Beyond the Mean: Advancing the Analytic Outer Density Profile

Ericka Florio
DAMTP, University of Cambridge and FORTH, University of Crete
Cambridge, UK and Heraklion, GR
eaf49@cam.ac.uk

Abstract

The mode of the double distribution describes the "most probable" galactic outer density profile, and the departure of this profile from galactic simulations can be used to infer cosmological parameters. This python library calculates and visualises the double distribution and the most probable profile, for a given cosmological model.

Keywords

turnaround radius, large scale structure, dark matter halos, precision cosmology.

1 Introduction

Galaxies represent a key observational window with which we can test our current standard model of cosmology, the ΛCDM model [2]. The structure of the immediate environment around galaxies can be used to infer the parameters of ΛCDM. While this technique is less well-known than LSS methods, it provides an avenue by which some of the ΛCDM parameters[1] can be measured independently from each other – a particularly relevant feature given current debates in cosmology based on parameter degeneracy, such as the evolving Dark Energy debate [1, 3].

The galactic environment can be calculated analytically via excursion set theory, which models the smoothing of the matter density beyond the galactic collapse scale as a random walk in Fourier space [7]. This random walk produces a double probability distribution (here simply called the "double distribution") describing the number density of structures as a function of the mass smoothing scale m and the local linearly-extrapolated overdensity $\tilde{\delta}_l$. The mode of this distribution then gives the most probable profile, which can be compared to simulations [4] and foreseeably could be measured in the sky. The departure of this most probable profile from the simulated/measured profile provides one way by which the turnaround radius can be measured [5]. The turnaround radius is the distance at which matter transitions from inward collapse to cosmological expansion and the point most sensitive to cosmological parameters.

Additionally, higher-order statistics of this distribution may additionally provide physical insight, such as **and I need reasons here.** The double distribution is skewed, and so the inter-quartile range is used to measure its spread. The analytic derivation of the IQR for a given smoothing scale m is given in the Appendix.

> *This repository calculates the double distribution, the most probably profile and other important quantities related to them, for a given late-time cosmological model as defined by a choice of Ω_m and Ω_L.*

[1]Specifically, the present-day matter density Ω_m and the present-day DE density, Ω_L

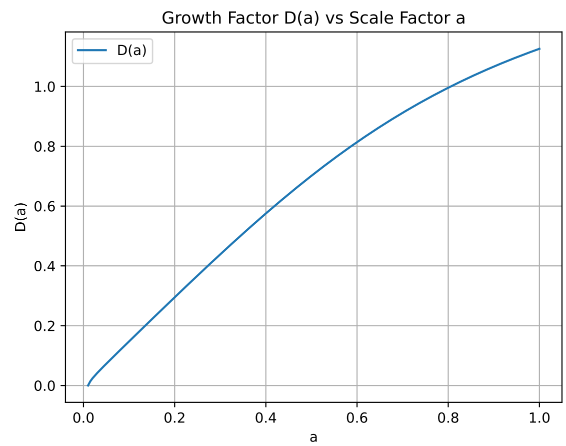

Figure 1: The growth factor as a function of the scale factor, calculated using Eqn. (1) **for concordance ΛCDM cosmology where $\Omega_m = 0.27$ and $\Omega_L = 1 - \Omega_m$.**

2 Methodology

In this section I will describe the mathematics behind each subprogram in this repository. The main variables I will use are:

- n: comoving dimensionless density of dark matter.
- β: the clustering radius for a cloud surrounding an object of mass m, such $\beta \geq 1$ and the mass of the cloud/cluster system is βm.
- δ_l: the (possibly non-linear) matter perturbation, $\delta\rho/\rho$.
- $\tilde{\delta}_l$: the density perturbation *linearly evolved* forward to some time.
- $\tilde{\delta}_c(a)$: the collapse density at scale factor a. Thought to be a Universal constant given a particular cosmology.
- $S(m)$: the variance of the over-density field.
- γ: the power of m where $S(m)$ is assumed to follow a power law in m.

