MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe September 22, 2018

Prof. Alan Guth

Lecture Notes 3

THE DYNAMICS OF NEWTONIAN COSMOLOGY

INTRODUCTION:

The dynamics of the universe on the large scale seems to be controlled by gravity, and so any theoretical work in cosmology rests heavily on the foundations of a theory of gravity. Among professionals, the dynamics of cosmology is always treated in the context of the relativistic theory of gravity developed by Einstein in 1915 — the theory which is known by the misleading name of "General Relativity". We believe that, at the classical level, general relativity is almost certainly the correct theory of gravity. (At extraordinarily high energies, like those encountered during the first 10^{-45} second after the big bang, a quantum theory of gravity would be required. At present, however, gravity at the quantum level is not well-understood. Many physicists believe that string theory is likely to be the correct quantum theory of gravity, but there are many questions about string theory that have not yet been answered.) In cosmology, general relativity is necessary to make sure that the possibly non-Euclidean geometry of the universe is being treated correctly. General relativity is also required to give an accurate treatment of the gravitational effect of electromagnetic radiation (e.g., light), which is significant in the early universe and which is certainly a relativistic phenomenon. However, a good deal of cosmology can be understood strictly in terms of Newtonian gravity, and in these notes we will explore cosmology in that context. Even the gravitational effects of electromagnetic radiation can be inferred correctly by using Newtonian physics combined with some well-motivated guesses.

The universe is believed to be homogeneous, so the key problem is to understand the gravitational dynamics of a homogeneous distribution of mass. We will consider a distribution of mass with infinite extent, with a uniform mass density ρ.

This is a subtle problem, and in fact Isaac Newton himself got it wrong. Newton assumed that since the mass distribution is symmetric about any point, the gravitational field at any point must vanish, since there is no preferred direction in which it could point. He therefore believed that a static configuration of "fixed stars" could exist in equilibrium. Newton discussed this issue in a series of letters he wrote to the young theologian, Richard Bentley, during 1692-93*

---

* The original letters are still kept at Trinity College, Cambridge, and are published in H. W. Turnbull, ed., *The Correspondence of Isaac Newton, Volume III, 1688-1694* (Cambridge University Press, Cambridge, England, 1961, p. 233). They are also reprinted

> *"As to your first query, it seems to me that if the matter of our sun and planets and all the matter of the universe were evenly scattered throughout all the heavens, and every particle had an innate gravity toward all the rest, and the whole space throughout which this matter was scattered was but finite, the matter on the outside of this space would, by its gravity, tend toward all the matter on the inside and, by consequence, fall down into the middle of the whole space and there compose one great spherical mass. But if the matter was evenly disposed throughout an infinite space, it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered at great distances from one to another throughout all that infinite space. And thus might the sun and fixed stars be formed, supposing the matter were of a lucid nature.*† *" (December 10, 1692)*

The point of view that Newton described in his response to Bentley apparently represents a departure from his earlier reasoning. Previously Newton had believed that the fixed stars occupied a finite region in an infinite void, but now he realized that such a configuration would be driven by gravity to collapse. If the stars were distributed uniformly over the infinity of space, however, Newton concluded that static equilibrium could be maintained.*

---

in Milton K. Munitz, ed., *Theories of the Universe: From Babylonian Myth to Modern Science* (The Free Press, New York, 1957, p. 211). Best of all, thanks to Google Books and the Newton Project, the complete letters from Newton to Bentley are now available online: http://books.google.com/books?id=8DkCAAAAQAAJ&pg=PA201 and http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00254, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00255, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00256, and http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00258.

† By "lucid nature," Newton was apparently referring to the distinction that he supposed exists between the "lucid matter" of the sun and stars, and the "opaque" matter of the earth and other planets. The continuation of the text shows Newton's thoughts on this issue, and also on the role of divine intervention in the creation of the solar system: *"But how the matter should divide itself into two sorts, and that part of it which is to compose a shining body should fall down into one mass and make a sun and the rest which is fit to compose an opaque body should coalesce, not into one great body, like the shining matter, but into many little ones; or if the sun at first were an opaque body like the planets or the planets lucid bodies like the sun, how he alone should be changed into a shining body whilst all they continue opaque, or all they be changed into opaque ones whilst he remains unchanged, I do not think explicable by mere natural causes, but am forced to ascribe it to the counsel and contrivance of a voluntary Agent."*

* Newton's involvement in this problem was discussed in a fascinating article by Edward Harrison, "Newton and the Infinite Universe," *Physics Today*, February 1986, p. 24, which is available online with an MIT certificate at http://scitation.aip.org.

