The spherical collapse model

We will in the following consider what is perhaps the simplest model for the formation of non-linear (gravitationally bound) structures. Imagine that in an Einstein-de Sitter (that is, spatially flat and matter-dominated) universe we have a spherical region with density higher than the density of the background (which is equal to the critical density). According to general relativity, the evolution of this spherical overdensity will be independent of the background evolution, so that it will evolve exactly like a sub-universe with density higher than the critical density. For a positively curved matter-dominated universe the Friedmann equations have the parametric form

$$R_p = A(1 - \cos\theta) \tag{1}$$

$$t = B(\theta - \cos\theta), \tag{2}$$

where I have denoted the scale factor of the sub-universe by R_p, and the constants A and B are given by

$$A = \frac{\Omega_{m0}}{2(\Omega_{m0} - 1)} \tag{3}$$

$$B = \frac{\Omega_{m0}}{2H_0(\Omega_{m0} - 1)^{3/2}}, \tag{4}$$

where $\Omega_{m0} > 1$ is the density parameter of this sub-universe, and H_0 its Hubble constant. According to equations (1) and (2) the spherical region will expand until $\theta = \pi$, when $R_p = R_p^{max} = 2A$, then turn around and collapse, formally reaching $R_p = R_p^{min} = 0$ and infinite density when $\theta = 2\pi$. We have

$$R_p^{max} = 2A = \frac{\Omega_{m0}}{\Omega_{m0} - 1} \tag{5}$$

$$t_{max} = \pi B = \frac{\pi\Omega_{m0}}{2H_0(\Omega_{m0} - 1)^{3/2}}. \tag{6}$$

At that time the density of the spherical region compared to the EdS background is

$$\frac{\rho}{\rho_0} = \frac{\Omega_{m0}\rho_{c0}\frac{1}{(R_p^{max})^3}}{\rho_{c0}\frac{1}{a^3}} = \Omega_{m0}\left(\frac{a}{R_p^{max}}\right)^3. \tag{7}$$

Here a is the scale factor of the background given by $a = \left(\frac{3}{2}H_0 t\right)^{2/3}$, so that

$$a^3(t_{max}) = \left(\frac{3}{2}H_0 t_{max}\right)^2 = \frac{9\pi^2}{16}\frac{\Omega_{m0}^2}{(\Omega_{m0}-1)^3}. \tag{8}$$

Therefore

$$\begin{aligned}\frac{\rho}{\rho_0} &= \Omega_{m0}\frac{9\pi^2}{16}\frac{\Omega_{m0}^2}{(\Omega_{m0}-1)^3}\frac{(\Omega_{m0}-1)^3}{\Omega_{m0}^3} \\ &= \frac{9\pi^2}{16} \approx 5.55. \end{aligned} \tag{9}$$

So the spherical perturbation starts to collapse when its density has reached 5.55 times the background density.

In this simple-minded model matter has no internal pressure, so there is nothing stopping the spherical blob to collapse to infinite density. If the redshift at which the region reaches maximum size is z_{max}, the redshift at which it has collapsed completely, z_c, is given by

$$\frac{1+z_c}{1+z_{max}} = \frac{a_{max}}{a_c} = \left(\frac{t_{max}}{t_c}\right)^{2/3} = \frac{1}{2^{2/3}}, \tag{10}$$

i.e.,

$$1+z_c = \frac{1+z_{max}}{2^{2/3}}. \tag{11}$$

In real life collapse will, of course, stop before infinite density is reached. The end result will be a system which satisfies the virial theorem, such that the kinetic energy E_k and the potential energy E_p satisfy

$$E_k = -\frac{1}{2}E_p. \tag{12}$$

We can find the dimensions of the structure by the following argument: At z_{max} the sphere momentarily stands still before starting to collapse, so all its energy is potential energy,

$$E = E_p = -\frac{3GM^2}{5r_{max}}, \tag{13}$$

where r is its physical size. When it has collapsed to $r = r_{max}/2$, we have

$$E_p = -\frac{3GM^2}{5r_{max}/2} = -\frac{6GM^2}{5r_{max}}, \tag{14}$$

and from conservation of energy we find

$$E_k = E - E_p = -\frac{3GM^2}{5r_{max}} + \frac{6GM^2}{5r_{max}} = \frac{3GM^2}{5r_{max}} = -\frac{1}{2}E_p(r_{max}/2). \qquad (15)$$