2.1 The growth factor

2.1.1 Analytic form. The growth factor calculates the evolution of linear density perturbations $\tilde{\delta}_l$ given the perturbation at some reference point, $\tilde{\delta}_{l,0}$, via the relation

$$\tilde{\delta}_l = \tilde{\delta}_{l,0} \frac{D(a)}{D(a_0)}.$$

The growth factor is found by evaluating the integral

$$D(a) = \frac{\sqrt{a(1-\Omega_m-\Omega_L)+\Omega_m+\Omega_L a^3}}{a^{3/2}} \times \int_0^a \left(\frac{x}{x(1-\Omega_m-\Omega_L)+\Omega_m+\Omega_L a^3}\right)^{3/2} dx \quad (1)$$

for a particular cosmology, defined by (Ω_m, Ω_L). Note that I use Ω_L where other authors often use Ω_Λ, as in the repository `Omega_L` is used to represent this variable.

This growth factor can be evaluated analytically for two special cases: Einstein-de-Sitter (EdS) cosmology, $\Omega_m = 1$, $\Omega_L = 0$; and free ΛCDM cosmology, $\Omega_m + \Omega_L = 1$. In the case of EdS, $D(a) \propto a$. For free ΛCDM, $D(a) \propto A\left((2\omega)^{1/3}a\right)$ where $\omega = \Omega_L/\Omega_m$ and

$$A(x) = \frac{(x^3+2)^{1/2}}{x^{3/2}} \int_0^x \left(\frac{u}{u^3+2}\right)^{3/2} du.$$

Performing the analytic derivation of this, I found the constant of proportionality to be $\Omega_L(2\omega)^{-2/3}$ and confirmed this with the numerical solution [**NTS**: Add this plot when fixed]. Vaso's notes [6] claim there is no constant of proportionality.

2.1.2 Implementation. The program `growth-factor.py` performs one or more of the following four functions:

(1) Calculate $D(a_i)$, where a_i is defaulted to `pms.a_f`, by numerically integrating Eqn. (1), and print this value to terminal along with the integration error.
(2) Plot $D(a)$ along the scale factor range given by `pms.a_i`, `pms.a_f`, `pms.num_steps`.
(3) Calculate the EdS approximation to $D(a)$ and plot the approximation against the numerical solution.
(4) Calculate the free ΛCDM approximation to $D(a)$ and plot the approximation against the numerical solution. [**NTS**: This is not working right now...]

Running this program with the concordance ΛCDM cosmology should produce a plot of $D(a)$ matching Fig. 1.

2.2 The double distribution

2.2.1 Analytic form. The double distribution is a probability distribution that describes the way that the most massive structures in the Universe are arranged according to their mass scale and their overdensity. It can be derived by counting the number of structures which appear during a random walk in Fourier space where some cut-off mass scale, corresponding to the scale at which overdensities collapse non-linearly to produce galaxies, acts as an absorbing wall. The full derivation is presented in [7]. Here I simply quote the result,

$$\frac{dn}{dm d\tilde{\delta}_l}(m, \tilde{\delta}_l, \beta, a) = \frac{\rho_{m,0}}{m}\left[\frac{\tilde{\delta}_{0,c}(a)-\tilde{\delta}_l}{[S(m)-S(\beta m)]^{3/2}}\right]\left|\frac{dS}{dm}\right|_m \times \exp\left[-\frac{(\tilde{\delta}_{0,c}(a)-\tilde{\delta}_l)^2}{2[S(m)-S(\beta m)]}\right] \times \frac{\exp\left[-\frac{\tilde{\delta}_l^2}{2S(\beta m)}\right] - \exp\left[-\frac{(\tilde{\delta}_l-2\tilde{\delta}_{0,c}(a))^2}{2S(\beta m)}\right]}{2\pi\sqrt{S(\beta m)}}. \quad (2)$$

