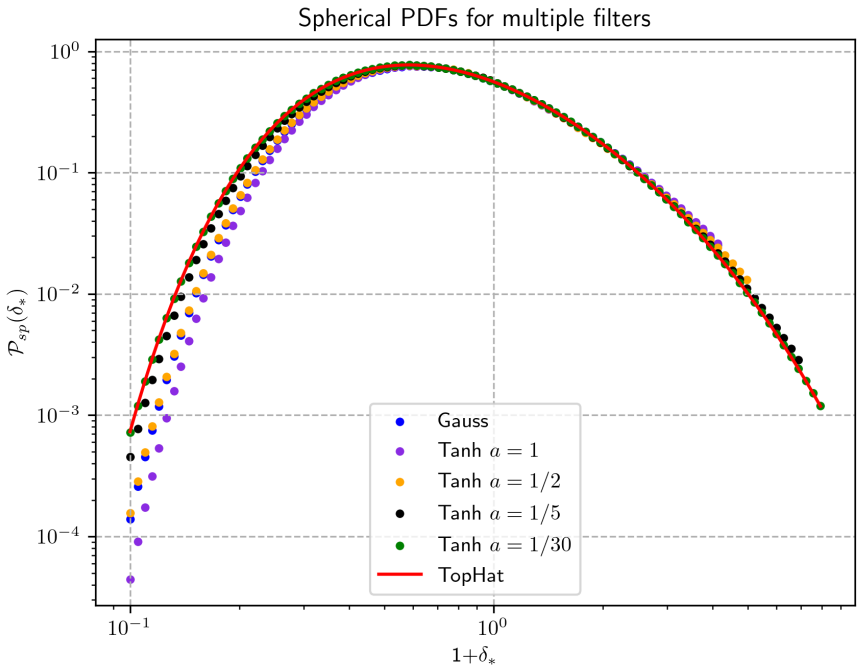

Figure 4.7: Spherical PDF for the Gaussian and Tanh filters compared to the TopHat at $z = 0$. The PDF's are cut at $\delta_* = \delta_c$. Spherical PDF's at different redshift may constructed according to (2.0.17), where the monopole prefactor and action (3.0.1) are divided by $g(z)$ and $g^2(z)$ respectively.

virialize at a significantly lower δ_* compared to the TopHat, cutting off the domain of its PDF much sooner. This is further confirmed by the gradual growth of δ_c in the Tanh functions as it goes from its Gaussian approximation to its TopHat limit. This early virialization may be explained by the way Gaussian and TopHat filters smooth over the density field in a given cell. Whereas TopHat filters would uniformly average the field within the cell, Gaussian filters weigh the matter content near the center of the cell more heavily compared to its edges. On average, the cell would effectively contain more matter towards its center, inducing stronger gravitational interactions between matter. The matter in the cell thus virializes more easily compared to a

uniformly spread field like in the case of the TopHat.

In Fig. 4.6, we plot the non-linear saddle-point profiles for an overdense (δ_*=2.944) and underdense ($\delta_* = -0.596$) cell for the TopHat, Tanh, and Gaussian filters. Looking at the overdense case, we see that the centers of the profile can vary drastically from one another. Profiles of Gaussian-like filters are shown to be an order of magnitude larger than the TopHat filter. This should not be too surprising, however, as we already saw that Gaussian-like filters lead to faster virialization than TopHat-like filters. This is related to the fact that Gaussian filters have smaller effective radius, r_*, for the the same linear variance (see Table 4.1). Hence, the corresponding density profiles are more non-linear. Moving towards the tail of the profiles, we see that all the curves begin to qualitative match with one another. Looking at the underdense case, we see that all filters produce qualitatively same profiles across the whole plot. Even the center of the curves match with one another, peaking at around ~ -0.7. This suggests that the non-linear profile of voids exhibits some universality to the choice of filter.

Fig. 4.7 compares the spherical PDF for multiple filters. From it, we see that Gaussian-like filters lead to a very similar PDF at overdense regions as the TopHat. The difference is more pronounced at underdensities, where Gaussian-like filters lead to smaller probability, by a factor of a few. However, we also see that the PDF's are not significantly different from one another, with all curves being qualitatively the same.

Chapter 5

Perturbative examination of the PDF

In this section, we closely follow the approach of [24] and examine analytic properties of the PDF at small density contrasts, $|\delta_*| \ll 1$, using perturbative methods.

Having now constructed a pipeline for the spherical PDF, we now want to include the effects coming from the aspherical contributions. While this is typically done using comparisons with N-body simulations (see chapter 6), analytic expressions can be derived for it within the perturbative regime. As discussed in [24], it is shown that the aspherical prefactor can be written as

$$\mathcal{A}_{asp} = \mathcal{A}_{tot}\sqrt{\mathcal{D}_0}, \tag{5.0.1}$$

where

$$\mathcal{D}_0 = \det[\mathbb{I} + 2\hat{\lambda}\sqrt{P}Q_0\sqrt{P}], \tag{5.0.2}$$

and

$$\mathcal{A}_{tot} = \frac{1}{\sqrt{\mathcal{D}_{tot}}}, \tag{5.0.3}$$

with

$$\mathcal{D}_{tot} = \det[\mathbb{I} + 2\hat{\lambda}\sqrt{P}Q_{tot}\sqrt{P}], \tag{5.0.4}$$

and

$$Q_{tot}(\mathbf{k}_1, \mathbf{k}_2) = \frac{1}{2}\frac{\partial^2 \bar{\delta}_W}{\partial\delta_L(\mathbf{k}_1)\partial\delta_L(\mathbf{k}_2)}. \tag{5.0.5}$$

Here, $\mathbb{I}$ refers to the operator given by (2.0.19) while the quantities $\sqrt{P}Q_i\sqrt{P}$ in (5.0.2) and (5.0.4) describe the operators given by $\sqrt{P(k_1)}Q_i(k_1, k_2)\sqrt{P(k_2)}$, with $P(k)$ being the linear power spectrum. The quantity $\mathcal{A}_{tot}$ is referred to as the *total prefactor* and is generated from all multipole fluctuations. Note that (5.0.1) holds for all δ_* and not just small values. By using standard cosmological perturbation theory (SPT) [8] and expanding up to second order in δ_*, it was shown that SPT corrections to the total prefactor correspond to 1-loop corrections of the linear power spectrum. While this is done assuming a TopHat filter, the effects of the loop correction for arbitrary filters will be examined here. By further considering effective field theory (EFT) corrections and examining the counterterm contributions, we then estimate the sensitivity that different filters have to the short-scale physics.

Note that our prefactor code also outputs the value of $\mathcal{D}_0$ since $Q_0(k_1, k_2)$ is needed for the monopole prefactor as well.

5.0.1 Perturbative expansion of the total prefactor

After applying SPT corrections and expanding Q_{tot} up to its first two terms, we find that [24]

$$Q_{tot}(\mathbf{k}_1, \mathbf{k}_2) = F_2(\mathbf{k}_1, \mathbf{k}_2) W(|\mathbf{k}_{12}| r_*) + 3 \int_{\mathbf{q}} F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}) W(|\mathbf{k}_{12} + \mathbf{q}| r_*) \hat{\delta}_L(\mathbf{q}), \quad (5.0.6)$$

where we denote $\mathbf{k}_{1..n} = \mathbf{k}_1 + ... + \mathbf{k}_n$ and F_2, F_3 are the SPT kernels (see [8]). Using the trace equation and assuming small δ_*, we can now approximate $\mathcal{D}_{tot}$ as

$$\begin{aligned} \mathcal{D}_{tot} &= \exp\left[\mathrm{Tr}\, \ln\left(\mathbb{I} + 2\hat{\lambda}\sqrt{P} Q_{tot} \sqrt{P}\right)\right] \\ &\approx \exp\left[2\hat{\lambda}\mathrm{Tr}\left(Q_{tot} P\right) - 2\hat{\lambda}^2 \mathrm{Tr}\left(Q_{tot} P Q_{tot} P\right)\right]. \end{aligned} \quad (5.0.7)$$

Looking for the first non-zero contributions to the trace terms, we find that

$$\mathrm{Tr}\left(Q_{tot} P\right) = 3 \int_{\mathbf{k}} \int_{\mathbf{q}} F_3(\mathbf{q}, -\mathbf{q}, \mathbf{k}) W(k r_*) \hat{\delta}_L(\mathbf{k}) P(q), \quad (5.0.8a)$$

$$\mathrm{Tr}\left(Q_{tot} P Q_{tot} P\right) = \int_{\mathbf{k}_1} \int_{\mathbf{k}_2} F_2^2(\mathbf{k}_1, \mathbf{k}_2) P(k_1) P(k_2) |W(\mathbf{k}_{12} r_*)|^2. \quad (5.0.8b)$$