The fallacy of Newton's argument was not really understood until the beginning of the 20th century. When Einstein first developed his theory of general relativity, he very quickly tried to apply it to the universe as a whole, and at first he was rather shocked to learn that the theory did not allow a static solution. According to the mathematics of the theory, an initially static configuration would lead to a universal collapse, as each particle of matter in the universe attracted all of the others. Einstein chose to modify general relativity by adding a "cosmological term" — a kind of universal repulsion — so that a static solution would be possible. In hindsight, one can see that the same reasons which preclude a static solution in the theory of general relativity (without a cosmological constant) apply also to the Newtonian case.

The nonexistence of a static equilibrium for an infinite homogeneous distribution of mass can be seen very easily by using some mathematics that was unavailable to Newton. Newton formulated his law of universal gravitation in the language of an inverse square force law, but we now know how to reformulate such a law in terms of flux integrals. Just as Coulomb's law implies Gauss's law, Newton's inverse square law of gravity gives rise to a Gauss's law of gravity:

$$\vec{E} = \frac{q}{r^2}\hat{r} \quad \text{implies} \quad \oint \vec{E} \cdot d\vec{a} = 4\pi q_{\text{enclosed}} \tag{3.1}$$

$$\vec{g} = -\frac{GM}{r^2}\hat{r} \quad \text{implies} \quad \oint \vec{g} \cdot d\vec{a} = -4\pi G M_{\text{enclosed}} \ , \tag{3.2}$$

where $\vec{g}$ is the gravitational acceleration vector, and the integrals are over an arbitrary "Gaussian" surface. If Eq. (3.2) is applied to a uniform distribution of mass, then clearly $M_{\text{enclosed}} > 0$ for any Gaussian surface that encloses a nonzero volume. Thus the left hand side must also be nonzero, and so one cannot have $\vec{g} = 0$, as a static universe would demand.

Another formulation of Newtonian gravity takes the form of a gravitational Poisson's equation:

$$\nabla^2 \phi = 4\pi G \rho \ , \ \text{where } \vec{g} = -\vec{\nabla}\phi \ , \tag{3.3}$$

and ρ is the mass density. Here ∇^2 is the Laplacian,

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \ ,$$

---

libproxy.mit.edu/content/aip/magazine/physicstoday/article/39/2/10.1063/1.881049 or for purchase at http://scitation.aip.org/content/aip/magazine/physicstoday/article/39/2/10.1063/1.881049.

and $\vec{\nabla}$ is the gradient,

$$\vec{\nabla} \equiv \hat{\imath}\,\frac{\partial}{\partial x} + \hat{\jmath}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}\ .$$

In this formalism one can also see that $\vec{g} = 0$ implies $\phi = \textit{constant}$, which in turn implies $\rho = 0$, so a static universe is possible only if it is empty.

Historically, I believe that the inconsistency of the static universe was overlooked in the context of Newtonian mechanics because Newtonian gravity is usually described in terms of an action at a distance. In this formulation, the relevant issues are subtle. General relativity, on the other hand, is always formulated in terms of local differential equations analogous to Eq. (3.3), and in this formulation the result is unmistakable.

I have now discussed the reasons why a homogeneous mass distribution must produce a gravitational field, but I have not yet discussed what goes wrong if one tries to calculate the force on a given particle by summing the Newtonian gravitational forces caused by all the other particles. Since these other particles extend with uniform density to infinity in all directions, it seems obvious that the integration over the mass distribution cannot pick out any preferred direction, and therefore must give no gravitational force. The problem with using this method, however, is that the integration is ambiguous. We will show that, due to the poor convergence properties of the integral, the integration has no unique answer, but instead can give any answer that one wants, depending on the order with which the different regions of the integration volume are included.

To see how this can happen, let us first consider some general properties of integrals. Suppose that $f(x)$ is a function such that

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \tag{3.4}$$

converges, in the sense that

$$\lim_{L\to\infty} \int_{-L}^{L} f(x)\,\mathrm{d}x \tag{3.5}$$

exists. Suppose, however, that

$$\int_{-\infty}^{\infty} |f(x)|\;\mathrm{d}x \tag{3.6}$$

diverges (i.e., is infinite). Such integrals are called *conditionally convergent*, and in general their value is ambiguous. The answer depends on the order in which the different regions of the x-integration are added up. Conversely, if the integral (3.6) converges, then the integral (3.4) is called absolutely convergent, and its value is independent of the order in which the different regions of integration are added.

