The Halo Model

João Victor Rebouças

October 10, 2025

Abstract

Didactical notes about the halo model for the nonlinear matter distribution. These notes are accompanied by a simple Python code that calculates the halo model power spectrum.

1 Introduction

Cosmological linear perturbation theory is able to make predictions for an important phenomenon: the growth of matter inhomogeneities δ_m. The results from the linear approximation imply that the amplitude of Fourier modes $\delta_m(k, a)$ is frozen in superhorizon regimes, $k \ll \mathcal{H}$, where $\mathcal{H}$ is the conformal Hubble factor. Conversely, after it enters the horizon, $k \gg \mathcal{H}$, the amplitude grows proportionally scale factor a. During this growth, nonlinear gravitational effects become important, and the linear perturbation theory predictions are inaccurate.

The spherical collapse is a simple model for nonlinear gravitational growth of a local overdensity. Its core idea is to investigate the growth of a spherical overdense region in an otherwise homogeneous Universe. Despite its simplicity, it provides the entry point for a whole class of nonlinear models of large-scale structure due to one key prediction: the formation of dark matter haloes. These haloes are identified as small-scale regions of space that concentrate much more matter (baryonic and dark) than its surroundings. To illustrate the scales: typical cosmological scales are of the order $1\text{Mpc} - 1\text{Gpc}$; while typical haloes have sizes of tens, hundreds of kpc. In contrast, their typical densities reach hundreds of times the mean density of the Universe. Dark matter haloes have become a central aspect of large-scale structure formation theory, and their existence is supported by observations of rotational curves in galaxies as well as N-body simulations. By modelling the population of haloes, we can obtain predictions for nonlinear matter perturbations, and, with additional information, create models for the galaxy distribution across space.

These notes are organized as follows: Section 2 describes the spherical collapse model, in particular deriving the critical linear overdensity δ_c for the formation of haloes as well as the virialized overdensity Δ_{vir}; Section 3 formalizes the distribution of haloes and derives the nonlinear matter power spectrum in the halo model; Section 4 discusses some technical implementation details; Appendix A contains mathematical derivations of results shown throughout the discussion.

2 Spherical Collapse

2.1 Theory

Suppose that, at a high enough redshift where the Universe is almost homogeneous, we delimit a spherical region of comoving radius R. How much mass does it enclose? Well, at that time,

the Universe has a mean density of $\rho_m(a_i) = \rho_{\rm cr}\Omega_m a_i^{-3}$, where $\rho_{\rm cr}$ is the critical energy density $\rho_{\rm cr} = 3H_0^2/8\pi G$, $\Omega_m \approx 0.3$ is the relative abundance of matter today, and a_i is the scale factor at initial time. Therefore,

$$\rho_{\rm cr}\Omega_m a_i^{-3} = \frac{M}{V_i} = \frac{M}{4\pi r_i^3/3}, \tag{1}$$

where M is the mass enclosed in the region and $r_i = a_i R$ is the physical radius. Therefore, we obtain the relation between the comoving radius and the enclosed mass,

$$M = \frac{4\pi\rho_{\rm cr}\Omega_m R^3}{3}. \tag{2}$$

This equation defines a specific radius, the Lagrangian radius $R_L^3 = 3M/4\pi\rho_{\rm cr}\Omega_m$, associated with a mass M. This radius can be interpreted as following: if a collapsed object such a halo has a mass M, then, before its formation, the matter was scattered in a region of comoving radius R_L.

Now, let's suppose the following scenario: at a_i we take a spherical region of the homogeneous Universe with comoving radius R, and we shrink it to a smaller radius $r_i = a_i R_i < a_i R_L$. This way, we create an overdense region. In this process, we have conserved the total mass M. Therefore, the density ρ_i of the spherical region is

$$\frac{\rho_i}{\bar{\rho}_i} = 1 + \delta_i = \left(\frac{a_i R_L}{r_i}\right)^3. \tag{3}$$

Due to the spherical symmetry of the problem, the spherical overdensity will have its shape conserved; moreover, the total mass is conserved. However, its physical radius r change with time and, in consequence, its density ρ. To find an appropriate evolution equation for r, we simply notice that, inside the spherical overdense region, homogeneity and isotropy still apply, and therefore Friedmann's equations also apply. In particular the second Friedmann equation for the *local* scale factor $\tilde{a}$ inside the sphere,

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

The physical radius is proportional to this local scale factor. Assuming dark energy is negligible, we obtain

$$\begin{aligned} \frac{\ddot{r}}{r} &= -\frac{4\pi G}{3}\frac{M}{4\pi r^3/3}, \\ \ddot{r} &= -\frac{GM}{r^2}. \end{aligned} \tag{5}$$

One can find a parametric solution for this equation in terms of an auxiliary variable θ,

$$r(\theta) = A(1 - \cos\theta), \tag{6a}$$
$$t(\theta) = B(\theta - \sin\theta), \tag{6b}$$

where A and B are constants. In Appendix A.1, I verify that this indeed provides a solution for Equation 5.

The constants A and B can be interpreted by checking the shape of the solution. First, since $dt/d\theta \geq 0$, then θ is monotonically increasing with time, with $\theta = 0$ being equivalent with $t = 0$. The evolution of r is interesting: for small values of θ, r is small and increasing up until $\theta = \pi$,

or equivalently $t = \pi B$, when r reaches its maximum value of $r_{\rm ta} = 2A$; we call this value the *turnaround radius*, and the instant when it happens, $t_{\rm ta} = \pi B$, the *turnaround time*. After that, the radius starts decreasing, up until $\theta = 2\pi$, equivalent to $t = 2t_{\rm ta}$, where $r = 0$. At this point, we say that the overdensity has *collapsed*. We can thus rewrite the solution in terms of $r_{\rm ta}$ and $t_{\rm ta}$,

$$r(\theta) = \frac{r_{\rm ta}}{2}(1 - \cos\theta), \tag{7a}$$

$$t(\theta) = \frac{t_{\rm ta}}{\pi}(\theta - \sin\theta). \tag{7b}$$

The interpretation of the $r \to 0$ limit requires some caution. In practice, this means that the radius is small enough that our underlying assumptions are not valid anymore. At some point, internal processes that occur in this overdense region contribute to stabilize its size to a finite value; the spherical collapse model neglects any influence from the internal structure. Essentially, the overdense region has collapsed into a very small region of space where matter is gravitationally bound: a dark matter halo. Inside this halo, due to the gravitational potential, matter virializes and cools down, favoring the formation of stars and galaxies.