By this point, we can begin to perturbatively expand the saddle-point quantities, $\hat{\lambda}$ and $\hat{\delta}_L(k)$, using the Euler-Lagrange equations (2.0.15). Writing $\hat{\lambda}$ in the following power series,

$$\hat{\lambda} = \sum_{i=1} a_i \delta_*^i, \quad (5.0.9)$$

and expanding $\bar{\delta}_W$ up to its first term in SPT as

$$\bar{\delta}_W = \int_{\mathbf{k}} W(kr_*)\delta_L(\mathbf{k}), \tag{5.0.10}$$

then the Euler-Lagrange equation (2.0.15a) tells us that $\hat{\delta}_L(k)$, up to leading order in δ_*, is given by

$$\hat{\delta}_L(k) = -a_1\delta_* P(k)W(kr_*). \tag{5.0.11}$$

Substituting (5.0.11) into the the constraint condition, $\delta_* = \bar{\delta}_W$, and expand $\bar{\delta}_W$ up to $\mathcal{O}(\delta_*)$, we can solve for a_1 to get

$$a_1 = -\frac{1}{\sigma^2_{L,r_*}}. \tag{5.0.12}$$

Putting all of this together, we find that $\mathcal{D}_{tot}$, expanded to second order in δ_*, is given by

$$\mathcal{D}_{tot} = \exp\left[-\delta_*^2 \frac{\sigma^2_{\text{1-loop}}}{\sigma^4_{L,r_*}}\right], \tag{5.0.13}$$

where

$$\sigma^2_{\text{1-loop}} = \int_{\mathbf{k}} P_{\text{1-loop}}(k)|W(kr_*)|^2, \tag{5.0.14a}$$

$$P_{\text{1-loop}}(k) = \int_{\mathbf{q}} \left(6F_3(\mathbf{k},-\mathbf{q},\mathbf{q})P(|\mathbf{k}-\mathbf{q}|)P(k) + 2F_2^2(\mathbf{k}-\mathbf{q},\mathbf{q})P(k)P(q)\right). \tag{5.0.14b}$$

This is exactly the 1-loop correction to the power spectrum which appears in SPT. Using this result, we then find that the total prefactor, up to second order, is given

by

$$\mathcal{A}_{tot} = \exp\left[\frac{\delta_*^2}{2}\frac{\sigma^2_{\text{1-loop}}}{\sigma^4_{L,r_*}}\right]. \tag{5.0.15}$$

Perhaps unsurprisingly, we find that the resulting equations have the exact same form as that of the TopHat filter, but with $\sigma^2_{\text{1-loop}}$ computed using the new filter. Note that (5.0.15) has zero derivative at $\delta_* = 0$. We will use the fact in the next section cross check our formalism.

5.0.2 One-loop corrections to the variance

While SPT can successfully describe linear density perturbation associated with large length, it is known to fail when describing short distance physics where perturbations become become highly non-linear. However, EFT of LSS allows one to probe further into the mildly non-linear regime by introducing a correction term, known as the counterterm, to the SPT expansion. In this subsection, we will discuss the sensitivity of the counterterm to the choice of filter. In particular, we want to see if smoother filters reduce the contributions of the counterterm to the PDF. We will find that this is not the case.

The counterterm is given by

$$\delta_c(\mathbf{k}) = -\gamma(z)k^2\delta_L(\mathbf{k}), \tag{5.0.16}$$

where $\gamma(z)$ is a redshift dependent coefficient which scales with g as[1] $\gamma(z) \sim g^2(z)$. Note that the EFT is applicable only if $\gamma k^2 \ll 1$. We note here that $\gamma(z)$ is some free coefficient that can only be found by fitting the model to data or N-body simulations.

[1] In reality, γ scales as $\gamma(z) \sim g^{8/3}(z)$ in the mildly non-linear regime, but is included here as it is less than $\mathcal{O}(g^3)$.

This comes as a consequence of using EFT, which introduces additional unknown coefficients into your expansion meant to capture the effects coming from the short-scale physics in the system. This, however, comes at the cost of not knowing their analytic expression, and so must instead be found through data or simulations [2].

Because the counterterm enters through how the non-linear density field, $\delta(\mathbf{k})$, relates to its linear field, $\delta_L(\mathbf{k})$, the Euler-Lagrange equations, and thus the saddle-point configuration, $\hat{\delta}_L(\mathbf{k})$, changes. Following a similar process as the last subsection, we expand $\hat{\lambda}$ as $\hat{\lambda} = \sum_{i=1} \tilde{a}_i \delta_*^i$. However, as we expand $\bar{\delta}_W$ up to its first term, it now picks up the contributions from the counterterm, resulting in the expression

$$\bar{\delta}_W = \int_{\mathbf{k}} W(kr_*)\delta_L(\mathbf{k})(1 - \gamma k^2). \tag{5.0.17}$$

Substituting this into the Euler-Lagrange equation we find that

$$\hat{\delta}_L(k) = -\tilde{a}_1 \delta_* P(k) W(kr_*)(1 - \gamma k^2). \tag{5.0.18}$$

Using the constraint condition, $\delta_* = \bar{\delta}_W$, we can solve for $\tilde{a}_1$ to get

$$\begin{aligned} \tilde{a}_1 &= -\frac{1}{\sigma^2_{L,r_*}} \left(\frac{1}{1 - \frac{\gamma \Sigma^2_{r_*}}{\sigma^2_{L,r_*}}} \right) \\ &\approx -\frac{1}{\sigma^2_{L,r_*}} \left(1 + \frac{2\gamma \Sigma^2_{r_*}}{\sigma^2_{L,r_*}} \right), \end{aligned} \tag{5.0.19}$$

[2] One can actually find analytic expressions for these coefficients assuming we know the full short-scale physics. However, this is usually not the case and so, they must be measured instead.

	Gaussian	Tanh $a = 1$	Tanh $a = 1/2$	Tanh $a = 1/5$	Tanh $a = 1/30$
$\Sigma^2_{r_*,i}/\Sigma^2_{r_*,TH}$	0.9369	1.0009	0.9263	0.8967	0.9794

Table 5.1: Ratios of the counterterm variance for different filters compared to the TopHat.

where in the last line, we expanded assuming small γk^2. Here we define

$$\Sigma^2_{r_*} = \int_{\mathbf{k}} k^2 P(k) |W(kr_*)|^2. \tag{5.0.20}$$

Putting it all together, we find that our modified saddle-point profile, expanded to leading order in δ_* and γk^2, is given by

$$\hat{\delta}_L(k) = \frac{\delta_*}{\sigma^2_{L,r_*}} P(k) W(kr_*) \left(1 - \gamma k^2 + \frac{2\gamma \Sigma^2_{r_*}}{\sigma^2_{L,r_*}}\right). \tag{5.0.21}$$

In order to calculate the contribution of the counterterm on the PDF, we repeat the calculation of $\mathcal{A}_{tot}$ in the previous subsection, but now using (5.0.21). Looking at the leading exponential in the PDF, we find that

$$-\frac{1}{g^2} \int_{\mathbf{k}} \frac{|\hat{\delta}_L(\mathbf{k})|^2}{2P(k)} \approx -\frac{\delta_*^2}{2g^2 \sigma^2_{L,r_*}} - \frac{\delta_*^2 \gamma \Sigma^2_{r_*}}{2g^2 \sigma^4_{L,r_*}}, \tag{5.0.22}$$

which gains a non-zero contribution at order g^2.

Using (5.0.15) and (5.0.22), then we find that part of the PDF is given by

$$\mathcal{A}_{tot} \exp\left[-\frac{1}{g^2} \int_{\mathbf{k}} \frac{|\hat{\delta}_L(\mathbf{k})|^2}{2P(k)}\right] \approx \exp\left\{-\frac{\delta_*^2}{2g^2 \sigma^2_{L,r_*}} + \frac{\delta_*^2}{2\sigma^4_{L,r_*}} \left(\sigma^2_{\text{1-loop}} - \frac{2\gamma \Sigma^2_{r_*}}{g^2}\right)\right\}. \tag{5.0.23}$$

Thus, we may define the renormalized, 1-loop linear density variance

$$\sigma^2_{\text{1-loop, ren}} = \sigma^2_{\text{1-loop}} - \frac{2\gamma}{g^2}\Sigma^2_{r_*}. \tag{5.0.24}$$

By generalizing our equations, we can now explore to what extent the aspherical prefactor of PDF's with different filters is sensitive to short-scale counterterms. In order to calculate the short-scale dependence, we must know the value of γ. As mentioned earlier, this can only be done by fitting our model to data or N-body simulations. In an attempt to remove our dependence on this nuisance parameter, we will instead examine ratios of the counterterm variances coming from different filters[3]. The results are shown in Table. 5.1. Here we see that most cases differ $\sim 10\%$ between the TopHat case, with the Tanh $a = 1/5$ case being the most different. Based off the filters explored, we see that the short-scale physics of the PDF has somewhat of a weak dependence on the filters used. Thus, despite having smoother boundaries, all filters are equally sensitive to the short-scale physics.

[3]We note here that γ does not depend on the filter, so it cancels in the above ratio.

Chapter 6

Aspherical contributions from N-body simulations

With a pipeline to construct the spherical PDF of an arbitrary window function numerically, we now want to discuss the effects coming from the aspherical fluctuations. These are encapsulated by the aspherical prefactor, $\mathcal{A}_{asp}$, (2.0.22). Computing this prefactor semi-analytically is difficult as it requires solving a coupled set of partial differential equations with initial conditions. This has been done for the case of the TopHat [24]. Generalizing the method for arbitrary filters is expected to increase the complexity of the equations, and so this approach is not considered in this work. As such, we instead extract the aspherical prefactor from N-body simulations. To this end, we use the high-resolution (HR) `Quijote` dataset [30]. The HD dataset consists of 100 realizations of a given ΛCDM cosmology which evolve 1024^3 cold dark matter (CDM) particles in a cosmological volume of 1 $(h^{-1}\text{Gpc})^3$ with a gravitational softening length of 25 h^{-1} kpc. Initial conditions for all simulations were generated

at $z = 127$ using second-order Lagrangian perturbation theory (2LPT). The simulation uses the following cosmological parameters, which closely follow the Planck 2018 constraints: $\Omega_{\rm m}$=0.3175, $\Omega_{\rm b}$=0.049, h =0.6711, n_s=0.9624, σ_8=0.834, $w = -1$, and massless neutrinos. The simulation saves snapshots at redshifts 0, 0.5, 1, 2, and 3.