As a simple example, consider the function

$$f(x) = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \;, \end{cases} \tag{3.7}$$

the integral of which clearly satisfies the properties of conditional convergence as described above. To illustrate the ambiguity of the integral, note first that

$$\lim_{L\to\infty} \int_{-L}^{L} f(x)dx = 0. \tag{3.8}$$

This limit is trivial, since the integral is zero for any value of L. Now let's add the contributions in a different order, starting at some arbitrary point $x = a$. We take the region of integration larger and larger, but always centered on $x = a$. That is, we can define the integral

$$\lim_{L\to\infty} \int_{a-L}^{a+L} f(x)dx \;. \tag{3.9}$$

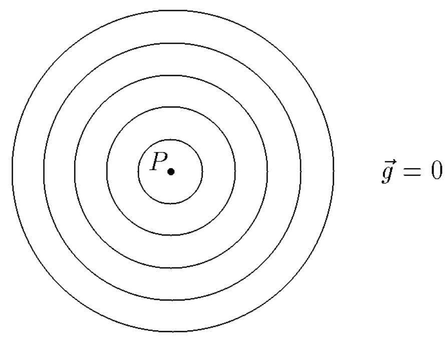
You should be able to convince yourself that the integral is equal to $2a$ for any $L \geq a$, and therefore the limit is $2a$. Since we can choose a to be anything we like, we can get any answer that we like. Note that the integrals shown as (3.5) and (3.9) are both ways of giving precise meaning to the integral (3.4), so one concludes that the integral (3.4) is ambiguous. Mathematically one can (and usually does) ***define*** the integral (3.4) to be the expression (3.5), but one must keep in mind that this is an arbitrary choice that is unlikely to have physical meaning. When x represents a spatial coordinate, as it does here, then the expressions (3.5) and (3.9) differ only by the choice of where the origin of the coordinate system is placed, while physically this choice is completely arbitrary.

For an infinite distribution of mass with uniform density ρ, the gravitational acceleration at a point P is given formally by the integral

$$\vec{g}(P) = \int G\rho\, d^3\vec{r}\,' \, \frac{\vec{r}\,' - \vec{r}_P}{|\vec{r}\,' - \vec{r}_P|^3} \;, \tag{3.10}$$

where $\vec{r}_P$ is a vector from the origin to the point P. We will see that this integral is conditionally convergent, and therefore $\vec{g}(P)$ can have any value, depending on the order in which the contributions from different values of $\vec{r}\,'$ are added. Newton's law of gravity says nothing about the order in which the contributions should be added, since in normal situations vector addition is commutative.

To see how this integral behaves, suppose we first determine the value of $\vec{g}$ at an arbitrary point P by summing the contributions from spherical shells that are centered at P:



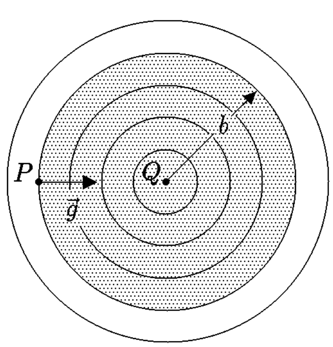
In this case one can argue by symmetry that each shell contributes exactly zero to $\vec{g}$, and hence the sum must be zero.

The integral clearly converged (in fact it vanished!), but it is not absolutely convergent. If we inserted absolute value signs in the integrand, we can evaluate the integral by transforming to polar coordinates with P at the origin. We find a linear divergence:

$$\int G\rho\, d^3\vec{r}\,' \frac{1}{\left|\vec{r}\,' - \vec{r}_P\right|^2} = 4\pi G\rho \int r'^2\, dr' \frac{1}{r'^2} = \infty\ .$$

Thus the integral is convergent but not absolutely convergent, so it is conditionally convergent.

To see the ambiguity that we expect due to the conditional convergence, we need to carry out the integration of Eq. (3.10) with a different ordering. Spherical shells are still very convenient, but suppose we choose spherical shells centered around a different origin. To see what we find, let us calculate $\vec{g}$ at P by summing the contributions from spherical shells which are centered at some other point Q, located a distance b away:



The gravitational field due to a thin spherical shell of mass is well known — inside the shell the field vanishes identically, and outside the shell the field is the same as it would be if the same mass were concentrated at the point in the center of the sphere. For the shells centered at Q, note that the point P will lie inside the shell for all shells with radius $r > b$. These shells will therefore give no contribution to the gravitational field at P. The shells with $r < b$, on the other hand, which are shown with shading in the diagram above, will produce a gravitational field at P. Specifically, they will produce a field at P which is the same as the field that would be produced if the entire mass (for $r < b$) were concentrated at Q. Thus, by this method of summation we find

$$\vec{g} = \frac{GM}{b^2}\hat{e}_{QP} \ , \tag{3.11}$$

where $M = \frac{4\pi}{3}b^3\rho$ is the combined mass of all the shells with $r < b$, and $\hat{e}_{QP}$ is a unit vector pointing from P to Q. So the answer we get depends on the order of summation. Since we could have chosen the point Q to be any distance and in any direction, we could have gotten any answer we wanted.

