Spherical Collapse

João Victor Rebouças

September 2025

Abstract

Didactical notes about the spherical collapse model of structure formation.

1 Introduction

Cosmological linear perturbation theory is able to provide an important prediction: the growth of matter inhomogeneities δ_m. Long story short, 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.

In this context, the spherical collapse model is a simple model for nonlinear structure formation in the Universe. Its core idea is to investigate the growth of a spherical region of overdensity 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. By small-scale, I mean the following: 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.

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

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

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, 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 $t_{\rm ta}$,

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

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

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, we see that the turnaround radius and time are related by

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

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

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

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

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

where a is the global scale factor. Therefore,

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

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 B. We can use Equation 29 to write the scale factor as a function of time and Equation 8 to find that

$$\frac{\rho}{\bar{\rho}} = \frac{6M \times 6\pi G t^2}{\pi r_{\text{ta}}^3 (1 - \cos\theta)^3},$$

$$\frac{\rho}{\bar{\rho}} = \frac{36 G M t_{\text{ta}}^2 (\theta - \sin\theta)^2}{\pi^2 r_{\text{ta}}^3 (1 - \cos\theta)^3},$$

$$\frac{\rho}{\bar{\rho}} = 1 + \delta = \frac{9}{2} \frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3}. \tag{13}$$

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,

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

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, $t = 2t_{\mathrm{ta}}$, the linear overdensity would be[1]

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

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,

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

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

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

[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.

More accurately, between virialization ($\theta = 3\pi/2$) and collapse ($\theta = 2\pi$), the radius 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. Again, this number is important when studying the halo model and predicting statistical properties of haloes.

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.

3 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.

A Verifying the parametric solution

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

$$\frac{dr}{d\theta} = A\sin\theta, \tag{22}$$

$$\frac{dt}{d\theta} = B(1-\cos\theta) = \frac{B}{A}r. \tag{23}$$

Therefore, we can write

$$\dot{r} = \frac{dr}{dt} = \frac{\frac{dr}{d\theta}}{\frac{dt}{d\theta}} = \frac{A\sin\theta}{B(1-\cos\theta)}. \tag{24}$$

Furthermore,

$$\frac{d\dot{r}}{d\theta} = \frac{A}{B}\frac{(1-\cos\theta)\cos\theta - \sin\theta\sin\theta}{(1-\cos\theta)^2} = -\frac{A}{B}\frac{1}{1-\cos\theta} \tag{25}$$

Finally,

$$\begin{aligned}
\ddot{r} &= \frac{\frac{d\dot{r}}{d\theta}}{\frac{dt}{d\theta}} = -\frac{A}{B}\frac{1}{1-\cos\theta}\frac{1}{B(1-\cos\theta)}, \\
\ddot{r} &= -\frac{A}{B^2}\frac{1}{(1-\cos\theta)^2}, \\
\ddot{r} &= -\frac{A^3}{B^2}\frac{1}{r^2}.
\end{aligned} \tag{26}$$

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

$$\frac{A^3}{B^2} = GM. \tag{27}$$

B Scale factor during matter domination

The Friedmann equation reads

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho_{\rm cr}\Omega_m a^{-3}}{3} = Ca^{-3}. \tag{28}$$

Therefore,

$$\begin{aligned}
\dot{a}^2 &= Ca^{-1}, \\
\dot{a} &= C^{1/2}a^{-1/2}, \\
a^{1/2}da &= C^{1/2}dt, \\
\frac{2}{3}a^{3/2} &= C^{1/2}t + D, \\
a &= C^{1/3}(3/2)^{2/3}t^{2/3}, \\
a^3 &= 9Ct^2/4 = 9 \times 8\pi G\rho_{\rm cr}\Omega_m/3 \times 1/4 \times t^2, \\
a^3 &= 6\pi G\rho_{\rm cr}\Omega_m t^2, && (29) \\
\rho_{\rm cr}\Omega_m a^{-3} &= \frac{1}{6\pi G t^2}. && (30)
\end{aligned}$$