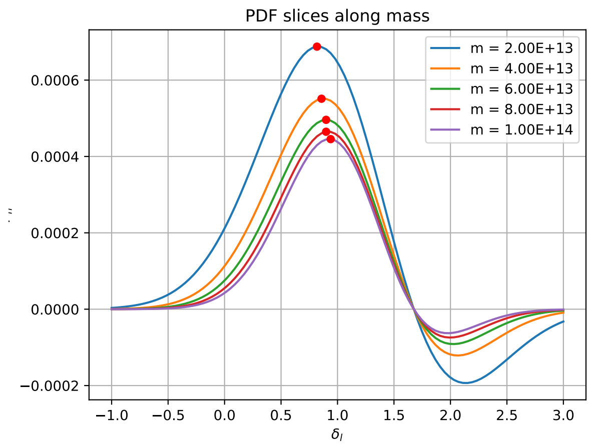

Figure 2: The double distribution Eqn. (3) plotted with respect to the mass scale and linearly extrapolated overdensity, sliced along a range of masses (see legend) with the mode of each slice shown as a red dot. No enforcement of positivity is used here, and it is obvious that above some critical overdensity the assumptions behind the PDF become invalid, as the PDF becomes negative.

We will neglect the exponential term

$$\exp\left[-\frac{(\tilde{\delta}_l - 2\tilde{\delta}_{0,c}(a))^2}{2S(\beta m)}\right]$$

which is called the "structure-in-structure" term, and which accounts for the fact that encountered structures at a particular step may be enclosed in even larger structures. Our focus will be on measuring the turnaround radius, which characterises the largest possible structures, and so this term will be irrelevant for our purposes. This gives our working double distribution,

$$\frac{dn}{dm d\tilde{\delta}_l}(m, \tilde{\delta}_l, \beta, a) = \frac{\rho_{m,0}}{m}\left[\frac{\tilde{\delta}_{0,c}(a)-\tilde{\delta}_l}{[S(m)-S(\beta m)]^{3/2}}\right]\left|\frac{dS}{dm}\right|_m \times \exp\left[-\frac{(\tilde{\delta}_{0,c}(a)-\tilde{\delta}_l)^2}{2[S(m)-S(\beta m)]}\right] \cdot \frac{\exp\left[-\frac{\tilde{\delta}_l^2}{2S(\beta m)}\right]}{2\pi\sqrt{S(\beta m)}}. \quad (3)$$

In order to make contact with Fig. (1) of Ref. [4], we plot this double distribution as a function of the rescaled energy density $\rho/\bar{\rho}_m$ and of the clustering parameter β, instead of the mass m and overdensity $\tilde{\delta}_l$. This transformation is accomplished by using the definition

$$\delta_l = \frac{\rho}{\bar{\rho}_m} - 1$$

to translate ρ into the non-linear overdensity δ_l, and using the scaling relation predicted by spherical collapse[2] to translate this into the linearly extrapolated overdensity, which can then be used directly to find the PDF. As the PDF depends on β, we simply vary

[2]See Sec. 2.3 and Eqn. (5)

β and fix m to an estimate of the virial mass, M_{200}, as given by Sec. 4 of Ref. [4].

2.2.2 Calculation of the collapse time. Most aspects of the double distribution can be set using parameters, but the calculation of $S(m)$ and $\tilde{\delta}_{0,c}(a)$ are significantly more involved. The calculation of $S(m)$ is covered in Sec. 2.3. Here I describe the method used to find $\tilde{\delta}_{0,c}(a)$, which is laid out in Ref. [6] and is given without derivation.

$\tilde{\delta}_{0,c}(a)$ measures the overdensity of a structure which collapses at time a, linearly extrapolated to the present. It naturally depends on the time of collapse, denoted a_{coll}. Given a cosmology, the quantity a_{pta} can be calculated exactly. The collapse time is related to a_{pta} via

$$\int_0^{a_{coll}} \sqrt{\frac{y}{\omega y^3 - \phi y + 1}} dy = \int_0^{a_{pta}} \sqrt{\frac{x}{\omega x^3 - \kappa x + 1}} dx$$

where $\omega = \Omega_L/\Omega_m$, $\phi = (\Omega_m + \Omega_L - 1)/\Omega_m$ and $\kappa \leq 3\omega^{1/3}/2^{2/3}$. I implement a numerical root finding method in order to solve this equation for a_{coll} at a given precision. Then

$$\tilde{\delta}_{c,0} = \frac{3\Omega_m(\kappa - \phi)}{2} D(a_{coll}) \tag{4}$$

where $D(a)$ is the growth factor given by Eqn. (1).