So the virial theorem is satisfied when the spherical region has collapesed to half its maximum size, and we say that it has *virialized* and become stable. It is now 8 times as dense as when it started to collapse. The time when this occurs si found by requiring

$$R_p^{vir} = \frac{1}{2}R_p^{max}, \qquad (16)$$

that is,

$$A(1 - \cos\theta) = A, \qquad (17)$$

and we must have $\theta > \pi$ to correspond to times after the onset of collapse. The solution is then $\theta = 3\pi/2$, giving

$$\begin{aligned} t_{vir} &= B\left(\frac{3\pi}{2} + 1\right) = \pi B\left(\frac{3}{2} + \frac{1}{\pi}\right) \\ &= \left(\frac{3}{2} + \frac{1}{\pi}\right) t_{max} \approx 1.81 t_{max} \end{aligned} \qquad (18)$$

Since turnaround the background EdS density has decreased further by a factor

$$\left(\frac{a_{max}}{a_{vir}}\right)^3 = \left(\frac{t_{max}}{t_{vir}}\right)^2 = \frac{1}{1.81^2}. \qquad (19)$$

At virialization, then, the spherical perturbation has a density which is larger than the background density by a factor of

$$5.55 \times 8 \times 1.81^2 \approx 145. \qquad (20)$$

In many books and papers you will find that the authors take t_{vir} to be the time when $R_p = 0$, so that $t_{vir} = t(\theta = 2\pi) = 2t_{max}$. In that case the background density has dropped by a factor $1/2^2 = 1/4$, and the ratio of the density of the blob to the background becomes

$$5.55 \times 8 \times 4 \approx 178. \qquad (21)$$

The main message is the same in both cases: Perturbations form graviationally bound structures when they become 150-200 times as dense as the background. This simple model is consistent with results from N-body simulations: Galaxies and clusters of galaxies separate out as distinct graviationally bound structures when their densities are at least 100 times greater than the background density.

We can estimate the redshift at which virialization happens in a general CDM model, assuming that the same results as above remain approximately true even when the background density differs from the critical. Denote the density parameter of the background by Ω_{m0}^{bg}. Our approximate criterion for virialization is then

$$\rho_{vir} \geq 100 \times \frac{3\Omega_{m0}^{bg}}{8\pi G}(1+z_{vir})^3. \tag{22}$$

We take the blob to be made up of particles with velocity dispersion σ, and use the virial theorem. Since the calculation is approximate, we allow ourselves to be careless with factors of order 1, and write the virial theorem as

$$\frac{1}{2}M\sigma^2 = \frac{1}{2}\frac{GM^2}{R}, \tag{23}$$

which gives

$$R = \frac{GM}{\sigma^2}. \tag{24}$$

The density of the spherical region at virialization is then

$$\rho_{vir} = \frac{M}{\frac{4\pi}{3}R^3} = \frac{M}{\frac{4\pi}{3}\frac{G^3M^3}{\sigma^6}} = \frac{3\sigma^6}{4\pi G^3 M^2}. \tag{25}$$

So the condition for virialization becomes

$$\frac{3\sigma^2}{4\pi G^3 M^2} \geq 100 \times \frac{3H_0^2\Omega_{m0}^{bg}}{8\pi G}(1+z_{vir})^3, \tag{26}$$

which, after solving for $1+z_{vir}$ and inserting numbers, gives

$$1+z_{vir} \leq 0.47\left(\frac{\sigma}{100\ \mathrm{km\,s^{-1}}}\right)^2\left(\frac{M}{10^{12}\ M_\odot}\right)^{-2/3}(\Omega_{\mathrm{m0}}^{bg}h^2)^{-1/3}. \tag{27}$$

Using the value presently favoured by observations, $\Omega_{m0}^{bg}h^2 = 0.13$, we finally have

$$1+z_{vir} \leq 0.93\left(\frac{\sigma}{100\ \mathrm{km\,s^{-1}}}\right)^2\left(\frac{M}{10^{12}\ M_\odot}\right)^{-2/3}. \tag{28}$$

This is a crude result, but it gives some insights into when structures of different sizes formed. A large galaxy like the Milky Way has $\sigma \sim 300\ \mathrm{km\,s^{-1}}$ and $M \sim 10^{12}\ M_\odot$, and we find $1 + z_{vir} \leq 8$. Thus we would be surprised to find a fully-formed large galaxy at redshifts significantly larger than 10. Rich galaxy clusters have $\sigma \sim 1000\ \mathrm{km\,s^{-1}}$ and $M \sim 10^{15}\ M_\odot$, which gives $1 + z_{vir} \leq 0.93$. This is clearly inaccurate, but still indicates that clusters of galaxies formed fairly recently.