From the N-body PDF, we can estimate the aspherical prefactor as

$$\mathcal{A}_{asp}(\delta_*) = \frac{\mathcal{P}_{\text{N-body}}(\delta_*)}{\mathcal{P}_{\text{sp}}(\delta_*)}. \tag{6.0.1}$$

By this point, it is important to note that the aspherical prefactor can be used to further validate our numerical pipeline described above. As shown in [24], the aspherical prefactor is expected to be redshift independent for any choice of window function[1]. This property can be used to cross-validate the simulations themselves.

In order to construct the N-body PDF, we use the `Pylians` library [32]. First, the algorithm distributes particle masses over the grid space using the Cloud-in-Cell mass assignment scheme (MAS). Then, by initializing our filter choice and sphere radius, r_*, as per Table 4.1, `Pylians` constructs non-overlapping spheres[2] over a given simulation and computes a histogram of the smoothed cell densities. Note that if one uses overlapping sphere, one increases the statistics used in constructing the PDF, thereby reducing the statistical noise. However, this would also increase the covariance between cells [33], which we choose to ignore here. The histogram is then translated into the PDF by a proper normalization. This process is repeated using 10 snapshots and averaged over, giving us $\mathcal{P}_{\text{N-body}}(\delta_*)$. The error of the PDF extracted

[1] A small dependence comes from the counterterm and two-loop corrections.

[2] Non-overlapping in this case refers to the fact that the center of each cell is separated by a distance of $2r_*$. In reality, as the cells are smoothed over using Gaussian and Tanh filters, they pick up contributions from outside the cell (unlike TopHat filters), causing them to overlap.

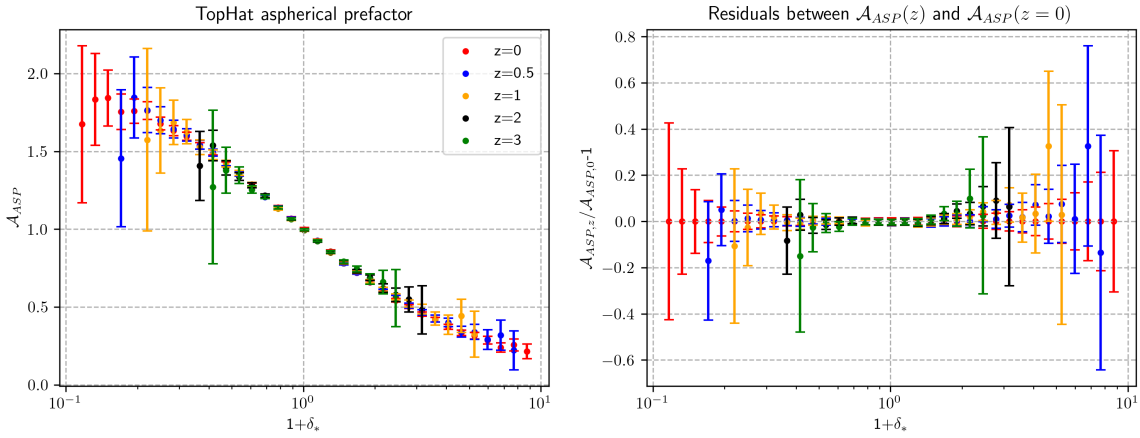

Figure 6.1: *Left panel:* The aspherical prefactor of the TopHat filter over redshifts $z = 0, 0.5, 1, 2, 3$. *Right panel:* Difference between aspherical prefactor at different redshifts compared to the $z = 0$ case.

from simulations is modeled as the Poisson noise within each δ_* bin.

In order to further verify our numerical pipeline, we examine the aspherical prefactor for the TopHat filter first. The results are shown in Fig. 6.1. There, we see that, within error bars, the aspherical prefactor is redshift independent across different values of δ_*, as expected. This agreement between theory and simulation further confirms the validity of our pipeline.

The resulting aspherical prefactor for all other filters can be found in Fig. 6.2. Note that we do not show plots for $a = 1/30, 1$ since they essentially coincide with the TopHat and Gaussian cases. We observe a mild redshift dependence of the residuals which becomes stronger for the Gaussian filter. The residuals follow a quadratic curve near $\delta_* = 0$. We also note that the prefactor at $\delta_* = 0$ drifts away from its theoretical value of $\mathcal{A}_{asp}(\delta_* = 0) = 1$ as redshift grows[3]. We expect this behavior to come from two-loop corrections to the PDF. This is because 1-loop contributions can

[3]This condition for $\mathcal{A}_{asp}$ comes from ensuring that the PDF remains normalized in the limit $g^2 \to 0$. However, this relation only holds true up to 1-loop corrections.

only influence second derivatives and higher of the PDF at $\delta_* = 0$ [25], suggesting that any shifting effects must come from higher loop corrections. Furthermore, it was found in [25] that effects from two-loop corrections became more noticeable on smaller sized cells with radii $r_* = 5$ and 7.5 Mpc$/h$. The results found in this work appear to be consistent with these results as our Gaussian filter has $r_* \sim 5$ Mpc$h$.

We also compare the results coming from our numerical aspherical prefactor with the perturbative, analytical equations derived in chapter 5. From (5.0.1), we see that the total prefactor can be written as

$$\mathcal{A}_{tot} = \mathcal{A}_{asp}/\sqrt{\mathcal{D}_0}.$$

Given that both $\mathcal{A}_{asp}$ and $\mathcal{D}_0$ can be obtained numerically, we now have a way to compare our analytic expression for $\mathcal{A}_{tot}$ (5.0.15) to numerical results $\mathcal{A}_{asp}/\sqrt{\mathcal{D}_0}$. This can be seen in Fig. 6.3 where we have computed $\mathcal{A}_{asp}/\sqrt{\mathcal{D}_0}$ for the Tanh $a = 1/30$ (left panel) and Gaussian (right panel) filters. According to (5.0.15), we expect that at $\delta_* = 0$, the derivative of the total prefactor should be zero. Looking at the Tanh $a = 1/30$ filter first, we see that this appears to hold true. This shows that the numerical results are in agreement with theoretical expectations. In principle, EFT allows one to calculate the coefficient in front of the δ_*^2 term in this ratio. This however requires measuring the counterterm $\gamma(z)$ from the N-body power spectrum. This measurement is outside the scope of our work.

Looking at the results for the Gaussian filter, we find some discrepancy as it appears to have a small but non-zero derivative at $\delta_* = 0$. The disagreement between theory and results are expected to come from two-loop corrections to the PDF which, as discussed earlier, were more prominent in the Gaussian-like filters.