From the results of Appendix A.1, we see that the turnaround radius and time are related by

$$\frac{\pi^2 r_{\rm ta}^3}{8 t_{\rm ta}^2} = GM. \tag{8}$$

With the result for the radius, we can write the density of the overdense region as

$$\rho = \frac{M}{V} = \frac{M}{4\pi r^3/3} = \frac{6M}{\pi r_{\rm ta}^3 (1 - \cos\theta)^3}. \tag{9}$$

From linear theory, we are more used to calculate the overdensity $\delta = \rho/\bar{\rho} - 1$. The background energy density is given by

$$\bar{\rho} = \rho_{\rm cr} \Omega_m a^{-3}, \tag{10}$$

where a is the global scale factor. Therefore,

$$\frac{\rho}{\bar{\rho}} = \frac{6M}{\pi r_{\rm ta}^3 (1 - \cos\theta)^3 \Omega_m \rho_{\rm cr} a^{-3}}. \tag{11}$$

In a matter-dominated Universe, we can solve Friedmann's equation to find how the scale factor changes with time,

$$a \propto t^{2/3}. \tag{12}$$

A proper derivation is shown in Appendix A.2. We can use Equation 87 to write the scale factor as a function of time and Equation 8 to find that

$$\begin{aligned}
\frac{\rho}{\bar{\rho}} &= \frac{6M \times 6\pi G t^2}{\pi r_{\rm ta}^3 (1 - \cos\theta)^3}, \\
\frac{\rho}{\bar{\rho}} &= \frac{36 G M t_{\rm ta}^2 (\theta - \sin\theta)^2}{\pi^2 r_{\rm ta}^3 (1 - \cos\theta)^3}, \\
\frac{\rho}{\bar{\rho}} &= 1 + \delta = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3}.
\end{aligned} \tag{13}$$

We can compare this result to linear theory, where the overdensities grow linearly with the scale factor,

$$\delta_L = \delta_i a/a_i. \tag{14}$$

One way to do this is to expand Equation 13 for small times, where the overdensity is small. At lowest order in θ, $\theta - \sin\theta \approx \theta^3/6$ and $1 - \cos\theta \approx \theta^2/2$. Plugging these into Equation 13, we obtain $\delta \approx 0$, an indication that overdensities are small and linear theory applies. Going to the next-to-leading order, we have

$$\begin{aligned}
\theta - \sin\theta &\approx \frac{\theta^3}{3!} - \frac{\theta^5}{5!} = \frac{\theta^3}{6}\left(1 - \frac{\theta^2}{20}\right), \\
1 - \cos\theta &\approx \frac{\theta^2}{2} - \frac{\theta^4}{24} = \frac{\theta^2}{2}\left(1 - \frac{\theta^2}{12}\right), \\
1 + \delta &\approx \frac{(1-\theta^2/20)^2}{(1-\theta^2/12)^3} \approx \left(1 - \frac{\theta^2}{10}\right)\left(1 + \frac{\theta^2}{4}\right) \approx 1 + \frac{3\theta^2}{20} + \mathrm{O}(\theta^4), \\
\delta &\approx \delta_L = \frac{3\theta^2}{20}.
\end{aligned} \tag{15}$$

Notice that, for small θ,

$$\begin{aligned}
t &= \frac{t_{\rm ta}}{\pi}(\theta - \sin\theta) \approx \frac{t_{\rm ta}}{\pi}\frac{\theta^3}{6}, \\
\theta^2 &= \left(\frac{6\pi t}{t_{\rm ta}}\right)^{2/3}.
\end{aligned} \tag{16}$$

Since $a \propto t^{2/3}$, the overdensity at early times is proportional to the scale factor, as expected from linear theory. Now, suppose that an initial overdensity δ_i is evolved in two ways: using the linear model and the spherical collapse model. By the time the overdensity collapses in the nonlinear model, $t = 2t_{\rm ta}$, the linear overdensity would be[1]

$$\delta_L(2t_{\rm ta}) = \frac{3}{20}(12\pi)^{2/3} \approx 1.68647. \tag{17}$$

This critical overdensity value has important meaning when discussing the halo model using the excursion set theory, or Press-Schechter formalism. If a local linear overdensity grow above this value, this signifies the formation of a dark matter halo.

Finally, when we assume a homogeneous overdensity, we are implicitly assuming that its internal structure is negligible and doesn't play a role in its dynamics. For small radius, this is not true, since the trapped particles virialize, stabilizing the radius in a nonzero value (but small compared to cosmological scales). According to the virial theorem, when the system reaches a dynamical equilibrium, its total kinetic and potential energies are related by

$$U = -2K. \tag{18}$$

The total energy is conserved. At turnaround time, the kinetic energy is zero. Therefore,

$$\begin{aligned}
E = U(t_{\rm ta}) &= K(t_{\rm vir}) + U(t_{\rm vir}) \\
U(t_{\rm ta}) &= U(t_{\rm vir})/2, \\
U = -\frac{3GM^2}{5r} &\implies r(t_{\rm ta}) = r_{\rm ta} = 2r(t_{\rm vir}) \implies \theta_{\rm vir} = 3\pi/2,
\end{aligned} \tag{19}$$

[1] Notice that we cannot simply evaluate Equation 15 at $\theta = 2\pi$. The variable θ is nonlinearly related to time t, and thus it would not yield the correct evolution of the linear overdensity.

which we can plug into Equation 13 to obtain the overdensity at virialization time,

$$\Delta_{\rm vir} = 1 + \delta(\theta_{\rm vir}) \approx 148. \tag{20}$$

A subtle consideration is that, between virialization ($\theta = 3\pi/2$) and collapse ($\theta = 2\pi$), the radius of the halo stays the same while the background density decreases with time. Therefore, the virial overdensity is a bit higher,

$$\Delta_{\rm vir} = \frac{9}{2}\frac{(2\pi)^2}{1} = 18\pi^2 \approx 178. \tag{21}$$

This is a gross estimate of the overdensity enclosed by a dark matter halo. When studying simulations, it is common to define the halo as a region enclosing $\Delta_{\rm vir} = 200$ times the background density.