Thus we can conclude that the integral which determines $\vec{g}$ is ill-defined. By summing the gravitational force using concentric spherical shells centered at different points Q, we can get any answer we want. But what about the simple symmetry argument, which says that the gravitational force must be zero because there is no preferred direction for it to point? It was this argument that Newton found persuasive in the letter to Bentley cited earlier. Newton might phrase the reasoning in the following way: If there is to be a force on the mass located at P, then the force would have to point in some direction. But since all directions are identical in this problem, the force must vanish. To convince Newton that he was wrong, we would have to persuade him that this problem is very special, because there is no way to define an inertial reference frame. Ordinarily one can define an inertial frame by imagining test particles at infinite distances from all others — the inertial frames are those in which these test particles have constant velocities. In the problem of an infinite uniform mass distribution, however, there is no place to put these test particles. Thus, one cannot measure the absolute acceleration of any particle, but instead one must settle for measuring the **relative** acceleration of one particle with respect to another. One can decide, for example, to measure the accelerations of all particles relative to P. One then finds, as we will see later, that all the accelerations point toward P, and that the acceleration of any given particle has a magnitude which is proportional to the distance from P. If one had chosen to measure all accelerations relative to Q, one would have found a similar pattern centered on Q.

THE MATHEMATICAL MODEL:

The approach that I will follow here is a bit more involved than that used in most textbooks, but it also leads to a stronger result. Most textbooks simply assume that a uniform distribution of mass will remain uniform, but here we will show that the inverse square law of gravity leads to this result. Most other force laws would not.

In order to make the problem of an infinite uniform mass distribution well defined, it is necessary to treat the concept of infinity carefully. Specifically, the safest way to think about infinity is to think of it as a **limit** of finite quantities. The easiest approach, which we will use, is to treat the mass distribution as a uniform sphere of radius R_{max}. Only at the end of the calculation will we take the limit $R_{\text{max}} \to \infty$.

When we choose to use a sphere of mass to define our problem, it is important to ask if the answer would have turned out differently if we had chosen some other shape. I will not try to demonstrate the answer to this question, but I will tell you what it is. Many shapes, such as any of the regular polyhedra (tetrahedron, cube, octahedron, dodecahedron, or icosahedron), would give the same answer in the limit in which the size approaches infinity. If we had used a rectangular solid, however, the answer would have been different. [For those students who have learned about multipole expansions, I mention that it is only the quadrupole moment of the shape that matters in the limit of infinite size.] The solution obtained from the rectangular solid would correspond to an anisotropic model of the universe, in which the gravitational field would be different in different directions. General relativity also allows for the possibility of anisotropic homogeneous solutions, but I have never explored how closely the properties agree. Since our universe is highly isotropic, we are justified in using the sphere to formulate our problem.

We will treat the matter as a nonrelativistic dust of particles which can move freely, with gravity supplying the only significant force. The assumption that the universe is dominated by nonrelativistic matter and that gravity is the only significant force appear to be valid assumptions for our own universe for most of its history, but not for all of it. Recall that in the context of relativity, energy and mass are really the same thing, related by the celebrated formula $E = mc^2$, where c is the speed of light. In the early universe there was a high density of energy in electromagnetic radiation, and this energy density can be expressed as a mass density by dividing it by c^2. For the first approximately 50,000 years of cosmic history, the mass density of the universe was dominated by the electromagnetic radiation and highly relativistic particles, both of which lead to significant pressure forces. These pressure forces, in turn, lead to a contribution to the gravitational force, since general relativity implies that pressures as well as mass or energy densities can serve as the source of a gravitational field. Cosmologists call this early period "radiation-dominated", and the period in which the universe is dominated by nonrelativistic dust is called "matter-dominated". Starting in about 1998, astronomers have been gathering evidence that for the past 5 billion years or so the expansion of the universe has not been slowing as it would in a matter-dominated universe, but instead it has been accelerating. These observations were a big surprise to most of us, and they suggest that the universe today is dominated by a nonzero energy density in the vacuum — which is equivalent to what Einstein called the cosmological constant — or some form of peculiar matter that behaves very similarly. The term "dark energy" has been coined to describe this form of energy, which remains rather mysterious as the name suggests.

