Application: Evolution of the Universe

Last time we discussed the formation of stars, as an example of a dynamical system. Now we'll set our sights on a larger scale: the universe itself. A full, correct treatment of the evolution of the universe requires general relativity. However, we can gain substantial insight using a Newtonian treatment of the problem. In doing so we'll largely follow Eve Ostriker's notes.

First, we'll borrow some assumptions from cosmology. We will assume that, to a first approximation, the universe is homogeneous. Perhaps this assumption seems obvious now, but if you think about it, it's ludicrous. You and I are roughly 10^{31} times denser than the average density of the universe. The Milky Way galaxy as a whole is 10^{5-6} times denser than the universe as a whole, depending on your cutoff radius for the Galaxy. **Ask class:** how can we get away with assuming homogeneity? We can do so for the whole universe because at length scales $l > 100$ Mpc or so, the universe really does appear to be distributed in a pretty homogeneous way, and it gets more so for larger scales. In addition, the cosmic microwave background provides direct evidence that at earlier times the universe was even more homogeneous. The reason that homogeneity (and isotropy) were assumed initially was that this made it easier to solve the equations. However, it is the observational confirmation that makes this acceptable. That's the difference between aesthetic or simplifying assumptions like the cosmological principle, and assumptions such as made by Ptolemy about the circularity of planetary orbits. Both are made for theoretical reasons (or reasons of preference), but as always it is the observations that decide.

Anyway, let us now consider a finite sphere carved out of the background universe, and assume that it is homogeneous and uniform. Let us also assume that the particles in this sphere are cold, i.e., they have no random motion and therefore no pressure. **Ask class:** can they think of examples of matter like that? Dark matter is hypothesized to have these properties, but a more familiar example is dust. We can then ask the Newtonian question: how does this sphere of pressureless matter evolve in time, given some initial radial expansion?

Assume that the sphere has a time-dependent radius $R(t)$, and that although at any given instant the density is uniform over the sphere, that value of the density also changes with time: $\rho = \rho(t)$. Then the sphere has a mass

$$M = \frac{4}{3}\pi\rho(t)R^3(t) = \text{const.} \tag{1}$$

Ask class: what is the gravitational acceleration at the surface? It is

$$\begin{aligned} \mathbf{g} &= -(GM/R^2(t))\hat{R} \\ \ddot{R} &= -GM/R^2\ . \end{aligned} \tag{2}$$

Ask class: what is the total energy per mass (kinetic plus potential) of a particle at the surface at any given instant? It is $E = \frac{1}{2}\dot{R}^2 - GM/R$. Since the situation is time-dependent it is not obvious whether the energy of this particle is conserved in time, so let's check:

$$dE/dt = d\left(\frac{1}{2}\dot{R}^2 - GM/R\right)/dt = \dot{R}\ddot{R} - \frac{d}{dt}\left(\frac{GM}{R}\right) = -\frac{GM}{R^2}\dot{R} + \frac{GM}{R^2}\dot{R} = 0\ . \tag{3}$$

Therefore, the energy of each particle is conserved. We can rewrite the energy in terms of density:

$$E = \frac{1}{2}\dot{R}^2 - \frac{G}{R}\frac{4}{3}\pi R^3\rho(t) = \frac{\dot{R}^2}{2}\left[1 - \rho(t)/\left(\frac{3}{8\pi G}(\dot{R}/R)^2\right)\right]\ . \tag{4}$$

Ask class: by analogy with orbits, how does the behavior of the system depend on the term in brackets? If the total energy is negative, the system is bound and will eventually recollapse. If the energy is positive, the system is unbound and will expand forever. If the energy is zero, it's on the margin but will expand forever (albeit with asymptotically zero speed). From the expression in brackets this means that there is a *critical density*

$$\rho_{\rm crit} \equiv \frac{3}{8\pi G}(\dot{R}/R)^2 \tag{5}$$

such that if $\rho > \rho_{\rm crit}$ the sphere will eventually recollapse, whereas if $\rho \leq \rho_{\rm crit}$ the sphere will expand forever. By a happy coincidence, this is also the critical density in general relativity if the universe is matter-dominated (i.e., radiation and cosmological constants or its variants are unimportant). Therefore, if you know ρ and $\dot{R}/R$ observationally, you know the evolution of the universe in this matter-dominated scenario. The ratio $H(t) = \dot{R}/R$ is called the Hubble constant. Why, you may ask, is it called a constant if it varies? It's historical: in the local universe (all that was visible in Hubble's time), $\dot{R}/R$ is the same in all directions and therefore the bulk flow of the universe is *spatially* constant. The value now is written H_0, and is often expressed as $H_0 \equiv 100h$ km s^{-1} Mpc^{-1}; note that this has units of inverse time, as it must. The value of h is roughly 0.7, from Cepheid measurements and from observations of the cosmic microwave background. This implies that

$$(\rho_{\rm crit})_0 = \frac{3}{8\pi G}H_0^2 \approx 10^{-29}\ \text{g cm}^{-3} \approx 1.4\times 10^{11}\ M_\odot\ \text{Mpc}^{-3}\ . \tag{6}$$