2.2.3 Enforcing positivity. The double distribution relies on some fundamental assumptions which can fail in certain sections of its domain. For instance, this distribution is not expected to produce physical results close to the collapse scale. It is in this region that the distribution can become negative, as shown in Fig. 2.

In order to calculate reasonable sample statistics on this distribution, we have allowed the user to force the double distribution to be positive. This behaviour is controlled by the `enforce_positive_pdf` flag. When this flag is turned on, all negative values of Eqn. (3) are discarded and 0 is returned.

2.2.4 Implementation. The program `double-distribution.py` performs one or more of the following functions:

(1) If the parameter `plot_dimension` = 2, calculate Eqn. 3 as a function of $\rho/\bar{\rho}_m$ and β, and plot the results as a heat-map with marginal distributions.
(2) If `plot_dimension` = 1, calculate Eqn. 3 for a small set of β values, and plot these slices as a function of $\rho/\bar{\rho}_m$, along with the numerically calculated distribution mode.

If the distribution is forced to be positive, and the concordance ΛCDM model is chosen, then option 1 will produce a figure matching Fig. 3, and option 2 will produce a figure matching Fig. 4.

2.3 The density profile

The mode of the double distribution Eqn. (3) gives the most probable distribution of matter as a function of mass and overdensity. This can be used to derive the most probable outer density *profile*, that is, energy density as a function of radius. In order to perform this calculation, we must first be able to translate the linearly extrapolated overdensity $\tilde{\delta}_l$ into the non-linear overdensity δ_l, as this is related directly to the energy density. Assuming overdensities collapse spherically, Ref. [7] derives that

$$\tilde{\delta}_l \approx \tilde{\delta}_c[1 - (1 + \delta_l)^{-1/\tilde{\delta}_c}] \tag{5}$$

where $\tilde{\delta}_c$ is the overdensity of a collapsing structure which has been linearly extrapolated to the time it collapses, and which is thought to be constant within a given cosmology.

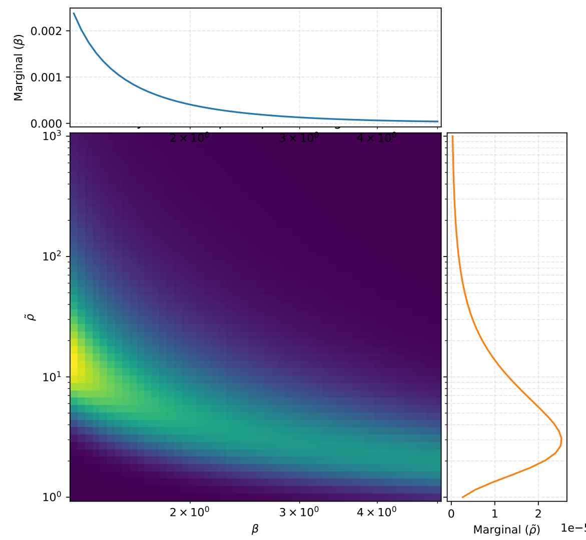

Figure 3: Heatmap of the double distribution Eqn. (3) as a function of the rescaled energy density $\tilde{\rho} \equiv \rho/\bar{\rho}_m$ and the clustering parameter β. The marginal distributions in $\tilde{\rho}$ (orange) and β (blue) are also shown.