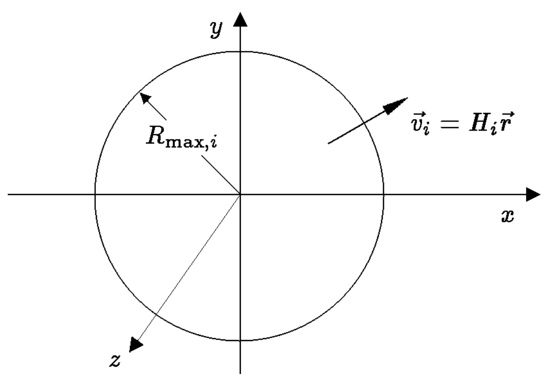
So, it now appears that the universe was radiation-dominated for the first 50,000 years, then became matter-dominated for about 9 billion years, and then "recently" became dark-energy-dominated, for about the last 5 billion years. Later on we will see how the pressure forces of the radiation-dominated period can be incorporated into the model, and we will also learn how to calculate the effects of vacuum energy. For now, however, we will confine our attention to the matter-dominated era.

We will begin the mathematical model of our idealized universe at some arbitrary initial time t_i. At that time we will assume that the universe consists of a sphere of matter with radius $R_{\mathrm{max},i}$, with uniform mass density ρ_i. For convenience we will introduce an x-y-z coordinate system, with the origin located at the center of the sphere. We will treat the matter as a nonrelativistic dust of particles which can move freely, with gravity supplying the only significant force.

The next step is to specify the initial velocity of each of the particles. In order to agree with the observed properties of the universe, we choose this initial velocity distribution according to Hubble's law: the particles at position $\vec{r}$ are given an initial velocity of the form

$$\vec{v}_i = H_i \vec{r} \;, \tag{3.12}$$

where H_i denotes the initial value of the Hubble "constant". Thus, the initial state of the model is described by the parameters ρ_i, H_i, and $R_{\mathrm{max},i}$.



The problem now is to calculate the evolution of this model, using Newton's law of gravity. Since each particle is started along a radial trajectory, and since the only forces will be radial, it follows that each particle continues to move along a radial trajectory. Thus we need only keep track of the radius of each particle as a function of time. We will follow an arbitrary particle with initial radius r_i, and we will denote its trajectory by $r(r_i, t)$. To compute the force on this particle due to all the other particles in the model universe, we can divide the mass distribution into thin spherical shells — with each shell centered on the origin and extending from some radius r to $r + dr$. We then use the result quoted earlier for the gravitational field of a thin spherical shell. One concludes that all shells with $r < r_i$ will produce a gravitational field at r_i equivalent to that of a point mass at the origin, while all shells with $r > r_i$ will contribute nothing at all to the gravitational field at r_i. The mass of all the shells with $r < r_i$ is given by

$$M(r_i) = \frac{4\pi}{3} r_i^3 \rho_i \ . \tag{3.13}$$

It is conceivable that at some point in the evolution of the system there could be a crossing of shells — that is, two trajectories $r(r_i, t)$ corresponding to two different values of r_i could cross. However, since initially the Hubble expansion is carrying each shell away from its neighbors, it is clear that a shell crossing will not happen until some nonzero time interval has elapsed. (We will in fact find that shell crossings never occur, but we have no way of knowing this before we start.) As long as no shell crossings have occurred, the mass interior to the shell which began at radius r_i is always equal to the expression for $M(r_i)$ given in Eq. (3.13), since mass is conserved. The gravitational acceleration acting at an arbitrary time t on a particle with initial radius r_i is then given by

$$\vec{g} = -\frac{GM(r_i)}{r^2(r_i, t)} \hat{r} \ , \tag{3.14}$$

where $\hat{r}$ denotes a unit vector in the radial direction. Taking the radial component of this vector equation and using Eq. (3.13), one has

$$\ddot{r} = -\frac{4\pi}{3} \frac{G r_i^3 \rho_i}{r^2} \ , \tag{3.15}$$

where $r \equiv r(r_i, t)$, and an overdot denotes a derivative with respect to t. The initial condition on the velocity given in Eq. (3.12) can be rewritten in this notation as

$$\dot{r}(t{=}t_i) = H_i r_i \ . \tag{3.16}$$

Finally, the initial value of $r(r_i, t)$ is given by

$$r(r_i, t_i) = r_i \ . \tag{3.17}$$

The mathematical problem is then to solve Eq. (3.15), subject to the initial conditions of Eqs. (3.16) and (3.17).