Finally, it is important to remark that in these derivations, we assume that the Universe is matter-dominated, a good approximation for redshifts $z > 1$. However, for haloes forming at low redshifts $z < 1$, we need to include the effects of dark energy acceleration in the spherical collapse equation 5. This leads to a small decrease in δ_c at low redshifts. While our results are analytical, numerical methods can be used to solve Equation 5 including dark energy.

2.2 Takeaways

- The Spherical Collapse model is the simplest nonlinear model for structure formation, and its equation can be derived using the Friedmann equation or simply a Newtonian equation of motion;
- Nonlinear gravitational effects make overdensities grow much faster compared to linear theory;
- The size of small overdensities initially grows due to the expansion of the Universe, and at some turnaround point it starts shrinking;
- Small overdensities collapse (*i.e.* $r \to 0$) in finite time, forming small-scale virialized structures: the dark matter haloes;
- If an overdense region is evolved using linear theory, if it gets to a critical value of $\delta_c \approx 1.686$, it marks the formation of a dark matter halo if it were evolved using a nonlinear model;
- At some point, the radius of the structure stabilizes due to its internal structure and virialization. The overdensity $\Delta_{\rm vir}$ in the virialized region is hundreds of times larger than the background density; while our calculations yield $\Delta_{\rm vir} \approx 178$, the community usually defines a halo as a region enclosing $\Delta_{\rm vir} = 200$ times the background density.

3 The Halo Model

The spherical collapse model describes a local overdensity; now, we turn our attention to global properties of the matter distribution, assuming all matter is enclosed in these haloes.

3.1 The Variance of Matter Inhomogeneities, $\sigma(R)$

A first definition that will be important to the discussion is the variance of the matter density field (or rather, the density contrast) $\delta(\mathbf{r})$ averaged over spheres of radius R, defined as

$$\sigma^2(R) = \langle \delta_R^2 \rangle = \int d\ln k\ \Delta^2(k)\tilde{T}^2(kR), \tag{22}$$

where

$$\delta_R(\mathbf{x}) = \int d^3y\, T(|\mathbf{x}-\mathbf{y}|)\delta(\mathbf{y}) \tag{23}$$

is the smoothed density field over a sphere of radius R,

$$T_R(\mathbf{x}) = \begin{cases} \frac{1}{4\pi R^3/3}, \text{ if } |\mathbf{x}| < R, \\ 0, \text{ otherwise.} \end{cases} \tag{24}$$

is the spherical top-hat function, $\tilde{T}(kR) = 3j_1(kR)$ is the spherical Bessel function, its Fourier transform, and $\Delta^2(k) = k^3P(k)/2\pi^2$ is the dimensionless matter power spectrum. In Appendix A.3 I derive Equation 22 from the definition of variance.

The intuition about $\sigma^2(R)$ is the following. First, the matter distribution is averaged, or smoothed, over spheres of radius R. Due to inhomogeneities, the resulting average differs from place to place. Then, we average this "smoothed" density field over all space. For a gaussian field, such as the linear matter density field, the average is zero: underdense and overdense regions contribute equally. It follows that the variance is equal to the average of the field squared. While variance is typically a measure of dispersion, $\sigma^2(R)$ denotes the average of δ_R^2, and therefore $\sigma(R)$ is a measure of typical fluctuations in the matter field: higher $\sigma(R)$ values imply higher δ^2 (both in under and overdensities). Historically, we measure this variance is a scale of $R = 8\text{Mpc}/h$, the size of a typical galaxy cluster.

We now bring back two ingredients from spherical collapse: the Lagrangian radius R for a mass M,

$$R_L^3 = \frac{3M}{4\pi\bar{\rho}}, \tag{25}$$

and the linear overdensity threshold for halo formation, $\delta_c \approx 1.686$. We then define the *peak height* ν as

$$\nu = \frac{\delta_c}{\sigma(R_L(M))}. \tag{26}$$

In the following, I denote $\sigma(R) = \sigma_R$, $\sigma(R_L(M)) = \sigma_M$ and $\delta_M = \delta_{R_L(M)}$. The variables R, M and ν are often interchanged.

3.2 The Power Spectrum

The key assumption in the halo model is that all matter lies within haloes. This implies that the spatial matter density can be expressed by a sum over all haloes,

$$\rho(\mathbf{x}) = \sum_i M_i U(|\mathbf{x}-\mathbf{x_i}|, M_i), \tag{27}$$

where i denotes the halo, M_i is the mass of the halo, $\mathbf{x_i}$ is the center of the halo, and $U(r, M)$ is the spherical halo profile, dependent on its mass M. The halo profile is normalized such that

$$\int d^3x\, U(|\mathbf{x}|, M) = 1. \tag{28}$$

We can rewrite the field as

$$\rho(\mathbf{x}) = \sum_i \int dM \int d^3x'\, MU(|\mathbf{x}-\mathbf{x}'|, M)\delta^D(M-M_i)\delta^D(\mathbf{x}'-\mathbf{x_i}). \tag{29}$$

This form is useful for the following reason. Remember that our density field is a *random field* because of the non-deterministic initial conditions from inflation. Each different initial condition configuration will lead to a different distribution of halo positions $\mathbf{x_i}$ and masses M_i. Therefore, all randomness is contained within those quantities.