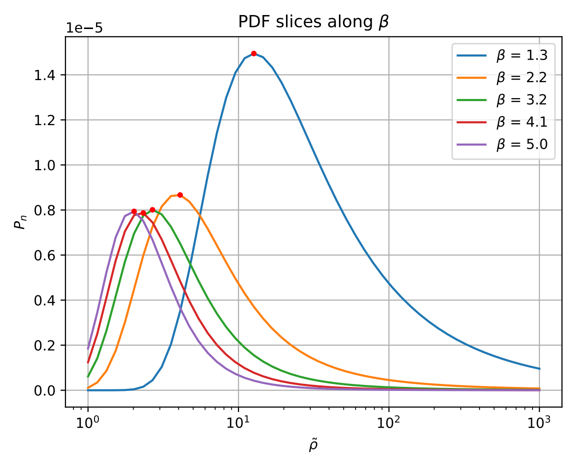

Figure 4: The double distribution shown in Fig. 3, sliced along the β axis at the highest and lowest values, as well as 25%, 50% and 75% of the total range (see the legend). The mode (red dot) of each distribution is calculated by finding the maximum of each vector of PDF values.

By calculating the maximum of the double distribution, Ref. [4] derives that the most probable profile has a linearly extrapolated overdensity which is a simple scaling of the collapse overdensity extrapolated to the present,[3]

$$\tilde{\delta}_l = \tilde{\delta}_{0,c}(a)\frac{S(\beta m)}{S(m)}.$$

Using Eqn. (5), we can then derive

$$1 + \delta_l = \left[1 - \frac{S(\beta m)}{S(m)}\right]^{-\tilde{\delta}_c}. \tag{6}$$

In order to make further progress, we must decide on a functional form of $S(m)$.

2.3.1 Power law approximation. The simplest guess we can make is that $S(m)$ follows a power law,

$$S(m) = S_0\left(\frac{m}{m_0}\right)^{-\gamma}. \tag{7}$$

In this work, I take $(S_0, m_0) = (S_8, m_8)$, which can be measured from cosmological surveys. Ref. [6] shows that this assumption holds for a small range of masses near our region of interest. It also gives us a great amount of analytic power, as the scaling relation we have derived is now *universal*, meaning it holds independent of the virial mass of the system. Under this approximation, the overdensity scaling relation becomes

$$1 + \delta_l = \left[1 - \beta^{-\gamma}\right]^{-\tilde{\delta}_c}.$$

Using the definition of overdensity and integrating over the sphere,[4] one can then derive the universal scaling relations for the most probable outer density profile,

$$\rho(\beta) = \rho_m(1 - \beta^{-\gamma})^{-\tilde{\delta}_c+1} \tag{8}$$

$$r^3(\beta) = (1+\delta_{ta})\left[1 - (1+\delta_{ta})^{-1/\tilde{\delta}_c}\right]^{1/\gamma}\beta(1-\beta^{-\gamma})^{\tilde{\delta}_c} \tag{9}$$

where $r \equiv R/R_{ta}$, the turnaround radius.

2.3.2 Relaxing the power law approximation. We can make a more general calculation of $S(m)$ by using its definition in terms of the power spectrum of perturbations. Assuming the power spectrum follows a power law, a common assumption across all areas of cosmology, the variance of the density field can be calculated by integrating this power law,

$$\begin{aligned} S(m) &\equiv \int_{k=0}^{k(m)} k^2|\delta_k|^2 dk \\ &= S_8\frac{\int_{k=0}^{k(m)} T^2(k)k^{n+2}dk}{\int_{k=0}^{k(m_8)} T^2(k)k^{n+2}dk} \end{aligned} \tag{10}$$

where we have used $|\delta_k|^2 \propto T^2(k)k^n$, and where n is the spectral index and is commonly set to 1 [6]. We then use Bardeen's transfer function for adiabatic CDM,

$$T_{CDM}(k) = \frac{\ln(1+2.34q)}{2.34q}\left[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4\right]^{-1/4}$$

such that the rescaled mode is $q \equiv k/\Omega_m h^2 Mpc^{-1}$. The function $k(m)$ describes the window that we use when stepping through Fourier space. Using a top-hat in Fourier space and solving for the cut-off mode, we find

$$k(m) = \left(\frac{6\pi^2\rho_{m,0}}{m}\right)^{1/3}. \tag{11}$$

2.3.3 Implementation. The program `density-profile.py` calculates the universal outer density profile given by Eqns. (8) and (9). Either the power law form of $S(m)$ or its value in terms of the transfer function may used to perform these calculations. This choice is controlled by the `power_law_approx` parameter.