Finally, if we assume equality in (28) and assume hydrostatic equilibrium is reached so that the temperature T satisfies $T \propto \sigma^2$, it can be shown that

$$\frac{k_B T}{7\ \mathrm{keV}} = \left(\frac{M}{10^{15} h^{-1}\ M_\odot}\right)^{2/3} (1 + z_{vir}), \tag{29}$$

which makes it seem reasonable that gas in galaxy clusters radiate X-rays.

We can analyse the spherical collapse model in another way to make the connection to linear perturbation theory by expanding equations (1) and (2) as power series in θ:

$$\begin{aligned} R_p &= A(1 - \cos\theta) \approx A\left(1 - 1 + \frac{1}{2}\theta^2 - \frac{1}{24}\theta^4\right) \\ &= A\left(\frac{1}{2}\theta^2 - \frac{1}{24}\theta^4\right) \qquad (30) \\ t &= B(\theta - \sin\theta) \approx B\left(\theta - \theta + \frac{1}{6}\theta^3 - \frac{1}{120}\theta^5\right) \\ &= B\left(\frac{1}{6}\theta^3 - \frac{1}{120}\theta^5\right). \qquad (31) \end{aligned}$$

If we keep just the first term in each expression, we have

$$R_p = \frac{1}{2}A\theta^2 \tag{32}$$

$$t = \frac{1}{6}B\theta^3 \tag{33}$$

from which we see that $R_p \propto t^{2/3}$, that is, the same expansion as the uniform background.

Since we have $R_p^{max} = 2A$ and $t_{max} = \pi B$, we can write equations (30) and (31) as

$$\frac{R_p}{R_p^{max}} \approx \frac{1}{4}\theta^2 - \frac{1}{48}\theta^4 \tag{34}$$

$$\frac{t}{t_{max}} = \frac{1}{\pi}\left(\frac{1}{6}\theta^3 - \frac{1}{120}\theta^5\right). \tag{35}$$

What we want to do is to find R_p as a functin of t. We start by rewriting equation (35) as

$$\theta^3 = 6\pi \frac{t}{t_{max}} + \frac{1}{20}\theta^5, \tag{36}$$

and imagine solving iteratively for θ as a function of t. The first guess is simply

$$\theta_{(0)}^3 = 6\pi \frac{t}{t_{max}}, \tag{37}$$

i.e.,

$$\theta_{(0)} = \left(6\pi \frac{t}{t_{max}}\right)^{1/3}. \tag{38}$$

In the next iteration we insert $\theta_{(0)}$ on the right-hand side of (36):

$$\begin{aligned} \theta_{(1)}^3 &= 6\pi \frac{t}{t_{max}} + \frac{1}{20}\left(6\pi \frac{t}{t_{max}}\right)^{5/3} \\ &= 6\pi \frac{t}{t_{max}} \left[1 + \frac{1}{20}\left(6\pi \frac{t}{t_{max}}\right)^{2/3}\right], \end{aligned} \tag{39}$$

so that

$$\theta_{(1)} = \left(6\pi \frac{t}{t_{max}}\right)^{1/3} \left[1 + \frac{1}{20}\left(6\pi \frac{t}{t_{max}}\right)^{2/3}\right]^{1/3}. \tag{40}$$

We are interested in the limit where $x \equiv 6\pi t/t_{max}$ is a small quanitity, so we can expand equation (39) and get

$$\theta_{(1)} \approx x^{1/3}\left(1 + \frac{1}{60}x^{2/3}\right) = x^{1/3} + \frac{1}{60}x, \tag{41}$$