For a long time it was "known" that the total matter density of the universe is equal to the critical density; if it weren't, the argument went, the universe would have either collapsed very rapidly or would have expanded so much that structure wouldn't have formed (it being very unlikely that the density was not the critical density but close enough to fool us after 10^{10} years). If you solve the equations you can show that if the universe was a little way from critical at some point (say, slightly subcritical density) then after the universe had expanded by an order of magnitude or two the density would be a *lot* subcritical. Since the universe did, in fact, expand by many orders of magnitude, having $\rho \approx \rho_{\rm crit}$ now would mean that it started off at $\rho = 0.9999\ldots\rho_{\rm crit}$. Much simpler to have had it start off at (and stay at!) $\rho_{\rm crit}$ from the beginning. A convenient dimensionless parameter for the density compared to the critical density is $\Omega_m \equiv \rho/\rho_{\rm crit}$, where the m subscript means "matter".

The problem, however, is that observers stubbornly persisted in not finding enough matter to make up the critical density. The total mass of stars in the universe is less than 1% of critical. Hot gas in clusters adds maybe another 1%, probably less. There is hot gas between galaxy clusters as well, but the total mass in baryons is only 4% of critical, based on constraints from light element nucleosynthesis shortly after the big bang. Observations of rotation of galaxies, of motion of galaxies and gas in clusters, and of gravitational lensing, suggest that there is another component: dark matter, making up maybe 23% of critical. What this dark matter is, no one can be sure, but a leading candidate is the lightest supersymmetric particle. In any case, the total is *way* short of what is needed for the critical density. More on that later.

If we follow the historical path and blithely ignore the evidence, we can try to solve the equations for $E=0$, where $\rho=\rho_{\rm crit}$. This is called a "flat" universe, where we are still assuming a matter-dominated cosmology. We then have

$$\begin{aligned} \dot{R}^2/2 - GM/R &= 0 \\ \dot{R} &= (2GM/R)^{1/2} \\ R^{1/2}dR &= (2GM)^{1/2}dt \\ R &= \left(\tfrac{3}{2}\right)^{2/3}(2GM)^{1/3}t^{2/3}\ . \end{aligned} \tag{7}$$

Therefore, $R/R_0 = (t/t_0)^{2/3}$, where R_0 and t_0 are respectively the current radius and age of the sphere (or our universe). This means that density $\rho \propto M/R^3 \sim t^{-2}$, in fact $\rho(t) = 1/(6\pi G t^2)$. We also see that the rate of expansion decreases in time, $\dot{R} \propto t^{-1/3}$. For a flat universe,

$$t_0 = \frac{1}{(6\pi G\rho_0)^{1/2}} = \frac{2}{3H_0}\ . \tag{8}$$

That's about 10^{10} yr, in rough agreement with constraints based on the oldest stars and other measures of the age of the universe.

Why would we focus attention on flat universes? For one thing, it's easier to treat than "open" ($E>0$) or "closed" ($E<0$) universes. However, there's a better reason. Rewriting our equation for the total energy, we find

$$\dot{R}^2 = 2E + 2GM/R\ . \tag{9}$$

If the universe expanded from a very small radius (as indicated by the cosmic microwave background and big bang nucleosynthesis), then at some point $|2GM/R| \gg |2E|$, since E is constant. Therefore, at early times the evolution of the universe was essentially governed by $\dot{R}^2 \approx 2GM/R$, i.e., it acted as if $E\approx 0$. This means that the flat universe case actually has been relevant for much of the history of the universe, regardless of the current value of E.