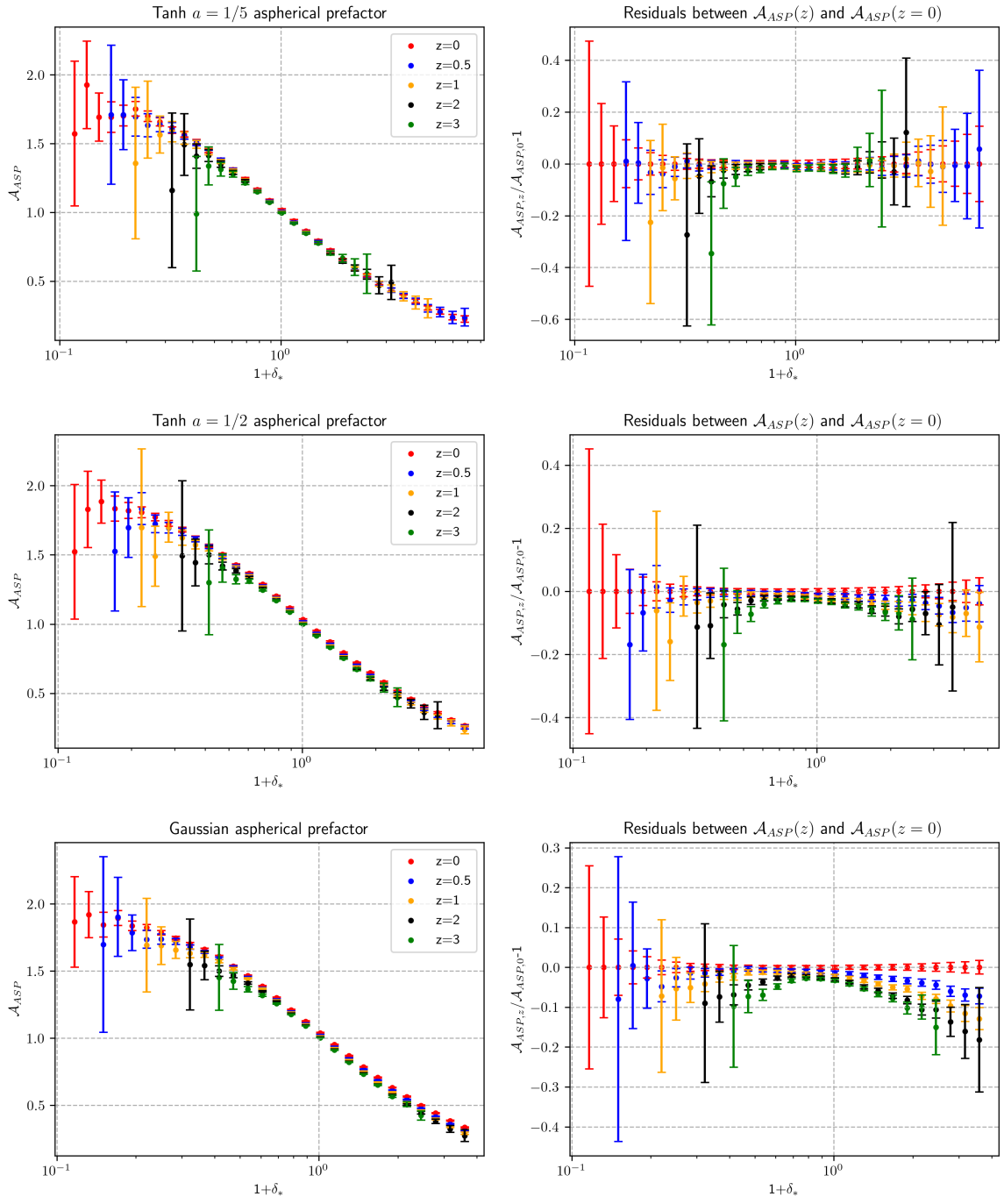

Figure 6.2: *Left panels:* Aspherical prefactor for multiple non-TopHat filters. *Right panel:* Compares the residuals of the prefactor at different redshift to $z = 0$. Note here that residuals come from the plots directly on its left.

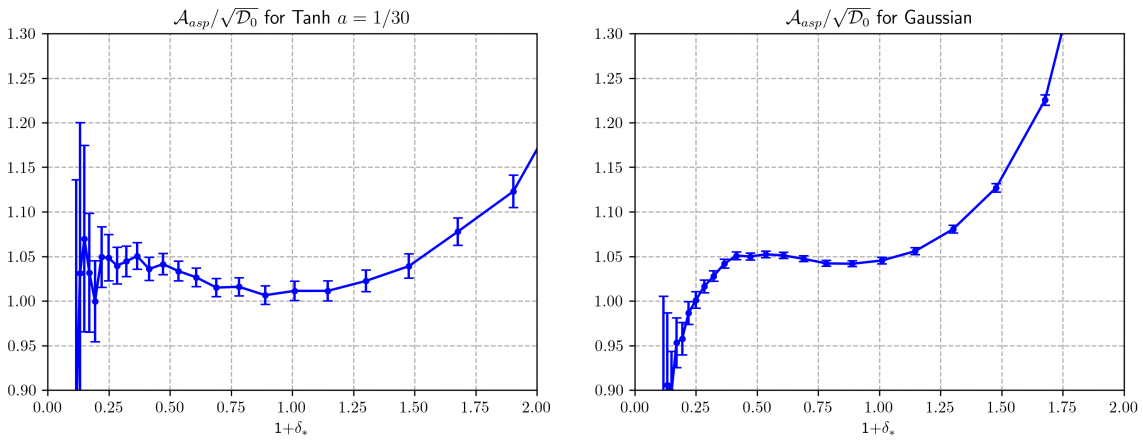

Figure 6.3: Plots of $\mathcal{A}_{asp}/\sqrt{\mathcal{D}_0}$ computed using the numerical results from the prefactor code and N-body comparison for redshift $z = 0$. Left panel displays the total prefactor for the Tanh $a = 1/30$ filter while the right panel comes from the Gaussian filter.

Chapter 7

Conclusion

In this work, we developed a numerical pipeline to construct the one-point PDF for matter densities given an arbitrary window function. By discretizing the system onto a lattice, the question of constructing numerical solutions of the PDF could then be rephrased as a multidimensional nonlinear constraint minimization problem. Our numerics were validated using analytic expressions for the PDF for the case of the Top-Hat filter. We found that the saddle-point profiles could be replicated numerically to within a sub-percent error. The overall spherical PDF, however, grew to a 3% error near its overdense tails. Here we also concluded that the box size of the lattice played an important role in accurately constructing the profiles while the number of points played the same role for the prefactor code. This implies that, while it would be possible to construct PDF's with sub-percent level errors, it would require high computational resources in order to run the code on a large and dense enough lattice.

By choosing the radii of cells such that the linear-matter variance for different PDF's were the same, we further examined the effects of several radial filters on the statistics of the matter fields. These filters included the Gaussian function, as well as a

family of curves which approximately interpolate between the Gaussian and TopHat. We found that overall, the spherical PDF's all qualitatively behaved the same, with no unique signatures found to distinguish one from the other.

While not examined in this paper, other popular filters, such as the Sharp-k function used in excursion set theory [19], can be implemented and studied using our code. Studying PDF's coming from sharp-k, however, are expected to be more computationally taxing when compared to the filters used in this paper. Due to the slow decrease and oscillatory features of the filter in position space, it is expected that a larger box size and number of lattice points would be needed to ensure the convergence of the profiles to the correct configuration. This would likely increase the runtime of the code to over 24 hours, requiring a significantly longer time to properly construct an accurate PDF.

The effects of the short-scale physics on different filters was also examined. By looking at the counterterm contributions coming from EFT corrections, it was found that for Gaussian-like filters, they were $\sim 10\%$ smaller compared to the TopHat case. This suggests that different, despite using qualitatively different filters, our PDF is not overly sensitive to the short-scale physics in our system.

Our PDF was compared with N-body simulations from the `Quijote` dataset. Results from this comparison showed that the aspherical contributions coming from the PDF was mostly redshift independent, agreeing with theoretical predictions. However, a small redshift dependent evolution in the aspherical prefactor was found in Gaussian-like filters, providing evidence for two-loop contributions.

As the runtime of our code currently sits on the order of tens of hours, further work is needed to improve the efficiency of our pipeline. This need for a faster runtime

comes from the requirement for quick and efficient code for Monte Carlo simulations for parameter extraction. Such simulations use a large number of randomly sampled sets of model parameters to make theoretical predictions and compare them with actual data. As our pipeline has a runtime on the order of hours, it cannot be used to efficiently construct and run Monte Carlo simulations, and so must be optimized. Whereas our current algorithm uses `Python` and runs on CPUs to solve for the saddle-point configuration, one could rewrite the minimization scheme in C/C++ code and use Nvidia's CUDA platform to GPU accelerate the process. Similarly, `JAX` [34] can be used to provide GPU accelerated options while using a `Python`-like syntax. This is expected to decrease the runtime of our code while also allowing us to construct a larger, denser lattice space needed to run our code with a high degree of accuracy.

Another important step for this research is to translate the matter PDF, which is not directly observable, to the statistics of galaxies. We plan to include the galaxy bias following the methods in [35, 36]. Constructing a galaxy PDF would also involve taking into account the effects of redshift-space distortion (RSD) [9].

It will also be useful to consider how this model relates to similar mass distribution schemes such as the excursion set theory [19]. As galaxy halo scales are typically of the order ~ 1 Mpc, but also represent extreme overdense regions of our universe, constructing a mass function for the small r_*, large δ_* limit of this theory may provide some insight on how it relates to or differs from other existing theories. However, there are various problems which must be addressed to solve this. This includes the treatment of virialized objects in this formalism and effects of baryonic physics on galactic scales. Currently, it is not clear how to incorporate either of these effects.

One potential application of our work deals with constructing void profile and

statics required to detect particle DM signals within extreme underdense regions of space [37, 38]. Other applications include relating constructing lensing models using spherical and cylindrical collapse functions [23] and relating our method to other statistical models, such as $k-$Nearest Neighbor (kNN) [39].

With semi-analytical methods able to study non-perturbative matter densities, we can go beyond the ΛCDM model and study signatures of primordial non-Gaussanity on the PDF. Another interesting direction would be to study the effects that long-range forces in the dark matter sector can have on the PDF through modifications of the spherical collapse dynamics. These are just two examples where our analytical methods can be useful, with more potential applications expected in the future.

Appendix A

Conventions

Following the same conventions as in [24], we define the Fourier transform of a field $\delta(\mathbf{x})$ as

$$\delta(\mathbf{x}) = \int_{\mathbf{k}} \delta(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{A.0.1}$$

where we define the integral measure over momentum space[1] as

$$\int_{\mathbf{k}} = \int \frac{d^3k}{(2\pi)^3}. \tag{A.0.2}$$

Similarly, we define the radial integral measure in momentum space to be

$$\int [dk] = \int_0^\infty \frac{dk k^2}{(2\pi)^3}, \tag{A.0.3}$$

[1] We emphasize here again that $\mathbf{k}$ describes the wavenumber defined from the Fourier transform in (A.0.1) and not any physical momenta in the system. However, we adopt this terminology for the thesis.

and its generalization to multiple wavenumbers as

$$\int [dk]^n = \int_0^\infty \prod_{i=1}^{n} \frac{dk_i k_i^2}{(2\pi)^3}. \tag{A.0.4}$$

In the scenario where we work with spherically symmetric fields, the Fourier transform turns into the following sine transform,

$$k\delta(k) = 4\pi \int dx x\delta(x) \sin(kx), \tag{A.0.5}$$

where x, k are the magnitudes of $\mathbf{x}, \mathbf{k}$ respectively.