We can now insert this in the expression for R_p:

$$\begin{aligned} \frac{R_p}{R_p^{max}} &= \frac{1}{4}\theta_{(1)}^2 - \frac{1}{48}\theta_{(1)}^4 \\ &= \frac{1}{4}x^{2/3}\left(1 + \frac{1}{60}x^{2/3}\right)^2 - \frac{1}{48}x^{4/3}\left(1 + \frac{1}{60}x^{2/3}\right)^4 \\ &\approx \frac{1}{4}x^{2/3}\left(1 + \frac{1}{30}x^{2/3}\right) - \frac{1}{48}x^{4/3} \\ &= \frac{1}{4}x^{2/3}\left(1 - \frac{1}{20}x^{2/3}\right). \end{aligned} \tag{42}$$

Inserting for x we get

$$\frac{R_p}{R_p^{max}} = \frac{1}{4}\left(6\pi\frac{t}{t_{max}}\right)^{2/3}\left[1-\frac{1}{20}\left(6\pi\frac{t}{t_{max}}\right)^{2/3}\right] \equiv R_{lin}. \tag{43}$$

The first factor is the background evolution, and the second describes the linear evolution of the perturbation. Both the perturbation and the outside universe are matter-dominated, so in both the density varies as the inverse cube of the scale factor. Writing the density of the blob in linear perturbation theory as $\rho = \rho_{bg}(1+\delta_{lin})$, we have

$$1+\delta_{lin} = \left(\frac{a}{R_{lin}}\right)^3, \tag{44}$$

so

$$\frac{R_{lin}}{a} = (1+\delta_{lin})^{-1/3} \approx 1 - \frac{1}{3}\delta. \tag{45}$$

Since $a = \frac{1}{4}(6\pi t/t_{max})^{2/3}$, we have

$$\frac{R_{lin}}{R_p^{max}} = \frac{1}{4}\left(6\pi\frac{t}{t_{max}}\right)^{2/3}\left(1-\frac{1}{3}\delta\right) \tag{46}$$

and combaring this to equation (43) we find

$$\delta = \frac{3}{20}\left(6\pi\frac{t}{t_{max}}\right)^{2/3}. \tag{47}$$

We can now find out what *really* happens, at least within this simplified model, when we extrapolate linear perturbation theory. For example, at when the spherical region starts to collapse, at $t = t_{max}$, linear perturbation theory predicts that $\delta_{lin}^{turn} = 3(6\pi)^{2/3}/20 \approx 1.06$. You may recall that the exact calculation gave that the density of the sphere was 5.55 times as dense as the background universe at this point Phrased in a different way, when linear perturbation theory gives a density contrast of 1.06, the structure has in reality began to collapse to form a bound object. Taking virialization to occur at $t = 2t_{max}$, the linear density contrast has at this point increased to

$$\delta_{lin}^{vir} \equiv \delta_c = \frac{3}{20}\left(6\pi\frac{2t_{max}}{t_{max}}\right)^{2/3} \approx 1.686. \tag{48}$$

Thus, a linear-theory density contrast of 1.686 corresponds to the time of complete gravitational collapse of a spherical perturbation. This value is used in analytical treatments of the growth of structure in the universe, such as the Press-Schechter formalism. The actual (in this model) non-linear density contrast at $t = 2t_{max}$ we found to be

$$1 + \delta_{nl}^{vir} \approx 178. \tag{49}$$

Virialization with Λ

Since redshifts of around 0.5 the universe has been accelerating, possibly because of the cosmological constant. This is not too far away from the time when clusters virialized according to the spherical collapse model, and it therefore makes sense to investigate the effect of Λ on virialization. We will follow the approach of Lahav, Lilje, Primack & Rees, MNRAS 251 (1991) 128.