We define the *halo mass function* as

$$n(M) = \left\langle \sum_i \delta^D(M - M_i)\delta^D(\mathbf{x} - \mathbf{x_i}) \right\rangle, \tag{30}$$

where the bracket $\langle\rangle$ denotes an average over all possible realizations of the initial matter density field. Remember that the Dirac deltas have dimension: $\delta^D(\mathbf{x}-\mathbf{x_i})$ has dimensions of inverse volume, while $\delta^D(M - M_i)$ has dimensions of inverse mass (or energy). Additionally, this function simply counts the haloes that are in position $\mathbf{x_i}$ and have mass M_i. Therefore, $n(M)$ represents a mean number density of haloes per unit mass.

In the literature, you can find $n(M)$ denoted as $\frac{dn}{dM}$, which may be more intuitive since the letter n usually means number density. Additionally, people may refer to the halo mass function in terms of the peak height, $f(\nu)$, which is related to $n(M)$ by a change of variables, $n(M)dM = f(\nu)d\nu$.

Statistical homogeneity relates this averages to *ensemble averages*, where we divide space into large enough regions without causal contact and assume those as independent, finite-size realizations of the Universe.

The average mass density can be calculated as

$$\bar{\rho} = \left\langle \sum_i \int dM \int d^3x' \, MU(|\mathbf{x} - \mathbf{x}'|, M)\delta^D(M - M_i)\delta^D(\mathbf{x}' - \mathbf{x_i}) \right\rangle, \tag{31}$$

$$\bar{\rho} = \int dM \int d^3x' \, MU(|\mathbf{x} - \mathbf{x}'|, M) \left\langle \sum_i \delta^D(M - M_i)\delta^D(\mathbf{x}' - \mathbf{x_i}) \right\rangle, \tag{32}$$

$$\bar{\rho} = \int dM \int d^3x' \, MU(|\mathbf{x} - \mathbf{x}'|, M)n(M). \tag{33}$$

$$\bar{\rho} = \int dM \, Mn(M). \tag{34}$$

The density contrast field is defined as

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}, \tag{35}$$

and its Fourier transform as

$$\delta(\mathbf{k}) = \int d^3x \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{36}$$

From these definitions, we arrive at the following expression[2]

$$\delta(\mathbf{k}) = \frac{1}{\bar{\rho}} \sum_i \int dM \int d^3x' \, MU(k, M)\delta^D(M - M_i)\delta^D(\mathbf{x}' - \mathbf{x_i})e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{37}$$

where $U(k, M)$ is the Fourier transform of the normalized spherical profile $U(r, M)$. With this expression, we are ready to calculate the power spectrum, given by

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = (2\pi)^3 \delta^D(\mathbf{k} + \mathbf{k}')P(k). \tag{38}$$

[2] As $\delta = \rho/\bar{\rho} - 1$, the -1 part contributes with a $\delta^D(\mathbf{k})$ that we are neglecting.

Inserting the expression into the brackets, we obtain

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = \frac{1}{\bar{\rho}^2}\Big\langle \sum_i \int dM \int d^3x M U(k,M) e^{-i\mathbf{k}\cdot\mathbf{x}} \delta^D(M-M_i)\delta^D(\mathbf{x}-\mathbf{x_i}) \sum_j \int dM' \int d^3x' M' U(k',M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} \delta^D(M'-M_j)\delta^D(\mathbf{x}'-\mathbf{x_j})\Big\rangle \tag{39}$$

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = \frac{1}{\bar{\rho}^2}\int dM \int d^3x M U(k,M) e^{-i\mathbf{k}\cdot\mathbf{x}} \int dM' \int d^3x' M' U(k',M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} \Big\langle \sum_i \sum_j \delta^D(M-M_i)\delta^D(\mathbf{x}-\mathbf{x_i})\delta^D(M'-M_j)\delta^D(\mathbf{x}'-\mathbf{x_j})\Big\rangle \tag{40}$$

This sum is broken into two terms: one in which $i = j$ and the other when $i \neq j$.

3.2.1 The 1-halo Term

When $i = j$, we are in the *1-halo regime*, where the correlation is calculated between points that are within the same halo. Therefore, this is the small-scale limit of the power spectrum.

The term

$$\Big\langle \sum_i \delta^D(M-M_i)\delta^D(\mathbf{x}-\mathbf{x_i})\delta^D(M'-M_i)\delta^D(\mathbf{x}'-\mathbf{x_i})\Big\rangle = n(M)\delta^D(M-M')\delta^D(\mathbf{x}-\mathbf{x}') \tag{41}$$

is proportional to the halo mass function: the other two deltas inside the bracket are redundant since they collapse to the same halo positions, and they only contribute when $M = M'$, $\mathbf{x} = \mathbf{x}'$. Inserting this into the correlator, we obtain

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle_{1h} = \frac{1}{\bar{\rho}^2}\int dM \int d^3x M U(k,M) e^{-i\mathbf{k}\cdot\mathbf{x}} \int dM' \int d^3x' M' U(k',M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} \times n(M)\delta^D(M-M')\delta^D(\mathbf{x}-\mathbf{x}'), \tag{42}$$

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle_{1h} = \frac{1}{\bar{\rho}^2}\int dM \int d^3x M^2 n(M) U(k,M) U(k',M) e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}}, \tag{43}$$