The power spectrum is defined as,

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}')\rangle = (2\pi)^3 P(k)\delta_D(\mathbf{k}+\mathbf{k}') \tag{A.0.6}$$

where δ_D is the Dirac delta function.

We also use the following definition for the spherical harmonics:

$$Y_0(\theta,\phi) = 1, \tag{A.0.7a}$$

$$Y_{\ell m}(\theta,\phi) = \frac{(-1)^{\ell+m}}{2^\ell \ell!}\left[\frac{2\ell+1}{4\pi}\frac{(\ell-|m|)!}{(\ell+|m|)!}\right]^{1/2} e^{im\phi}(\sin\theta)^{|m|}\left(\frac{d}{d\cos\theta}\right)^{\ell+|m|}(\sin\theta)^{2\ell} \tag{A.0.7b}$$

for $\ell > 0$ and $-\ell \leq m \leq \ell$. They obey the following relations,

$$\Delta_\Omega Y_{\ell m} = -\ell(\ell+1)Y_{\ell m}, \quad Y_{\ell m}(-\mathbf{n}) = (-1)^\ell Y_{\ell m}, \quad Y^*_{\ell m}(\mathbf{n}) = Y_{\ell,-m}(\mathbf{n}), \tag{A.0.8}$$

where Δ_Ω is the Laplacian on the 2D sphere. The spherical harmonics also obey the following orthogonal and normalization conditions when integrated over the 2D sphere

$$\int d\Omega Y_{\ell m} Y^*_{\ell' m'} = (4\pi)^{\delta_{0,\ell}} \delta_{\ell\ell'} \delta_{mm'} \tag{A.0.9}$$

where δ_{ij} is the Kronecker delta.

Fields in position and momentum space can be decomposed over spherical harmonics as,

$$\delta(\mathbf{x}) = \delta_0(r) + \sum_{\ell>0} \sum_{m=-\ell}^{\ell} \delta_{\ell m}(r) Y_{\ell m}(\mathbf{x}/r), \tag{A.0.10a}$$

$$\delta(\mathbf{k}) = \delta_0(k) + \sum_{\ell>0} \sum_{m=-\ell}^{\ell} (-i)^\ell \delta_{\ell m}(k) Y_{\ell m}(\mathbf{k}/k). \tag{A.0.10b}$$

Using (A.0.9), the fields have the properties

$$(\delta_{\ell m}(r))^* = \delta_{\ell,-m}(r), \quad (\delta_{\ell m}(k))^* = \delta_{\ell,-m}(k). \tag{A.0.11}$$

The position and momentum coefficients in (A.0.10) are related by,

$$\delta_{\ell m}(r) = 4\pi \int [dk] j_\ell(kr) \delta_{\ell m}(k) \tag{A.0.12}$$

where $j_\ell(x)$ is the spherical Bessel function of order ℓ. It is related to the Bessel function of the first kind as

$$j_\ell(x) = \sqrt{\frac{\pi}{2x}} J_{\ell+1/2}(x). \tag{A.0.13}$$

Spherical Bessel functions with product arguments of the form $j_\ell(kr)$ obey an orthogonal relation on the half-line given by

$$\int_0^\infty dr r^2 j_\ell(k'r) j_\ell(kr) = \frac{\pi}{2k^2}\delta_D(k-k'). \tag{A.0.14}$$

Appendix B

Gaussian fluctuations around saddle-point configuration

B.0.1 Derivation of kernel functions

As discussed in [24], contributions to the prefactor are found by examining small perturbations around the saddle-point configuration. Focusing on the fluctuations around , $\hat{\delta}_L(\mathbf{k})$, for now, we have $\delta_L(\mathbf{k}) = \hat{\delta}_L(\mathbf{k}) + \delta_L^{(1)}(\mathbf{k})$, then perturbations to the averaged density contrast, up to second order, will take the form

$$\bar{\delta}_W = \delta_* + \int_{\mathbf{k}} S(\mathbf{k})\delta_L^{(1)}(\mathbf{k}) + \int_{\mathbf{k}_1}\int_{\mathbf{k}_2} Q_{tot}(\mathbf{k}_1, \mathbf{k}_2)\delta_L^{(1)}(\mathbf{k}_1)\delta_L^{(1)}(\mathbf{k}_2). \tag{B.0.1}$$

Here, δ_* is the target density contrast, whose statistics we are trying to find (see chapter 4), and S, Q_{tot} are defined as

$$S(\mathbf{k}) = \left.\frac{\partial\bar{\delta}_W}{\partial\delta_L(\mathbf{k})}\right|_{\hat{\delta}_L}, \tag{B.0.2a}$$

$$Q_{tot}(\mathbf{k}_1, \mathbf{k}_2) = \frac{1}{2} \left. \frac{\partial^2 \bar{\delta}_W}{\partial \delta_L(\mathbf{k}_1) \partial \delta_L(\mathbf{k}_2)} \right|_{\hat{\delta}_L} . \tag{B.0.2b}$$

Expanding the fluctuations into spherical harmonics

$$\delta_L^{(1)}(\mathbf{k}) = \delta_{L,0}^{(1)}(k) + \sum_{\ell>0} \sum_{m=-\ell}^{\ell} (-i)^\ell \delta_{L,\ell m}^{(1)}(k) Y_{\ell m}(\mathbf{k}/k), \tag{B.0.3}$$

and using the orthogonality condition of $Y_{\ell m}$, we find that the monopole contributions coming from (B.0.3) can be rewritten as

$$\bar{\delta}_{W,0} = \delta_* + \int [dk] 4\pi S(k) \delta_{L,0}^{(1)}(k) + \int [dk]^2 16\pi^2 \left\langle Q_{tot}(\mathbf{k}_1, \mathbf{k}_2) \right\rangle_{\Omega_1,\Omega_2} \delta_{L,0}^{(1)}(k_1) \delta_{L,0}^{(1)}(k_2), \tag{B.0.4}$$

where $\langle . \rangle_{\Omega_1,\Omega_2}$ denotes the spherical average over Ω_1 and Ω_2. Note that we has used the fact that a spherically symmetric ansatz for $\delta_L(\mathbf{k})$ also implies that $S(\mathbf{k})$ is spherically symmetric [24]. Here we emphasize that the quantity $\langle Q_{tot}(\mathbf{k}_1, \mathbf{k}_2) \rangle_{\Omega_1,\Omega_2}$ now depends only on the magnitudes k_1 and k_2.

Had we originally expanded $\bar{\delta}_W$ over spherically symmetric configurations, $\hat{\delta}_L(k)$, (B.0.1) may be written as,

$$\bar{\delta}_W = \delta_* + \int [dk] 4\pi S(k) \delta_L^{(1)}(k) + \int [dk]^2 4\pi Q_0(k_1, k_2) \delta_L^{(1)}(k_1) \delta_L^{(1)}(k_2), \tag{B.0.5}$$

for some spherically symmetric kernel, $Q_0(k_1, k_2)$. Given that the monopole contribution (B.0.4) represents the spherically symmetric part of $\bar{\delta}_W$, then we see that $Q_0(k_1, k_2)$ and $Q_{tot}(\mathbf{k}_1, \mathbf{k}_2)$ are related to one another through the relation,

$$Q_0(k_1, k_2) = 4\pi \left\langle Q_{tot}(\mathbf{k}_1, \mathbf{k}_2) \right\rangle_{\Omega_1,\Omega_2} . \tag{B.0.6}$$

While this relation is not used in this paper, it provides a useful relation between the three-dimensional kernel and its spherically symmetric counterpart.

In order to find explicit expressions for $S(k)$ and $Q_0(k_1, k_2)$, we can consider taking variations of $\bar{\delta}_W$ with respect to spherically symmetric configurations, $\delta_L(k)$. Going up to second order, we find that[1]

$$\begin{aligned}\bar{\delta}_W[\delta_L] &= \bar{\delta}_W[\hat{\delta}_L] + d\bar{\delta}_W[\hat{\delta}_L] + d^2\bar{\delta}_W[\hat{\delta}_L] \\ &= \delta_* - \frac{16\pi^2}{r_*^4 k}\int [dk] dR\, R^2 \tilde{W}' \frac{f'}{(1+f)^{4/3}} j_1(kR)\, d\delta_L(k) \\ &\quad + \frac{32\pi^3}{r_*^3}\int [dk]^2 dR R^2 W_{\rm th}(k_1 R) W_{\rm th}(k_2 R) \left[\frac{\tilde{W}'' R^2}{9 r_*^2}\frac{(f')^2}{(1+f)^{8/3}}\right. \\ &\quad \left. + \frac{\tilde{W}' R}{3 r_* (1+f)^{4/3}}\left(-f'' + \frac{4}{3}\frac{(f')^2}{1+f}\right)\right] d\delta_L(k_1)\, d\delta_L(k_2),\end{aligned} \tag{B.0.7}$$

where the primes denote differentiation with respect to the argument of the function. Here we have omitted the argument of some functions, with $f = f(\hat{\delta}_L(R))$, where f is the spherical collapse dynamics defined by (2.0.11), and $\tilde{W} = \tilde{W}(R(1+f)^{-1/3}/r_*)$.