The cosmological constant can be treated in Newtonian theory as an additional contribution to the potential energy. For a spherical shell of radius R enclosing a mass m, the energy per unit mass is given by

$$\mathcal{E} = \frac{1}{2}\dot{R}^2 - \frac{Gm}{R} - \frac{1}{6}\Lambda R^2. \tag{50}$$

This expression can now be integrated to find the potential energy of a massive sphere. For a uniform sphere with density ρ and mass $M = 4\pi R^3\rho/3$ we have already found the contribution from the standard graviational potential, and it is equal to

$$E_{p,G} = -\frac{3GM^2}{5R}. \tag{51}$$

For the Λ term the contribution from a shell of radius x and thickness dx is

$$dE_{p,\Lambda} = -\frac{1}{6}\Lambda x^2 \times 4\pi x^2 \rho dx = -\frac{2\pi\rho}{3}\Lambda x^4 dx, \tag{52}$$

and integrating from 0 to R we find

$$E_{p,\Lambda} = -\frac{2\pi\rho\Lambda}{3}\int_0^R x^4 dx = -\frac{2\pi\rho\Lambda}{3}\frac{1}{5}R^5 = -\frac{1}{10}\Lambda M R^2. \tag{53}$$

For a collapsing sphere of radius R and mass M we have $\dot{R} = 0$ at turnaround, and denoting the maximum value of R by R_t we can write the total energy, which is conserved, as

$$E = E_{p,G,t} + E_{p,\Lambda,t} = -\frac{3GM^2}{5R_t} - \frac{1}{10}\Lambda M R_t^2. \tag{54}$$

It can be shown that for a potential energy $E_p \propto R^n$, a generalization of the virial theorem states that $E_k = \frac{n}{2} E_p$. Thus, in the final, virialized state our sphere satisfies

$$E_{k,f} = -\frac{1}{2} E_{p,G,f} + E_{p,\Lambda,f}. \tag{55}$$

Conservation of energy gives

$$E_{k,f} + E_{p,G,f} + E_{p,\Lambda,f} = E_{p,G,t} + E_{p,\Lambda,t}. \tag{56}$$

We use equation (55) to eliminate $E_{k,f}$ and find

$$\frac{1}{2} E_{p,G,f} + 2E_{p,\Lambda,f} = E_{p,G,t} + E_{p,\Lambda,t}, \tag{57}$$

and when we use (51) and (53) we find

$$\frac{3GM}{R_f} + 2\Lambda R_f^2 = \frac{6GM}{R_t} + \Lambda R_t^2, \tag{58}$$

where R_f is the radius of the sphere at virialization. Introducing the dimensionless variable $x = R_f/R_t$, we can write equation (58) as

$$\frac{3GM}{xR_t} + 2\Lambda x^2 R_t^2 = \frac{6GM}{R_t} + \Lambda R_t^2, \tag{59}$$

that is,

$$1 + 2\frac{\Lambda R_t^3}{3GM} x^3 = 2x + \frac{\Lambda R_t^3}{3GM} x. \tag{60}$$

The density of the sphere at turnaround is $\rho_t = 3M/4\pi R_t^3$, and we use this to introduce yet another dimensionless parameter:

$$\eta = \frac{\Lambda R_t^3}{3GM} = \frac{\Lambda}{4\pi G \rho_t}. \tag{61}$$

For realistic physical parameters, η turns out to be a small number. We can now write equation (60) as

$$1 + 2\eta x^3 = 2x + \eta x, \tag{62}$$

so we get the cubic equation

$$2\eta x^3 - (2 + \eta)x + 1 = 0. \tag{63}$$

For $\Lambda = 0$ we know that $x = 1/2$ (the radius at virialization was one half the maximum radius when the background was EdS.) We only expect Λ to give a small correction to this result, so we write

$$x = \frac{1}{2} + \epsilon, \tag{64}$$

where $\epsilon \ll 1$, insert in (63) and expand to linear order in ϵ. We then find the linear equation

$$-\epsilon \left(2 - \frac{1}{2}\eta\right) = \frac{1}{4}\eta, \tag{65}$$

with the solution

$$\epsilon = -\frac{\frac{1}{4}\eta}{2 - \frac{1}{2}\eta}, \tag{66}$$

and

$$x = \frac{1}{2} + \epsilon = \frac{1 - \frac{1}{2}\eta}{2 - \frac{1}{2}\eta}, \tag{67}$$

and finally

$$\frac{R_f}{R_t} = \frac{1 - \frac{1}{2}\eta}{2 - \frac{1}{2}\eta}. \tag{68}$$

For $0 < \eta < 1$ this ratio is $< \frac{1}{2}$, which means that the perturbation collapses to a *smaller* radius when Λ is present, The reason is that spherical shells will have to fall further in to acquire the velocity which will bring the system to an equilibrium.