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle_{1h} = \frac{1}{\bar{\rho}^2}(2\pi)^3\delta^D(\mathbf{k}+\mathbf{k}')\int dM M^2 n(M) U^2(k,M), \tag{44}$$

$$P_{1h}(k) = \frac{1}{\bar{\rho}^2}\int dM M^2 n(M) U^2(k,M). \tag{45}$$

One thing to notice is that

$$\lim_{k\to 0} U(k,M) = \int d^3x\, U(\mathbf{x},M) = 1. \tag{46}$$

Therefore, at large scales, the 1-halo term contributes as a constant term, the shot noise:

$$\lim_{k\to 0} P^{1h}(k) = \frac{1}{\bar{\rho}^2}\int dM M^2 n(M). \tag{47}$$

3.2.2 The 2-halo Term

Now, when $i \neq j$, we are summing contributions from different haloes, and we call this the *2-halo regime*. The term

$$\left\langle \sum_i \sum_{j \neq i} \delta^D(M - M_i)\delta^D(\mathbf{x} - \mathbf{x_i})\delta^D(M' - M_j)\delta^D(\mathbf{x}' - \mathbf{x_j}) \right\rangle = n(M)n(M')\xi(|\mathbf{x} - \mathbf{x}'|, M, M') \quad (48)$$

is the two-point correlation function between haloes of masses M and M'. Inserting this into the correlator, we find

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle_{2h} = \frac{1}{\bar{\rho}^2} \int dM \int d^3x M U(k, M) e^{-i\mathbf{k}\cdot\mathbf{x}} \int dM' \int d^3x' M' U(k', M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} \times n(M)n(M')\xi(|\mathbf{x} - \mathbf{x}'|, M, M'), \quad (49)$$

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle_{2h} = \frac{1}{\bar{\rho}^2} \int dM \int dM' \int d^3x \int d^3x' M M' n(M)n(M') U(k, M) U(k', M') \xi(|\mathbf{x} - \mathbf{x}'|, M, M') e^{-i\mathbf{k}\cdot\mathbf{x}} e^{-i\mathbf{k}'\cdot\mathbf{x}'}. \quad (50)$$

We perform the $\mathbf{x}'$ integral first:

$$\int d^3x' \xi(|\mathbf{x}' - \mathbf{x}|, M, M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} = \int d^3x' \xi(|\mathbf{x}' - \mathbf{x}|, M, M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} e^{-i\mathbf{k}'\cdot\mathbf{x}} e^{i\mathbf{k}'\cdot\mathbf{x}}, \quad (51)$$

$$\int d^3x' \xi(|\mathbf{x}' - \mathbf{x}|, M, M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} = \int d^3x' \xi(|\mathbf{x}' - \mathbf{x}|, M, M') e^{-i\mathbf{k}'\cdot(\mathbf{x}' - \mathbf{x}')} e^{-i\mathbf{k}'\cdot\mathbf{x}}, \quad (52)$$

$$\int d^3x' \xi(|\mathbf{x}' - \mathbf{x}|, M, M') e^{-i\mathbf{k}'\cdot\mathbf{x}'} = P_{hh}(k', M, M') e^{-i\mathbf{k}'\cdot\mathbf{x}}, \quad (53)$$

where the Fourier transform of the halo 2-point correlation function is the halo power spectrum. Therefore,

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle_{2h} = \frac{1}{\bar{\rho}^2} \int dM \int dM' \int d^3x M M' n(M)n(M') U(k, M) U(k', M') e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}} P_{hh}(k', M, M'), \quad (54)$$

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle_{2h} = (2\pi)^3 \delta^D(\mathbf{k} + \mathbf{k}') \frac{1}{\bar{\rho}^2} \int dM \int dM' M M' n(M)n(M') U(k, M) U(k', M') P_{hh}(k', M, M'), \quad (55)$$

$$P_{2h}(k) = \frac{1}{\bar{\rho}^2} \int dM \int dM' M M' n(M)n(M') U(k, M) U(k, M') P_{hh}(k, M, M'). \quad (56)$$

$$(57)$$

The only thing left is to relate $P_{hh}(k, M, M')$ to $P_L(k)$, the linear matter power spectrum. Since haloes are biased tracers of matter, we define

$$P_{hh}(k, M, M') = b(M)b(M')P_L(k), \quad (58)$$

where $b(M)$ is the linear halo bias. With this assumption, we obtain

$$P_{2h}(k) = \frac{1}{\bar{\rho}^2} \int dM \int dM' M M' n(M)n(M') b(M)b(M') U(k, M) U(k, M') P_L(k), \quad (59)$$

$$P_{2h}(k) = P_L(k) \left(\frac{1}{\bar{\rho}} \int dM\, M n(M) b(M) U(k, M) \right)^2 \quad (60)$$

The final power spectrum is the sum of the 1-halo and 2-halo terms. In the beginning of the discussion, I stated that all matter correlations can be written in terms of halo mass statistics; we managed to write the matter power spectrum in terms of the halo mass function $n(M)$, which characterizes the distribution of mass across the population of dark matter haloes.

3.3 Halo Mass Function

We turn our attention to the following question: how does the mass M distribute across all haloes? Assuming the smoothed matter field $\delta_M(\mathbf{x})$ is gaussian with variance σ_M^2, the probability of forming a halo is the probability of finding $\delta > \delta_c$, and therefore

$$P(\delta > \delta_c) = \frac{1}{\sigma_M\sqrt{2\pi}} \int_{\delta_c}^{\infty} d\delta e^{-\frac{\delta^2}{2\sigma_M^2}} = \frac{1}{2}\text{erfc}\left(\frac{\delta_c}{\sigma_M\sqrt{2}}\right), \tag{61}$$

where $\text{erfc}(x)$ is the complementary error function.

Let's recall the intuition for the Lagrangian radius: if a collapsed structure has mass M , then tracing this mass back to high redshifts, when the matter field was mostly homogeneous, this matter was spread over a radius R_L. Collapsed structures are those points in which $\delta \geq \delta_c$: in particular, $\delta = \delta_c$ would create a halo with mass M, and higher densities denote haloes with higher masses. Therefore, the probability $P(\delta > \delta_c)$ is equivalent to the probability of forming a halo with mass M_h greater than M, $P(M_h > M)$: the complementary cumulative distribution function (CDF) associated with $P(M)$.