If the power law approximation is used, the profile is plotted for $\gamma \in (0.55, 0.525, 0.5)$, as these are near the middle of the realistic range as shown by Fig. (A.1) of Ref. [4]. This should produce a figure similar to Fig. 5 when the concordance cosmology is used. If instead the transfer functions are used, a range of masses is chosen and the outer density profile is plotted for each of the resulting values of $k(m)$. In either case, these profiles should roughly align with each other.

The profile is calculated for $\beta \in (1.3, 10)$, as the assumptions made to derive this profile break down below $\beta \approx 1.3$. We use $\tilde{\delta}_c = 1.6865$ for EdS cosmology and $\tilde{\delta}_c = 1.6757$ for concordance ΛCDM cosmology, as given by Apdx. B of Ref. [4]. We further set $\delta_{ta} = 11\Omega_m$ [**NTS**: need to ask Vaso again for the motivation behind this number].

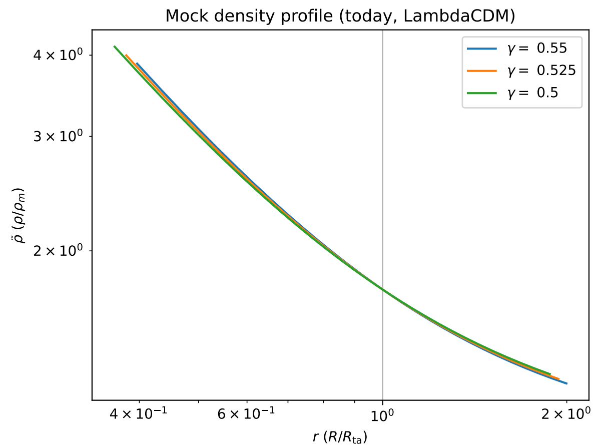

Figure 5: The most probable outer density profile, as a function of the rescaled density $\tilde{\rho}$ and the rescaled radius r, with a grid line showing the turnaround radius R_{ta}. This profile assumes a power law form of $S(m)$ and is plotted for a range of possible power law scalings γ (see legend).

3 Usage

This repository is publicly available and can be downloaded and unpacked according to the following instructions:

(1) `git clone https://github.com/the-florist/galactic-environment-statistics`
(2) `cd galactic-environment-statistics`
(3) `pip install -r requirements.txt`

[3]See Ref. [4] for more details on the derivation.
[4]See Apendix A of Ref. [5] for the full derivation.

The project folder contains one main program file, `main.py`, and three main modules: `growth_factor.py`, `density_profile.py` and `double_distribution.py`. To run one of these modules, run `python main.py i` where `i` is an integer from 1 to 3. This integer will choose which of the modules below to execute:

(1) The growth factor program: calculates the growth factor according to Sec. 2.1, with options to check this growth factor in the EdS case or the ΛCDM case.
(2) The density profile program: calculates the most probable density profile according to Sec. 2.3 for a range of smoothing scales.
(3) The double distribution program: calculates the double distribution according to Sec. 2.2 and plots this as both a 2D heatmap, and at five evenly-spaced slices along the mass axis.

All programs can be run using either a power-law scaling of $S(m)$, or by using the full form of $S(m)$ given by the transfer functions, Eqn. (10). All plots produced are stored in the `plots/` directory, which is created when any of the `src/` modules are called.