First, note that the dependence on r_i in these equations can be eliminated by a simple rescaling of the as yet unknown function $r(r_i, t)$. That is, define

$$u(r_i, t) \equiv r(r_i, t)/r_i \ . \tag{3.18}$$

Note that r_i does not depend on t, and it can therefore be treated as a constant as far as time derivatives are concerned. Eqs. (3.15)-(3.17) can then be rewritten as

$$\ddot{u} = -\frac{4\pi}{3}\frac{G\rho_i}{u^2} \ , \tag{3.19}$$

$$\dot{u}(t{=}t_i) = H_i \ , \tag{3.20}$$

$$u(r_i, t_i) = 1 \ . \tag{3.21}$$

These equations specify the initial value and time derivative of u, and its acceleration at all times, and they therefore completely determine the function. Since these equations do not involve r_i, it follows that $u(r_i, t)$ does not actually depend on r_i at all. This means that $u(r_i, t)$ is really just an overall scale factor, and we can define

$$a(t) \equiv u(r_i, t) \ . \tag{3.22}$$

Eq. (3.18) then becomes $r(r_i, t) = a(t)r_i$, which means that the particle locations at any given time t are given by a rescaling of their original positions, by the scale factor $a(t)$. Note also that the mean mass density inside a sphere of radius $r(r_i$,t) is given by

$$\rho(t) = \frac{M(r_i)}{\frac{4\pi}{3}r^3} = \frac{\frac{4\pi}{3}r_i^3\rho_i}{\frac{4\pi}{3}r^3} = \frac{\rho_i}{a^3(t)} \ , \tag{3.23}$$

and is also independent of r_i. The mass density thus remains completely uniform. Using Eqs. (3.22) and (3.23), Eq. (3.19) can be rewritten as

$$\boxed{\ddot{a} = -\frac{4\pi}{3}G\rho(t)a \ .} \tag{3.24}$$

Eq. (3.24) describes how the expansion of the scale factor is slowed down by the gravitational effects of the mass density $\rho(t)$.

We can now return to the issue of shell crossing, and see that it never occurs. From Eqs. (3.18) and (3.22) we know that $r(r_i, t) = a(t)r_i$, as long as our equations are valid. Thus, if the first shell crossing occurs at time t_{shell}, then the relation $r(r_i, t) = a(t)r_i$ must hold for all t between t_i and t_{shell}. But if $r(r_i, t) = a(t)r_i$ holds at time $t_{\text{shell}} - \epsilon$, for arbitrarily small $\epsilon > 0$, then there can be no shell crossing at $t = t_{\text{shell}}$, since $r(r_i, t) = a(t)r_i$ implies that no two shells with different values of r_i are about to touch.

The limit $R_{\rm max} \to \infty$ is now seen to be trivial. As discussed earlier in the section on "The Homogeneously Expanding Universe" in Lecture Notes 2, this kind of uniform expansion by an overall scale factor $a(t)$ appears to be absolutely homogeneous to the inhabitants of this idealized universe. Looking from the outside we see a sphere with a center and an edge, but someone living anywhere inside the sphere would simply see all of his neighbors receding in a Hubble pattern, with the Hubble expansion rate given by

$$\boxed{H(t) = \dot{a}/a \ .} \tag{3.25}$$

Only someone living so near to the edge that he could actually see it would have any way of knowing that the system was not globally homogeneous. Thus, the limit $R_{\rm max} \to \infty$ which we need to take is trivial. In fact, for observers on the interior of the sphere, nothing whatever depends on $R_{\rm max}$.

A CONSERVATION OF ENERGY EQUATION:

The equations of the last section completely determine the behavior of the model universe, so our only remaining task is to examine the consequences of these equations.

As with most Newtonian systems, conservation of energy is a useful concept. Conservation of energy is of course not an independent statement, but instead follows as a consequence of the Newtonian equations of motion. In this case Eq. (3.19) can easily be used to obtain such an equation. [Eq. (3.24), which gives the deceleration in terms of the mass density ρ, is more useful for most purposes. But it cannot be used by itself to give a conservation of energy equation, since the time dependence is not determined until one adds information about the time dependence of $\rho(t)$. One can of course combine Eq. (3.24) with Eq. (3.23) describing the evolution of $\rho(t)$, but this is equivalent to using Eq. (3.19).] The conservation equation is obtained from Eq. (3.19) by first replacing u by a, then bringing both terms to one side, and then multiplying by $\dot{a}$:

$$\dot{a}\left\{\ddot{a} + \frac{4\pi}{3}\frac{G\rho_i}{a^2}\right\} = 0 \ .$$

Using elementary calculus, the result can be rewritten as

$$\frac{dE}{dt} = 0 \ , \tag{3.26}$$

where

$$E = \frac{1}{2}\dot{a}^2 - \frac{4\pi}{3}\frac{G\rho_i}{a} \ . \tag{3.27}$$

E is not exactly an energy, and does not even have the right units to be an energy. However, if one considers a test particle of mass m that moves with the Hubble expansion starting at radius r_i, then the quantity $E_{\rm phys} \equiv m r_i^2 E$ is closely related to the energy of that particle. Specifically,

$$\begin{aligned} E_{\rm phys} &= m r_i^2 \left\{ \frac{1}{2} \dot{a}^2 - \frac{4\pi}{3} \frac{G\rho_i}{a} \right\} \\ &= \frac{1}{2} m \left(\dot{a} r_i \right)^2 - \frac{G m M(r_i)}{a r_i} \ , \end{aligned}$$

where $M(r_i)$ is given by Eq. (3.13). Then, recognizing that $a(t) r_i$ is the radius r of the test particle at time t, we can rewrite $E_{\rm phys}$ as

$$E_{\rm phys} = \frac{1}{2} m \left(\frac{{\rm d}r}{{\rm d}t} \right)^2 - \frac{G m M(r_i)}{r} \ . \tag{3.28}$$