Assuming all matter lies within haloes, we would expect $P(M > 0) = 1$. For low masses, or equivalent low R, the matter field tends to the homogeneous background field, and its variance diverges: $\lim_{R\to\infty} \sigma(R) = \infty$. Therefore, since $\text{erfc}(0) = 1$, we get $P(M > 0) = 1/2$, contradicting the assumption that all matter is within haloes. For this reason, an *ad-hoc* factor of 2 is included in this calculation, accounting for underdense regions: $P(M_h > M) = 2P(\delta > \delta_c)$.

Differentiating $P(M_h > M)$ with respect to M, we find the probability density of having a halo with mass M,

$$p(M) = -\frac{dP(M_h > M)}{dM} = -\frac{d}{dM}\text{erfc}\left(\frac{\delta_c}{\sigma_M\sqrt{2}}\right),$$

$$p(M) = \frac{2}{\sqrt{2\pi}} e^{-\frac{\delta_c^2}{2\sigma_M^2}} \frac{\delta_c}{\sigma_M^2}\frac{d\sigma_M}{dM}. \tag{62}$$

$$\tag{63}$$

Using the *peak height* $\nu = \delta_c/\sigma_M$, we can write this as

$$p(M)dM = \frac{2}{\sqrt{\pi}} e^{-\frac{\nu^2}{2}} d\nu = f(\nu)d\nu, \tag{64}$$

where $f(\nu)$ is the *halo mass function*. In our simple case, the halo mass function is the so-called *Press-Schechter* halo mass function,

$$f(\nu) = \frac{2}{\sqrt{\pi}} e^{-\nu^2/2}. \tag{65}$$

In current cosmology, it has been superseded by other functions, such as Sheth-Tormen, Tinker, among others, that agree more with simulations.

Finally, from the halo mass function normalization condition,

$$\int dM \, \frac{M}{\bar{\rho}} n(M) = 1, \tag{66}$$

we can interpret $p(M) = Mn(M)/\bar{\rho}$. Therefore, we obtain the useful change of variables from mass M to peak height ν,

$$\frac{M}{\bar{\rho}} n(M) dM = f(\nu) d\nu. \tag{67}$$

With this change of variables, the matter power spectrum can be written as

$$P_{\rm NL}(k) = P_{1h}(k) + P_{2h}(k), \tag{68a}$$

$$P_{1h}(k) = \frac{1}{\bar{\rho}} \int d\nu M f(\nu) U^2(k, M), \tag{68b}$$

$$P_{2h}(k) = P_L(k) \left(\int d\nu \, n(M) b(M) U(k, M) \right)^2. \tag{68c}$$

The nonlinear power spectrum as well as the separate 1-halo and 2-halo contributions are shown in Figure 1.

3.4 Takeaways

- The (infinite-dimensional) statistics of the matter field are related to the (1-dimensional) statistics of halo masses, encoded in the halo mass function $n(M)$ or $f(\nu)$;
- The nonlinear matter power spectrum can be written as a sum of two terms:
  - The 1-halo term dominates the small-scale matter power spectrum, and is substantially larger than the linear power spectrum. At large scales, the 1-halo term contributes as shot noise;
  - The 2-halo term dominates the large-scale matter power spectrum, and is proportional to the linear matter power spectrum.
- Spherical collapse implies a simple mass function, the Press-Schechter mass function, defined as $f(\nu) = \exp(-\nu^2/2)$. Nowadays, more advanced mass functions are available;
- The necessary ingredients to calculate the nonlinear matter power spectrum are the halo mass function, the normalized halo profile in Fourier space, and the halo bias, typically assumed linear.

4 Implementation

In this section, I discuss technical aspects regarding the implementation of the halo model.

A recent implementation of the halo model is given in pyhalomodel. I used their code to implement a simpler version, myhalomodel, which essentially strips pyhalomodel from all code that is not necessary for a simple didactical discussion.

The codes compute the integrals 68b and 68c. In principle, these integrals need to be computed from zero to infinity. In practice, one first defines a range of masses M and then converts to ν