Comparing (B.0.7) with (B.0.5), we find that $S(k)$ and $Q_0(k_1, k_2)$ are given by

$$S(k) = -\frac{4\pi}{r_*^4 k}\int dR R^2 \tilde{W}' \frac{f'}{(1+f)^{4/3}} j_1(kR), \tag{B.0.8a}$$

$$\begin{aligned}Q_0(k_1, k_2) = \frac{8\pi^2}{r_*^3}\int dR R^2 W_{\rm th}(k_1 R) W_{\rm th}(k_2 R) &\left[\frac{\tilde{W}'' R^2}{9 r_*^2}\frac{(f')^2}{(1+f)^{8/3}}\right. \\ &\left. + \frac{\tilde{W}' R}{3 r_* (1+f)^{4/3}}\left(-f'' + \frac{4}{3}\frac{(f')^2}{1+f}\right)\right].\end{aligned} \tag{B.0.8b}$$

[1] To avoid the overuse of δ, we will denote the variation of functional as d.

B.0.2 Monopole prefactor

Ref. [24] derives the results for the leading order corrections to the saddle-point configuration. Here we provide a summary of the method and its results before discussing the monopole prefactor.

Let us consider small fluctuations around the saddle-point configuration. We expand the Lagrange multiplier as $\lambda = \hat{\lambda} + \lambda^{(1)}$ and the linear density field into spherical harmonics following (B.0.3). Such fluctuations induces perturbations in $\bar{\delta}_W$ which, up to second order, can be written as

$$\begin{aligned}\bar{\delta}_W = \delta_* + \int [dk] 4\pi S(k) \delta^{(1)}_{L,0}(k) + \int [dk]^2 4\pi Q_0(k_1, k_2) \delta^{(1)}_{L,0}(k_1) \delta^{(1)}_{L,0}(k_2) \\ + \sum_{\ell>0,m} \int [dk]^2 Q_\ell(k_1, k_2) \delta^{(1)}_{L,\ell,m} \delta^{(1)}_{L,\ell,-m}(k_2),\end{aligned} \tag{B.0.9}$$

where $S(k), Q_0(k_1, k_2), Q_\ell(k_1, k_2)$ are some kernels, with $S(k)$ and $Q_0(k_1, k_2)$ having been discussed in the previous section.

Substituting the above perturbations into (2.0.6), we find that the PDF factorizes over the multipole numbers ℓ, resulting in (2.0.16). While the specific form for $\mathcal{A}_\ell$ is not relevant for this work, we find that the monopole prefactor has the form

$$\begin{aligned}\mathcal{A}_0 = \mathcal{N}_0^{-1} \int_{-i\infty}^{i\infty} \frac{d\lambda^{(1)}}{2\pi i g^2} \int \mathcal{D}\delta^{(1)}_{L,0} \exp\Bigg\{ -\frac{4\pi}{g^2} \Bigg[\int \frac{[dk]}{2P(k)} (\delta^{(1)}_{L,0}(k))^2 \\ + \lambda^{(1)} \int [dk] S(k) \delta^{(1)}_{L,0}(k) + \hat{\lambda} \int [dk]^2 Q_0(k_1, k_2) \delta^{(1)}_{L,0}(k_1) \delta^{(1)}_{L,0}(k_2) \Bigg] \Bigg\},\end{aligned} \tag{B.0.10}$$

where $\mathcal{N}_0$ is the normalization constant given by

$$\mathcal{N}_0 = \int \mathcal{D}\delta^{(1)}_{L,0} \exp\Bigg\{ -\frac{4\pi}{g^2} \int \frac{[dk]}{2P(k)} (\delta^{(1)}_{L,0}(k))^2 \Bigg\}. \tag{B.0.11}$$

Note that by this point, the resulting equations do not depend on our choice of filter, except in determining the explicit forms of the kernels $S(k)$, $Q_0(k_1, k_2)$. Noticing that both $\mathcal{A}_0$ and $\mathcal{N}_0$ are Gaussian path integrals, they can be solved and represented as functional determinants. After evaluating the integrals and simplifying, we get the following expression for the monopole prefactor,

$$\mathcal{A}_0 = \frac{1}{\sqrt{2\pi g^2}} \sqrt{\frac{\det\left(\frac{\mathbb{I}(k_1,k_2)}{4\pi P(k)}\right)}{-\det \mathcal{H}}}, \tag{B.0.12}$$

where $\mathcal{H}$ and $\mathbb{I}$ are given by (2.0.20) and (2.0.19) respectively.

Appendix C

Validation of the code

Using the lattice values (450,30) for $\delta_* \leq -0.4$ and (300,20) for $\delta_* > -0.4$, we show the results for the TopHat action and Lagrange multiplier at the saddle-point configuration in Fig. C.1 and C.2 respectively. There we see that the error between analytic and numerical results is less than 2% across all values of δ_*, except near $\delta_* = 0$ where the error quickly grows. This large error does not come from a failure of the numerics, but rather our use of the relative error for quantities which tend to zero, where small deviations may result in large residuals. As such, this shows that the pipeline is able to reconstruct the relevant statistics quantities to a good degree of accuracy.

If one uses the same lattice values (N, D) for all density contrasts in $[-0.9, 9]$, it was found that the underdense regions exhibited a slower convergence rate than overdense regions, regardless of the choice of filter used. This can be explained by our discretization procedure in (3.0.2). Recall that $R_{\max}$ is dependent on the choice of δ_*. As a result, for a fixed value of D, a small negative value of δ_* may result in a significantly smaller box size, $R_{\max}$, for underdense values compared to overdense values. In order to accurately represent the tails for extreme underdense profiles, we change

our lattice space to include a larger box size for underdense regions, while keeping the density of points the same across all δ_*. This suggests that box size is the key parameter needed to accurately capture the profiles. In fact, earlier test cases originally used lattice parameters of (450,15) for $\delta_* \leq -0.4$ and (300,10) for $\delta_* > -0.4$. The profiles from these cases were found to perform worse when compared to the standard parameters of (450,30) for $\delta_* \leq -0.4$ and (300,20) for $\delta_* > -0.4$, producing errors on the order of 10% near the tails of the profiles. Note that our optimal parameters differ from this earlier case by simply doubling the box size without increasing number of lattice points. We found that increasing the latter does not significantly reduce the error of the profiles, but dramatically increases the computation time. Thus, we expect the size of the lattice to be the important parameter for constructing accurate profiles.

This is in direct contrast with what happens in the prefactor code, however. After running the prefactor algorithm using the optimal lattice parameters, we find that the numerical TopHat solutions differ from the analytics up to 40% in the extreme overdense regions, as shown in Fig. C.3. However, if the number of lattice points is increased while holding the size of the box fixed, the relative error quickly drops. After examining multiple cases, the new lattice parameters $(\tilde{N}, D) = (1800, 30)$ for $\delta_* \leq -0.4$ and $(\tilde{N}, D) = (1800, 20)$ for $\delta_* > -0.4$ (Fig. C.4) were found to produce the smallest error among all our cases, peaking to around 3% near the extreme overdense tails.

As discussed in the main text, we found that non-TopHat filters required some convergence testing to ensure the saddle-point and prefactor code were computed to a high enough accuracy. We show in Fig. C.5 an example of this convergence testing

in the Gaussian case. Here, we consider four lattice parameter sets, given by

$$\text{Test 0:} \begin{cases} (450, 15) & \delta_* \leq -0.4 \\ (300, 10) & \delta_* > -0.4 \end{cases}, \qquad \text{Test 1:} \begin{cases} (450, 20) & \delta_* \leq -0.4 \\ (300, 15) & \delta_* > -0.4 \end{cases} \tag{C.0.1}$$

$$\text{Test 2:} \begin{cases} (360, 24) & \delta_* \leq -0.4 \\ (225, 15) & \delta_* > -0.4 \end{cases}, \qquad \text{Test 3:} \begin{cases} (450, 30) & \delta_* \leq -0.4 \\ (300, 20) & \delta_* > -0.4 \end{cases}.$$

There we see that as we compare consecutive trials, the amplitude in the errors decrease, going from a maximum of 4% in both the action and prefactor, to $\sim 2\%$.

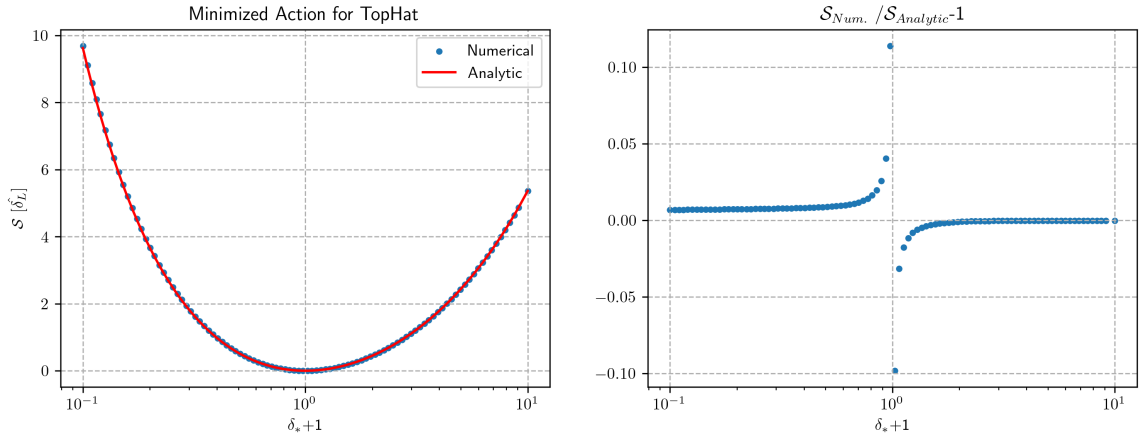

Figure C.1: *Left panel*: Comparisons between saddle-point action from the optimization algorithm (dots) and the analytic expression (solid line) as a function of δ_*. *Right panel*: Plots the residuals of the numerical results compared to the analytical.