The project folder also contains a `util/` folder, which stores the `parameters.py` and `functions.py` files. The parameters file controls all flags and physical parameters used in **any** of the programs described above, including the cosmological model parameters Ω_m and Ω_L. The functions file contains all calculations used in the program files, including all of the expressions described in Sec. 2. Within the program files, all references to the `parameters.py` module are abbreviated `pms`, and all references to the `functions.py` module are abbreviated `func`.

4 Future improvements

Below I list the planned improvements to this repository, sorted into three categories: known bugs, short-term improvements and long-term improvements

4.1 Known bugs

- Profile without power law: Currently the generation of the outer density profile using the transfer function form of $S(m)$ does not produce the correct answer. This is in part because the derivation of this profile is complicated, and I think I took one of the derivatives incorrectly. The units of this profile also need to be confirmed.
- Comparison 2: The second comparison check in the growth factor calculation is not producing the correct result. It did produce the correct result at one point, but a bug must have been introduced which must be found.

4.2 Short-term improvements

- Heat-map computation tracking: Sometimes when the heat-map Fig. 3 is generated for a large number of points, the program takes a while to run. I would like to implement a waiting function that will periodically spit out the fraction of the grid covered during the integration of the PDF, so the user get a sense of how long the program will take to complete.
- Convergence test: I would like to calculate sample statistics on the double distribution, and I would also like to confirm that these statistics converge with the resolution of the distribution. I have sort of implemented a basic convergence test, controlled by the `run_convergence_tests` flag. This was written to test the sample variance, but the sample variance currently cannot be calculated as the distribution is not correctly normalised. Once the normalisation factor is found, this convergence test functionality can be put back into use.

4.3 Long-term improvements

Watch this space...

Acknowledgments

Much of the theoretical work presented here was laid out by Prof. Vasiliki Pavlidou over the course of her career. I would like to thank Prof. Pavlidou for her collaboration on this project, and her mentorship during my visit to the University of Crete. My visit, and the work produced during that time, was generously funded by the GAPSTI fellowship in partnership with the Cambridge Centre for Doctoral Training. I would also like to thank Prof. Pavlidou's current and former students for their insightful comments on this work, in particular Anna Synaniis and Chrysostomos Sidiropoulos. This repository was created within the Cursor IDE, and in particular the plotting of the heat-map shown in Fig. 3 was written by Cursor and edited by the author. This work relies on the NumPy, SciPy and MatPlotLib python libraries.

References

[1] A.G. Adame and et al. 2025. DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations. *Journal of Cosmology and Astroparticle Physics* 2025, 02 (Feb. 2025), 021. doi:10.1088/1475-7516/2025/02/021
[2] Planck Collaboration. 2020. Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics* 641 (Sept. 2020), A6. doi:10.1051/0004-6361/201833910 arXiv:1807.06209 [astro-ph].
[3] George Efstathiou. 2025. Evolving dark energy or supernovae systematics? *Mon. Not. Roy. Astron. Soc.* 538, 2 (2025), 875–882. arXiv:2408.07175 [astro-ph.CO] doi:10.1093/mnras/staf301
[4] Giorgos Korkidis and Vasiliki Pavlidou. 2024. Mass scaling relations for dark halos from an analytic universal outer density profile. *Astronomy & Astrophysics* 690 (Oct. 2024), A71. doi:10.1051/0004-6361/202449704 Publisher: EDP Sciences.
[5] Giorgos Korkidis and Vasiliki Pavlidou. 2025. Hubble expansion signature on simulated halo density profiles - A path to observing the turnaround radius. *Astronomy & Astrophysics* 694 (Feb. 2025), A139. doi:10.1051/0004-6361/202450415 Publisher: EDP Sciences.
[6] Vasiliki Pavlidou. 2025. Vaso's notes on our easy-to-use, analytically derived, universal dark matter density profile of galaxy cluster exteriors. (2025). Unpublished manuscript.
[7] Vasiliki Pavlidou and Brian D. Fields. 2005. Double distribution of dark matter halos with respect to mass and local overdensity. *Physical Review D* 71, 4 (Feb. 2005). doi:10.1103/physrevd.71.043510