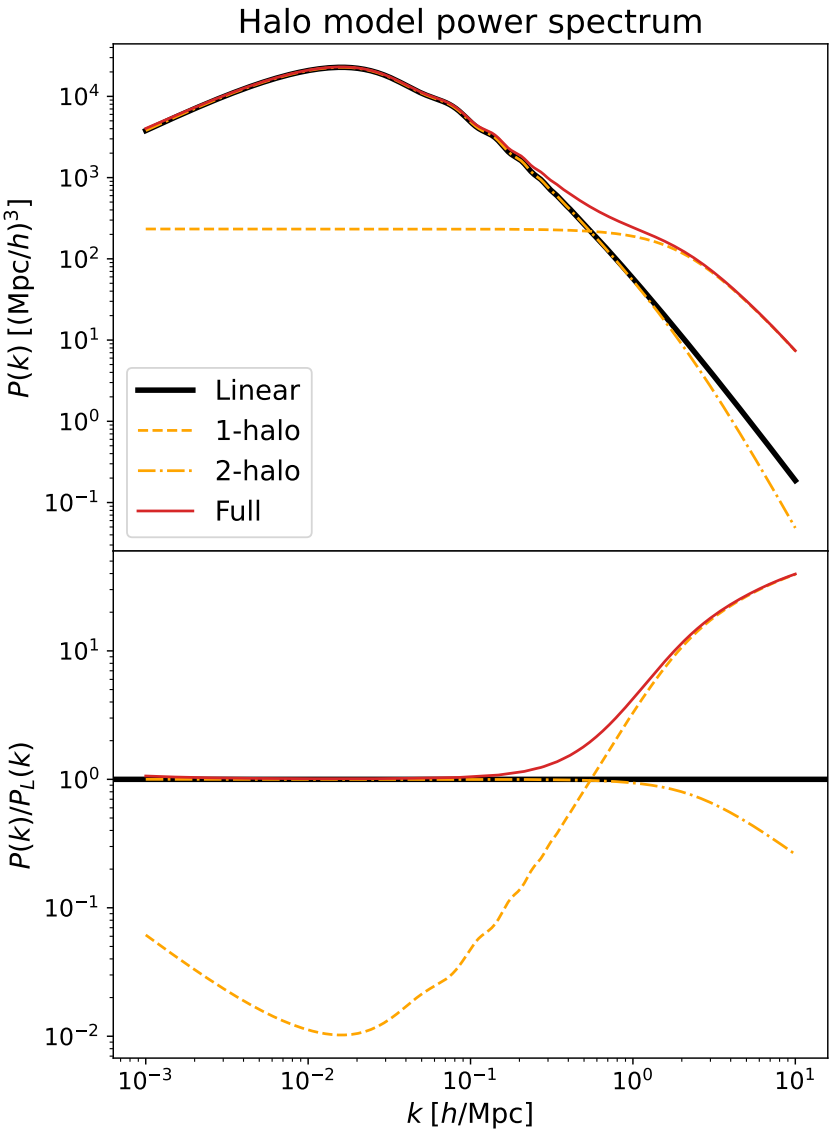

Figure 1: Matter power spectrum at $z = 0$ calculated using the halo model. The top panel shows the linear power spectrum (black, thick), the 1-halo term (orange, dashed), the two-halo term (orange, dashed-dotted) and the full nonlinear power spectrum (red). The bottom panel shows the same spectra divided by the linear one.

using equation 26. In the pyhalomodel, they use an array of 256 log-spaced masses from $10^9 M_\odot$ to $10^{17} M_\odot$.

One important aspect is the Fourier transform of the normalized halo profile. A common, accurate profile of dark matter haloes is given by the Navarro-Frenk-White (NFW) profile, given by

$$\rho(r) = \frac{\rho_s}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2}. \tag{69}$$

Since the halo profile only depends on the mass M, it follows that we must find r_s in terms of M. First, defining the halo as an overdense region with $\Delta = \Delta_{\rm vir} = 200$, we find the halo radius as

$$r_{\rm vir} = R_L(M)/\Delta_{\rm vir}^{1/3}. \tag{70}$$

While there are other definitions of halo, this is the one that we will assume. The NFW profile must be truncated at $r_{\rm vir}$, otherwise its integral is infinite. From simulations, we observe a relation between r_s and r_{200}. This relation can be parametrized by the halo concentration parameter, defined as

$$c = \frac{r_{\rm vir}}{r_s}. \tag{71}$$

From simulations, we observe

$$c \approx 7.85 \left(\frac{M}{2\times 10^{12} M_\odot}\right)^{-0.081} (1+z)^{-0.71}. \tag{72}$$

This provides r_s in terms of M. The parameter ρ_s is not necessary, since it only serves as a normalization constant.

Finally, the Fourier transform of the NFW profile is

$$U(k,M) = A\left\{\sin(kr_s)[{\rm Si}(kr_s+kr_{\rm vir}) - {\rm Si}(kr_s)] + \cos(kr_s)[{\rm Ci}(kr_s+kr_{\rm vir}) - {\rm Ci}(kr_s)] - \frac{\sin(kr_{\rm vir})}{k(r_s+r_{\rm vir})}\right\}, \tag{73}$$

where A is a normalization constant, ${\rm Si}(x)$ and ${\rm Ci}(x)$ are the sine and cosine integral functions defined as,

$${\rm Si}(x) = \int_0^x \frac{\sin(t)}{t} dt, \tag{74}$$

$${\rm Ci}(x) = -\int_0^x \frac{\cos(t)}{t} dt. \tag{75}$$

We find A imposing the normalization condition,

$$\lim_{k\to 0} U(k,M) = 1. \tag{76}$$

The non-trivial terms in the limit are:

$$\lim_{k\to 0}[{\rm Ci}(k(r_s+r_{\rm vir})) - {\rm Ci}(kr_s)] = \log\left(\frac{r_s+r_{\rm vir}}{r_s}\right), \tag{77}$$

$$\lim_{k\to 0} \frac{\sin(kr_s)}{k(r_s+r_{\rm vir})} = \frac{r_s}{r_s+r_{\rm vir}}. \tag{78}$$

Therefore,

$$\lim_{k\to 0} U(k,M) = A\left[\log\left(\frac{r_s+r_{\rm vir}}{r_s}\right) - \frac{r_s}{r_s+r_{\rm vir}}\right] = 1. \tag{79}$$

The normalized NFW profile is shown in Figure 2.

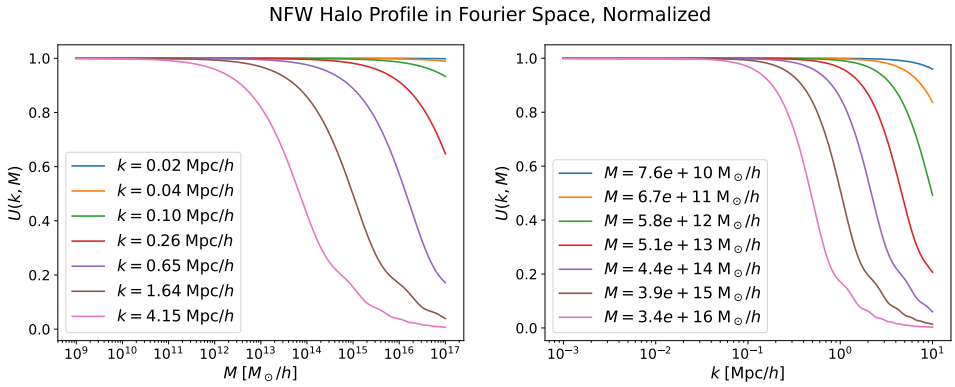

Figure 2: Normalized NFW halo profile in Fourier space. The left panel shows the mass dependence for several fixed k values, and the right panel is the converse.