This expression is the total energy of a particle of mass m moving radially in the gravitational field of a point particle of mass $M(r_i)$ located at the origin, where we have defined the zero of potential energy at infinity. If the test particle is at the edge of the sphere, with $r = R_{\rm max}(t)$, then this is the correct expression for the total energy of the test particle, since for all $r \geq R_{\rm max}(t)$, the gravitational effects of the sphere and a point mass are identical. If the test particle is at a smaller radius, however, then $E_{\rm phys}$ is still conserved, but it is not really the total energy. The mass that is located between r and $R_{\rm max}(t)$ would affect the amount of energy needed to bring the test particle from infinity, and hence would affect the potential energy of the test particle, but the effect of this mass is not included in $E_{\rm phys}$. Nonetheless, the mass that is located between r and $R_{\rm max}(t)$ does not affect the motion of the test particle, so we can define an analogue problem in which this mass is absent. That is, we can define an analogue problem in which $R_{{\rm max},i}$ is chosen so that the test particle is on the edge. For the analogue problem, $E_{\rm phys}$ is truly the total energy of the test particle. The motion of the test particle is the same for the analogue problem and the original problem, so we can understand the conservation of $E_{\rm phys}$ as a consequence of energy conservation for the analogue problem. It can also be shown, and you will have to opportunity to show on Problem Set 3, that E is proportional to the total enery, kinetic plus potential, of the entire sphere.

Using (3.23) to express ρ_i in terms of $\rho(t)$, Eq. (3.27) can be converted to the form

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi}{3} G\rho + \frac{2E}{a^2} \ . \tag{3.29}$$

It is more or less standard notation to introduce the variable k, defined by

$$k = -\frac{2E}{c^2} \ , \tag{3.30}$$

and then to rewrite Eq. (3.29) as

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2}\ .} \tag{3.31}$$

Eqs. (3.24) and (3.31) are the two key results of Newtonian cosmology. As long as the mass density is dominated by nonrelativistic matter, as has been the case for most of the history of our universe, these equations are both identical to the corresponding equations obtained from general relativity. They are called the Friedmann equations, after Alexander Friedmann, the Russian meteorologist who first derived the equations from general relativity in 1922. Some authors, however, including Barbara Ryden, use the term Friedmann equation only for Eq. (3.31).

UNITS:

In Lecture Notes 2 I talked about a comoving coordinate system, with coordinates measured in "notches". The scale factor $a(t)$ is then measured in m/notch. The concept of a notch is actually not used, to my knowledge, in any of the standard cosmology texts, but nonetheless I find it a very useful way of thinking — it helps to clarify what exactly the scale factor is, and when it is needed in an equation.

In the last two sections the concept of a notch did not appear, so now I would like to reinstate it. As written, one would infer that the quantity r_i, denoting the radial coordinate of a given particle at time t_i, is to be measured in meters. Note, however, that we used the coordinate r_i not merely to describe the position of the particle at t_i, but also as a permanent label of the trajectory $r(r_i, t)$. The coordinates r_i are in fact being used as comoving coordinates, and only at the special time t_i does the unit of these comoving coordinates correspond to the meter. It thus makes sense to rename the unit of r_i as a notch. The time t_i is then the time at which 1 notch corresponds to 1 m. The trajectory function $r(r_i, t)$ continues to be measured in meters, so by Eqs. (3.18) and (3.22), the scale factor $a(t)$ has the units of m/notch. The variable k then has the units of notch^{-2}.

Note, by the way, that we have still not defined the notch, since the time t_i is completely arbitrary. There are two common conventions. Some books, such as the text by Barbara Ryden, define $a = 1$ m/notch today. Many other books, however, adopt the convention that whenever $k \neq 0$, one defines the notch such that k has the numerical value of ± 1. These books tend to leave the notch arbitrary when discussing the $k = 0$ case.

NATURE OF THE SOLUTIONS:

The equations of Newtonian cosmology have now been written down, and our only remaining task is to examine the behavior of the solutions.