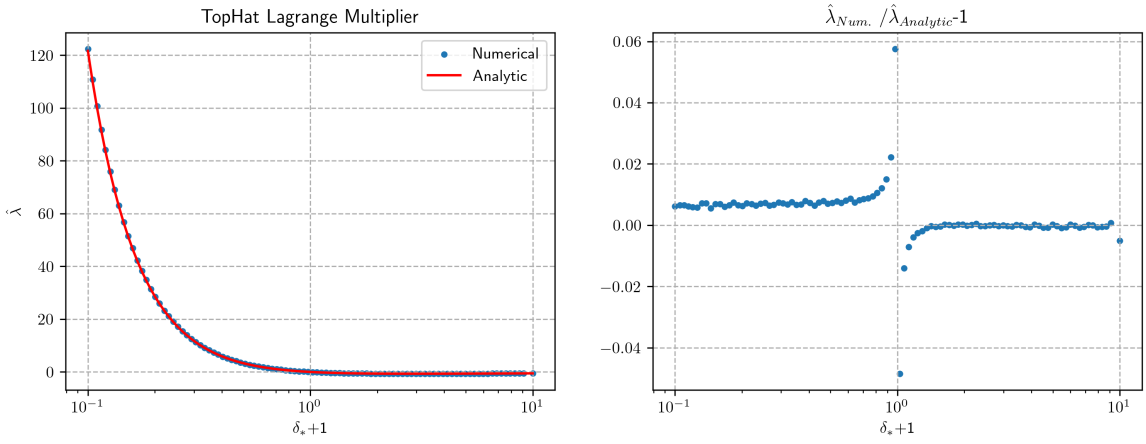

Figure C.2: *Left panel*: Comparisons between numerical saddle-point Lagrange multiplier (dots) and the analytic expression (solid line) as a function of δ_*. *Right panel*: Plots the residuals of the numerical results compared to the analytical.

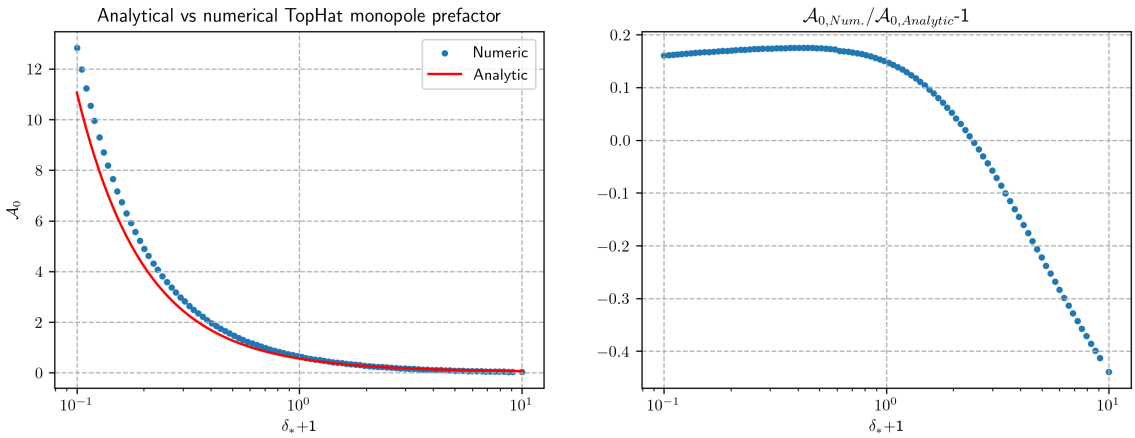

Figure C.3: Results of the prefactor code without increasing the density of points on the lattice *Left panel*: Comparisons between numerical spherical prefactor (dots) and the analytic expression (solid line) as a function of δ_*. *Right panel*: Plots the residuals of the numerical results compared to the analytical.

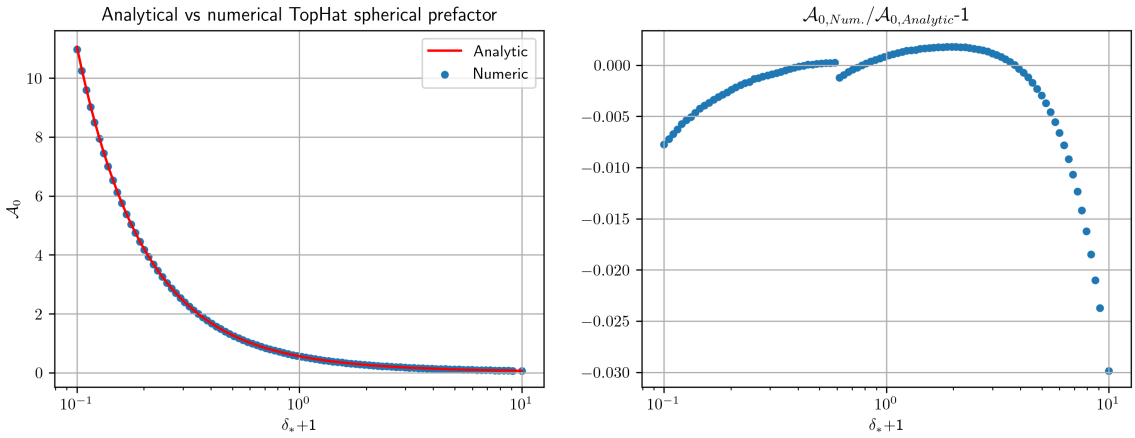

Figure C.4: Results of the prefactor code using optimal lattice parameters. *Left panel*: Comparisons between numerical spherical prefactor (dots) and the analytic expression (solid line) as a function of δ_*. *Right panel*: Plots the residuals of the numerical results compared to the analytical.

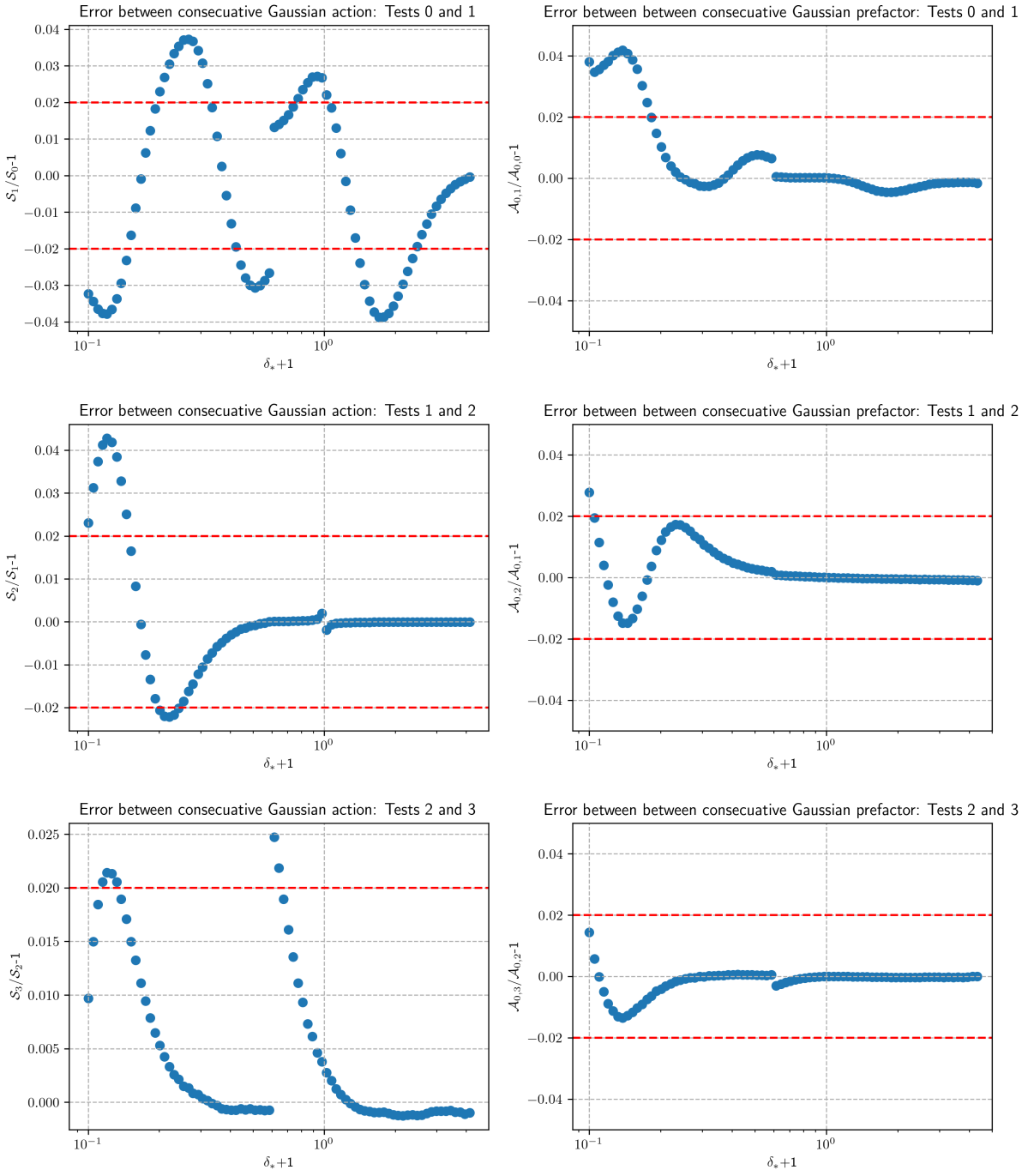

Figure C.5: Residuals of the action and prefactor between several tests cases following (C.0.1). Red dashed lines display the 2% error lines. The top row compares Tests 0 and 1, the middle compares tests 1 and 2, and the bottom compares tests 2 and 3. Note that the underdense regions appear to converge at a slower rate compared to the overdense regions, especially in the prefactor.