5 References

- https://arxiv.org/pdf/astro-ph/0206508
- https://arxiv.org/pdf/2007.01936
- https://arxiv.org/abs/2303.08752
- https://arxiv.org/abs/2508.10902

A Derivations

A.1 Verifying the parametric solution

First, I calculate the derivatives of r and t with respect to θ,

$$\frac{dr}{d\theta} = A\sin\theta, \tag{80}$$

$$\frac{dt}{d\theta} = B(1-\cos\theta) = \frac{B}{A}r. \tag{81}$$

Therefore, we can write

$$\dot{r} = \frac{dr}{dt} = \frac{\frac{dr}{d\theta}}{\frac{dt}{d\theta}} = \frac{A\sin\theta}{B(1-\cos\theta)}. \tag{82}$$

Furthermore,

$$\frac{d\dot{r}}{d\theta} = \frac{A}{B}\frac{(1-\cos\theta)\cos\theta - \sin\theta\sin\theta}{(1-\cos\theta)^2} = -\frac{A}{B}\frac{1}{1-\cos\theta} \tag{83}$$

Finally,

$$\begin{aligned}
\ddot{r} &= \frac{\frac{d\dot{r}}{d\theta}}{\frac{dt}{d\theta}} = -\frac{A}{B}\frac{1}{1-\cos\theta}\frac{1}{B(1-\cos\theta)}, \\
\ddot{r} &= -\frac{A}{B^2}\frac{1}{(1-\cos\theta)^2}, \\
\ddot{r} &= -\frac{A^3}{B^2}\frac{1}{r^2}. \qquad (84)
\end{aligned}$$

Therefore, the parametric equations are a solution of the spherical collapse equation if and only if

$$\frac{A^3}{B^2} = GM. \qquad (85)$$

A.2 Scale factor during matter domination

The Friedmann equation reads

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho_{\rm cr}\Omega_m a^{-3}}{3} = Ca^{-3}. \qquad (86)$$

Therefore,

$$\begin{aligned}
\dot{a}^2 &= Ca^{-1}, \\
\dot{a} &= C^{1/2}a^{-1/2}, \\
a^{1/2}da &= C^{1/2}dt, \\
\frac{2}{3}a^{3/2} &= C^{1/2}t + D, \\
a &= C^{1/3}(3/2)^{2/3}t^{2/3}, \\
a^3 &= 9Ct^2/4 = 9\times 8\pi G\rho_{\rm cr}\Omega_m/3\times 1/4\times t^2, \\
a^3 &= 6\pi G\rho_{\rm cr}\Omega_m t^2, \qquad (87) \\
\rho_{\rm cr}\Omega_m a^{-3} &= \frac{1}{6\pi G t^2}. \qquad (88)
\end{aligned}$$

A.3 The variance of the matter field

A.3.1 Spherical top hat

Define the spherical top hat function as

$$T(\mathbf{x}) = \begin{cases} \frac{1}{V}, \text{ if } |\mathbf{x}| < R; \\ 0, \text{ otherwise}, \end{cases} \qquad (89)$$

where $V = 4\pi R^3/3$. Its Fourier transform is given by

$$\begin{aligned}
\tilde{T}(\mathbf{k}) &= \int d^3x W(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \\
\tilde{T}(\mathbf{k}) &= \int r^2 \sin\theta \, dr d\theta d\phi \, W(r) e^{-ikr\cos\theta}, \\
\tilde{T}(\mathbf{k}) &= 2\pi \int r^2 \sin\theta \, dr d\theta \, W(r) e^{-ikr\cos\theta}, \\
\tilde{T}(\mathbf{k}) &= 2\pi \int r^2 \, dr du \, W(r) e^{-ikru}, \\
\tilde{T}(\mathbf{k}) &= 2\pi \int r^2 \, dr \, W(r) \frac{1}{-ikr} e^{-ikru}|_{u=-1}^{u=1}, \\
\tilde{T}(\mathbf{k}) &= 4\pi \int r^2 \, dr \, W(r) \frac{\sin(kr)}{kr}, \\
\tilde{T}(\mathbf{k}) &= \frac{4\pi}{k} \int dr \, rW(r) \sin(kr), \\
\tilde{T}(\mathbf{k}) &= \frac{3}{kR^3} \int_0^R dr \, r \sin(kr), \\
\tilde{T}(\mathbf{k}) &= \frac{3(\sin(kR) - kR\cos(kR))}{k^3 R^3} = 3j_1(kR),
\end{aligned}$$

where j_1 is the spherical Bessel function. Notice that, since the original top hat function is spherically symmetric, its Fourier transform shares this property.

A.3.2 The variance

First, we define the average square matter density in a sphere centered at $\mathbf{x}$ with radius R,

$$\delta_R(\mathbf{x}) = \int d^3y W(|\mathbf{x} - \mathbf{y}|)\delta(\mathbf{y}). \tag{90}$$

Since this is a convolution, it follows that

$$\delta_R(\mathbf{k}) = \tilde{W}(kR)\delta(\mathbf{k}). \tag{91}$$

Now, we average over all possible matter field realizations or, equivalently, over all spheres of radii R,

$$\sigma^2(R) = \langle \delta_R(\mathbf{x})^2 \rangle = \int \frac{d^3k_1}{(2\pi)^3} \int \frac{d^3k_2}{(2\pi)^3} \tilde{W}(k_1R)\tilde{W}(k_2R) \left\langle \delta_R(\mathbf{k_1})\delta_R(\mathbf{k_2}) \right\rangle . \tag{92}$$

From statistical homogeneity, $\langle \delta_R(\mathbf{k_1})\delta_R(\mathbf{k_2}) \rangle = (2\pi)^3 \delta^{(D)}(\mathbf{k_1} + \mathbf{k_2}) P(k_1)$. Therefore,

$$\begin{aligned}
\sigma^2(R) &= \int \frac{d^3k}{(2\pi)^3} \tilde{W}^2(kR) P(k), \\
\sigma^2(R) &= \int d\ln k \; \tilde{W}^2(kR) \Delta^2(k). \qquad (93) \\
& \qquad (94)
\end{aligned}$$