The solutions belong to three different classes, depending on whether E is positive, negative, or zero. The qualitative behavior can be seen most clearly from Eq. (3.27),

$$E = \frac{1}{2}\dot{a}^2 - \frac{4\pi}{3}\frac{G\rho_i}{a} \ .$$

If E is positive ($k < 0$), then one sees that da/dt can never vanish, since it gives the only positive contribution to the right-hand side. Thus an expanding universe with $k < 0$ would continue to expand forever. In a universe of this type there is not enough mass to reverse the expansion of the Hubble flow. Such a universe is called open. On the other hand, if E is negative ($k > 0$), then one sees that da/dt equals zero when

$$a = -\frac{4\pi G\rho_i}{3E} \ . \tag{3.32}$$

This universe reaches a maximum size, and then the pull of gravity overcomes the expansion and causes the universe to collapse into what is sometimes called a “big crunch”. A universe of this type is called closed. On the border between these two possibilities is the special case of $E = 0$ ($k = 0$). For reasons that will be discussed in Lecture Notes 5, such a universe is called flat.

The case $k = 0$ implies that the mass density ρ must have a special value ρ_c, which can be found from Eq. (3.31) (remembering that $\dot{a}/a = H$):

$$\boxed{\rho_c = \frac{3H^2}{8\pi G} \ .} \tag{3.33}$$

The quantity ρ_c is called the critical mass density — it is that mass density which puts the universe right on the borderline between eternal expansion and eventual collapse. Numerically, if one takes $H_0 = 100\,h_0$ km-s^{-1}-Mpc^{-1} (as in Eq. (3.3)) and $G = 6.674 \times 10^{-11}$ m^{3}kg^{-1}s^{-2}, one finds that

$$\rho_c = 1.88\,h_0^2 \times 10^{-26}\ \text{kg/m}^3 = 1.88\,h_0^2 \times 10^{-29}\ \text{g/cm}^3 \ , \tag{3.34}$$

where g is the abbreviation for gram. If $h_0 = 0.677$, which is the estimate from Planck 2018,* currently the best estimate, then $\rho_c = 8.6 \times 10^{-27}$ kg/m^3. The proton mass is

---

* N. Aghanim et al. (Planck Collaboration), “Planck 2018 results, VI: Cosmological parameters,” Table 2, Column 6, arXiv:1807.06209.

1.67×10^{-27} kg, which means that the critical mass density corresponds to about 5.2 protons per cubic meter. Although this mass density seems amazingly small, the actual mass density of the universe is known to be very close to it.

The ratio ρ/ρ_c is standardly denoted by the Greek letter Omega (Ω). If the dark energy is included, then the total Ω for our universe is now known to an accuracy of about half of a percent, and to within this accuracy it is equal to 1* [This is good news for theorists, because the inflationary model of cosmology predicts that $\Omega = 1$, and there is also a theoretical plausibility argument in favor of $\Omega = 1$ that can be made independently of inflation. We will return to these issues later in the course.] According to the Planck 2018 estimates† the dark energy is believed to comprise about 69% of Ω. Only 5% of Ω is due to "ordinary" matter, like the material that we are made out of. "Ordinary" matter is also called *baryonic* matter, since the bulk of its mass is that of protons and neutrons, which belong to a class of particles called baryons. The remaining 26% of the total mass density is *dark matter*. This is matter that is known to exist because of its gravitational effects on other matter, but which is not detected in any other way. The composition of the dark matter is unknown, but it is most likely in the form of a dilute gas of some so-far-undiscovered weakly interacting particle.

The time evolution of the $k = 0$ case is rather easy to calculate. From Eq. (3.27), one sees that $E = 0$ implies that

$$\frac{da}{dt} = \frac{\text{const}}{a^{1/2}} . \tag{3.35}$$

The value of the *const* will not be relevant, since it will depend on the arbitrary definition of the notch. One can integrate this equation by rewriting it as

$$a^{1/2} da = \text{const}\ dt , \tag{3.36}$$

which integrates to give

$$\frac{2}{3} a^{3/2} = (\text{const})t + c' . \tag{3.37}$$

The ambiguity of the constant of integration c' simply reflects our freedom to redefine the origin of time. It is traditional in big bang cosmology to define the zero of time to be the moment when the scale factor $a(t)$ vanishes — sometimes regarded as the instant of the big bang. One then has the following important result which holds for a flat, matter-dominated universe:

$$\boxed{a(t) \propto t^{2/3} .} \tag{3.38}$$

On Problem Set 2 you have explored the consequences of this behavior for the scale factor, and now you know how to derive it.

---

* Planck 2018 VI, op. cit., Table 4, Column 4.

† Planck 2018 VI, op. cit., Table 2, Column 6.