In full general relativity, when one has matter with a relativistic energy density u and pressure p, and a "cosmological constant" Λ, the acceleration equation becomes

$$\frac{\ddot{R}}{R} = -\frac{4}{3}\frac{\pi G}{c^2}(u + 3p) + \Lambda/3\ . \tag{10}$$

How does this change things? First, consider matter made up of particles with nonzero rest mass m. Assume that the temperature is low. **Ask class:** what is the approximate relativistic energy of a single particle? If the temperature is low, the energy is primarily the rest mass energy, so $E \to mc^2$. Then, the energy density becomes $u \approx c^2\rho$. The next thing we need is a comparison between the pressure and the energy density. **Ask class:** how can we do that? We can assume for starters that the particles form an ideal gas, so that $p = nkT = (\rho/m)kT$. Then

$$\frac{1}{c^2}(u + 3p) = \rho(1 + 3kT/mc^2)\ . \tag{11}$$

Recalling that the thermal speed is $v_{\rm th}^2 = 3kT/m$, this becomes

$$\frac{1}{c^2}(u + 3p) = \rho(1 + v_{\rm th}^2/c^2) \approx \rho\ . \tag{12}$$

Therefore, for slowly moving matter, energy density is much more important than pressure, and we get

$$\frac{\ddot{R}}{R} \approx -\frac{4}{3}\pi G\rho + \Lambda/3\ , \tag{13}$$

which is the same as the Newtonian equation except for the cosmological constant term. This first term therefore scales as R^{-3}, since the mass is constant and $\rho \sim M/R^3$.

If you consider radiation instead of cold (nonrelativistic) matter, you find eventually that instead

$$\frac{\ddot{R}}{R} = -\frac{8\pi G}{3}\frac{u_0}{c^2}(R/R_0)^{-4} + \Lambda/3\ , \tag{14}$$

where u_0 is the energy density of radiation at a time when the radius of the universe is R_0.

Now, in the simplest formulation, Λ is a true constant, independent of time or of the current size of the universe. With that in mind, and the dependences on radius of the radiation and cold matter, **Ask class:** what can one say about which term (matter, radiation, or cosmological constant) is most important at different epochs of the universe? The radiation term scales as R^{-4}, the matter term scales as R^{-3}, and the cosmological constant term scales as R^0. Therefore, at a small enough radius the radiation must be most important. If $\Lambda \neq 0$, then at a large enough radius, the cosmological constant term must be the most important. Those are the only things we can be certain of. It happens that in the actual universe, the matter term is most important for a range of redshifts, but it could have been differently (as far as we know). Recent data suggests that if you add up the contributions from all forms of matter and cosmological constant, you get $\Omega_{\rm total} = \Omega_m + \Omega_{\rm rad} + \Omega_\Lambda + \Omega_{\rm neutrino} + \ldots \approx 1$. This ends up meaning that in the general relativistic sense, the universe as a whole is close to flat. However, it also means that $\Omega_\Lambda \approx 0.73$, which is a bit of a surprise. People are still trying to understand the ramifications of this, and the nature of this so-called "dark energy".

As a final point, let's revisit the nature of gravitational instability, this time against a background of expanding spacetime. The current, quite successful, picture is that some process in the extremely early universe produced fluctuations in the density of matter; it's not exactly the same everywhere, although it was close early on. What happens when gravity does its thing? If we were in a static background, then for density ρ the timescale for growth of a density enhancement is on the order of $1/\sqrt{4\pi G\rho}$, where here we are putting in the 4π to be a little more precise. One would therefore expect that the density would grow as

$$\exp\left(\int^t \sqrt{4\pi G\rho}dt\right) \ , \tag{15}$$

that is, exponentially. What happens when we add an expanding universe? Then, as we showed for a matter-dominated phase (the relevant one for the growth of structure), the density changes like $\rho(t) = 1/(6\pi G t^2)$. Therefore, the growth becomes

$$\exp\left(\int^t \sqrt{4\pi G\rho(t)}dt\right) = \exp\left(\sqrt{\frac{2}{3}}\int\frac{dt}{t}\right) \approx t^{\sqrt{2/3}} \ , \tag{16}$$

that is, like a power law. Done correctly in general relativity, the power law is actually $t^{2/3}$ instead of $t^{\sqrt{2/3}}$, so if the density is a small perturbation from the average

$$\rho = \rho_{\rm crit}(1+\delta) \ , \tag{17}$$

then $\delta \propto R$. This slow growth means that it is challenging to grow galaxies! When the density enhancement becomes large ($\delta \sim 1$), then gravity is enough to overcome the expansion and start exponential collapse. Sometimes people talk about this as the matter "decoupling" from the background expansion; of course, the expansion is still going on, but it is a tiny perturbation. The solar system hasn't expanded by a factor of two in the last 5 billion years!