Bibliography

[1] M. S. Turner, "The road to precision cosmology," *Annual Review of Nuclear and Particle Science*, vol. 72, no. 1, pp. 1–35, 2022.

[2] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. Barreiro, N. Bartolo, S. Basak, *et al.*, "Planck 2018 results-vi. cosmological parameters," *Astronomy & Astrophysics*, vol. 641, p. A6, 2020.

[3] P. Bull *et al.*, "Beyond ΛCDM: Problems, solutions, and the road ahead," *Phys. Dark Univ.*, vol. 12, pp. 56–99, 2016.

[4] D. Fixsen, "The temperature of the cosmic microwave background," *The Astrophysical Journal*, vol. 707, no. 2, p. 916, 2009.

[5] D. J. Fixsen, E. Cheng, J. Gales, J. C. Mather, R. Shafer, and E. Wright, "The cosmic microwave background spectrum from the full cobe* firas data set," *The Astrophysical Journal*, vol. 473, no. 2, p. 576, 1996.

[6] G. Cabass, M. M. Ivanov, M. Lewandowski, M. Mirbabayi, and M. Simonović, "Snowmass white paper: Effective field theories in cosmology," *Physics of the Dark Universe*, vol. 40, p. 101193, 2023.

[7] Y. B. Zel'Dovich, "Gravitational instability: An approximate theory for large density perturbations.," *Astronomy and Astrophysics, Vol. 5, p. 84-89*, vol. 5, pp. 84–89, 1970.

[8] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, "Large-scale structure of the universe and cosmological perturbation theory," *Physics reports*, vol. 367, no. 1-3, pp. 1–248, 2002.

[9] M. M. Ivanov, "Effective field theory for large-scale structure," in *Handbook of Quantum Gravity*, pp. 1–48, Springer, 2023.

[10] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, "Cosmological non-linearities as an effective fluid," *Journal of Cosmology and Astroparticle Physics*, vol. 2012, no. 07, p. 051, 2012.

[11] 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, no. 9, pp. 1–40, 2012.

[12] E. Pajer and M. Zaldarriaga, "On the renormalization of the effective field theory of large scale structures," *Journal of Cosmology and Astroparticle Physics*, vol. 2013, no. 08, p. 037, 2013.

[13] A. A. Abolhasani, M. Mirbabayi, and E. Pajer, "Systematic renormalization of the effective theory of large scale structure," *Journal of Cosmology and Astroparticle Physics*, vol. 2016, no. 05, p. 063, 2016.

[14] R. E. Angulo, S. Foreman, M. Schmittfull, and L. Senatore, "The one-loop matter

bispectrum in the effective field theory of large scale structures," *Journal of Cosmology and Astroparticle Physics*, vol. 2015, no. 10, p. 039, 2015.

[15] T. Baldauf, L. Mercolli, M. Mirbabayi, and E. Pajer, "The bispectrum in the effective field theory of large scale structure," *Journal of Cosmology and Astroparticle Physics*, vol. 2015, no. 05, p. 007, 2015.

[16] T. Baldauf, L. Mercolli, and M. Zaldarriaga, "Effective field theory of large scale structure at two loops: The apparent scale dependence of the speed of sound," *Physical Review D*, vol. 92, no. 12, p. 123007, 2015.

[17] W. H. Press and P. Schechter, "Formation of galaxies and clusters of galaxies by self-similar gravitational condensation," *Astrophysical Journal, Vol. 187, pp. 425-438 (1974)*, vol. 187, pp. 425–438, 1974.

[18] J. M. Bardeen, J. Bond, N. Kaiser, and A. Szalay, "The statistics of peaks of gaussian random fields," *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 304, May 1, 1986, p. 15-61. SERC-supported research.*, vol. 304, pp. 15–61, 1986.

[19] J. Bond, S. Cole, G. Efstathiou, and N. Kaiser, "Excursion set mass functions for hierarchical gaussian fluctuations," *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 379, Oct. 1, 1991, p. 440-460. Research supported by NSERC, NASA, and University of California.*, vol. 379, pp. 440–460, 1991.

[20] A. R. Zentner, "The excursion set theory of halo mass functions, halo clustering, and halo growth," *International Journal of Modern Physics D*, vol. 16, no. 05, pp. 763–815, 2007.

[21] P. J. E. Peebles, *The large-scale structure of the universe*, vol. 96. Princeton university press, 1980.

[22] F. Bernardeau and P. Reimberg, "Large deviation principle at play in large scale structure cosmology," *Physical Review D*, vol. 94, no. 6, p. 063520, 2016.

[23] C. Uhlemann, C. Pichon, S. Codis, B. L'Huillier, J. Kim, F. Bernardeau, C. Park, and S. Prunet, "Cylinders out of a top hat: counts-in-cells for projected densities," *Monthly Notices of the Royal Astronomical Society*, vol. 477, no. 2, pp. 2772–2785, 2018.

[24] M. M. Ivanov, A. A. Kaurov, and S. Sibiryakov, "Non-perturbative probability distribution function for cosmological counts in cells," *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 03, p. 009, 2019.

[25] A. Chudaykin, M. M. Ivanov, and S. Sibiryakov, "Renormalizing one-point probability distribution function for cosmological counts in cells," *Journal of Cosmology and Astroparticle Physics*, vol. 2023, no. 08, p. 079, 2023.

[26] T. Baldauf, S. Codis, V. Desjacques, and C. Pichon, "Nonperturbative halo clustering from cosmological density peaks," *Physical Review D*, vol. 103, no. 8, p. 083530, 2021.

[27] K. Tokeshi, K. Inomata, and J. Yokoyama, "Window function dependence of the novel mass function of primordial black holes," *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 12, p. 038, 2020.

[28] T. Karim, M. Rezaie, S. Singh, and D. Eisenstein, "On the impact of the galaxy

window function on cosmological parameter estimation," *Monthly Notices of the Royal Astronomical Society*, vol. 525, no. 1, pp. 311–324, 2023.

[29] D. Blas, J. Lesgourgues, and T. Tram, "The cosmic linear anisotropy solving system (class). part ii: Approximation schemes," *Journal of Cosmology and Astroparticle Physics*, vol. 2011, no. 07, p. 034, 2011.

[30] F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys, *et al.*, "The quijote simulations," *The Astrophysical Journal Supplement Series*, vol. 250, no. 1, p. 2, 2020.

[31] H. Mo, F. Van den Bosch, and S. White, *Galaxy formation and evolution*. Cambridge University Press, 2010.

[32] F. Villaescusa-Navarro, "Pylians: Python libraries for the analysis of numerical simulations." Astrophysics Source Code Library, record ascl:1811.008, Nov. 2018.

[33] C. Uhlemann, O. Friedrich, A. Boyle, A. Gough, A. Barthelemy, F. Bernardeau, and S. Codis, "It takes two to know one: Computing accurate one-point pdf covariances from effective two-point pdf models," *arXiv preprint arXiv:2210.07819*, 2022.

[34] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: composable transformations of Python+NumPy programs," 2018.

[35] V. Desjacques, D. Jeong, and F. Schmidt, "Large-scale galaxy bias," *Physics reports*, vol. 733, pp. 1–193, 2018.

[36] N. Kokron, J. DeRose, S.-F. Chen, M. White, and R. H. Wechsler, "Priors on red galaxy stochasticity from hybrid effective field theory," *Monthly Notices of the Royal Astronomical Society*, vol. 514, no. 2, pp. 2198–2213, 2022.

[37] S. Arcari, E. Pinetti, and N. Fornengo, "Got plenty of nothing: cosmic voids as a probe of particle dark matter," *Journal of Cosmology and Astroparticle Physics*, vol. 2022, no. 11, p. 011, 2022.

[38] R. Voivodic, H. Rubira, and M. Lima, "The halo void (dust) model of large scale structure," *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 10, p. 033, 2020.

[39] A. Banerjee and T. Abel, "Nearest neighbour distributions: New statistical measures for cosmological clustering," *Monthly Notices of the Royal Astronomical Society*, vol. 500, no. 4, pp. 5479–5499, 2021.