Lesson 1 : The expanding (Newtonian) universe

Notes from Prof. Susskind video lectures publicly available on YouTube

Introduction

Cosmology is a very old subject. It goes back thousands of years in the past to the Babylonians and the Greeks. But we are not going to concern ourselves with the cosmological views of Antiquity. We shall start sometime in the second quarter of the twentieth century when Hubble[1] discovered that the universe is expanding.

Despite its ancient origin, the science of cosmology, as we know it today, is fairly recent. It even dates to well after Hubble. The discovery in 1964 of the three degree microwave radiation, also called cosmic microwave background (CMB), is its real beginning. The CMB was rapidly interpreted as a remnant of the Big Bang, which was itself first proposed in the nineteen twenties.

Before that, cosmology was in a certain sense less like physics and more like a natural science. Naturalists study this kind of things, that kind of things. They find a funny star over here, a galaxy that looks weird over there, and so forth. They classify their observations, name them, measure some characteristics about them.

The accuracy with which these things were known was so poor that it was very difficult to build significant physics from the observations. To be true, physicists were involved, however, because many of the things that astronomers observe are of course physical systems. They have momentum, energy. They have all the things that physical systems have. There is chemicals out there, so physical chemists were involved. There were sets of equations attempting to describe

1. Edwin Hubble (1889 – 1953), American astronomer.

and explain the universe of course. But they were wrong.

Right and accurate equations which satisfactorily agree with observations, and have predictive power, are relatively new. They span no more than Professor Susskind's career in physics, which is about fifty years. And that is what we are going to study in this course.

We will study the universe as a system. In other words we will describe it and predict its behavior with equations. If you don't like equations you hold the wrong book in your hands.

Isotropic and homogeneous universe

As always in physics, we start with observations. The first observation that we shall use may not be absolutely true. But it looks like it is approximately true. It is that, viewed from where we stand, the universe is *isotropic*.

By that we mean that in any direction we look, the universe looks roughly the same. It displays some spherical symmetry. Of course, if you look right at a star, it does not look exactly the same as if you look a small angle away where you see no star. But on the whole, averaging over patches in the sky and looking out far enough, so that we get away from the immediate foreground of our own galaxy, the universe looks pretty much the same in every direction. That is the meaning of being isotropic.

This leads to the next step in reasoning. If the universe is isotropic around us – with one exception that we shall see in a moment – then we can bet with a high degree of confidence that it is also pretty close to being homogeneous. This is stronger than being isotropic. Homogeneous means that it is the same in every place – again of course after some averaging. If you went out to, say, sixteen galaxies away from the Earth, and you looked around, what you would see would be pretty much the same as what we see from where we are.

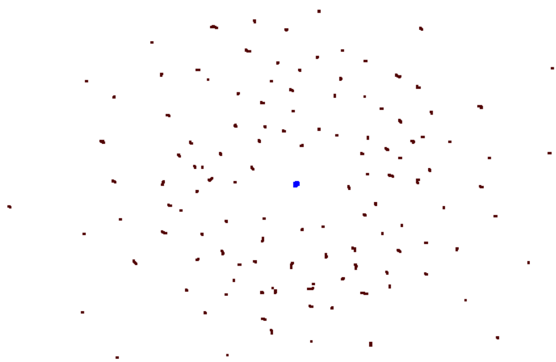

Figure 1 : Isotropic distribution of galaxies around the Earth, with spherical symmetry but *a priori* not necessarily homogeneity.

Why does being isotropic – which means the same in every direction – suggest that, more than isotropic, it is homogeneous ? The argument is very simple. Imagine that there is some distribution of galaxies as shown is figure 1.

Before developing the argument, let's remark by the way that at least in the first part of our study, it does not matter whether we call them galaxies or particles. For all practical purposes, in the first lessons, they are effectively just point masses distributed througout space in the universe. To some extent – that we will qualify in a moment – we can think of them like being the molecules of a gas in a large vessel.

It is useful to have some numbers in mind too. Within what we can see, there are about a hundred billion galaxies in the universe, that is 10^{11} galaxies that can in theory be seen by astronomers with their telescopes, be it in the visual electromagnetic range or in another range. And each galaxy comprises about 10^{11} stars. Altogether we are speaking of 10^{22} stars. If you remember from your chemistry course that Avogadro's number is approximately 6 x 10^{23}, we are talking about an order of magnitude of one mole of stars.

So why should the universe be homogeneous ? The simple reasoning goes like this : if viewed from point A, say the Earth, the universe was isotropic but not homogeneous, it would display some sort of rings or shells of matter like the rings of an onion centered at A. But then, viewed from a point B elsewhere in the universe, it would not look like the rings of an onion centered at B too. So either of two configurations are possible :

a) By accident, or design, we happen to be at the center of the universe, and it looks isotropic for us but not for other people in the universe.

b) It looks isotropic for anyone in the universe, then the only possibility is that it is homogeneous.

Therefore if we reject the idea of being at the center of the universe, as astronomers and physicists do, it must be homogeneous.

This homogeneity of course is *on average*. The same can be said of the air in the room : the density of air is uniform, but that is only true at the scale of a micrometer or more. At the molecular scale, there are fluctuations, places with more molecules, places with less, if only because at a molecule there is a molecule, and next to it there is none. Moreover molecules of oxygen or nitrogen move around incessantly. At room temperature, the average speed of a nitrogen molecule is about 500 meters per second. We shall see that for galaxies however the story is different. They don't skitter around in the universe like molecules, but seem to follow an organized grand movement which we shall study in depth.

The homogeneity of the universe is called the *cosmological principle*. This leads some people, when asked why the universe is homogeneous, to answer : but it is a principle! Remember, however, that it is true in last resort because isotropy has been observed to some degree of approximation and then a simple reasoning leads to homogeneity.

In certain scientific media some astronomers claim that there are structures out there which stretch over very big regions of the visible universe and contradict the cosmological principle. I don't know how to evaluate these claims. But what is certainly true is that the idea of complete uniformity is not exact. As we already pointed out, just the fact that there are galaxies means it is not exactly the same everywhere. In fact there are clusters of galaxies and superclusters of galaxies. So it appears that it is not really homogeneous. It tends to come in some sort of lumps.

But on some big enough scale, like a billion light years roughly, maybe a little less, if you average over that much, the universe seems homogeneous. So that will be our first basic assumption.

A moving grid to track galaxies

The initial model we shall construct is no longer accepted as is, today, at the beginning of the XXIst century. But it is a useful stepping stone in our study, because it is the first model based on observations and real physics. And it clarifies things a lot. So let's go back a few decades to sometime between the forties and the sixties.

The idea of a cosmological principle was itself put forward earlier, but at that time it was not based on observations and people had not any real right to put it forward. When it first appeared it was indeed only a principle. But then, with more and more astronomical investigations, it became a model stemming from observations. And finally the cosmic microwave background discovered in 1964 really nailed it.

The universe we are going to model is formed of galaxies that we can assimilate with particles in a homogeneous gas. Each galaxy on the whole is not electrically charged. It is electrically neutral. But it is not gravitationally neutral. So galaxies interact through Newtonian gravity and that is the only important force on big enough scale.

Gravity is pulling all the stuff together or is doing something to it. But it is a little bit confusing. Pick a galaxy A

in our uniform universe, see figure 2. What happens to it? What forces are exerted on it? Does it accelerate in some direction? But everything is homogeneous, so where does it move to?

We can introduce an origin and perpendicular axes. And for simplicity let's put ourselves at the origin O.

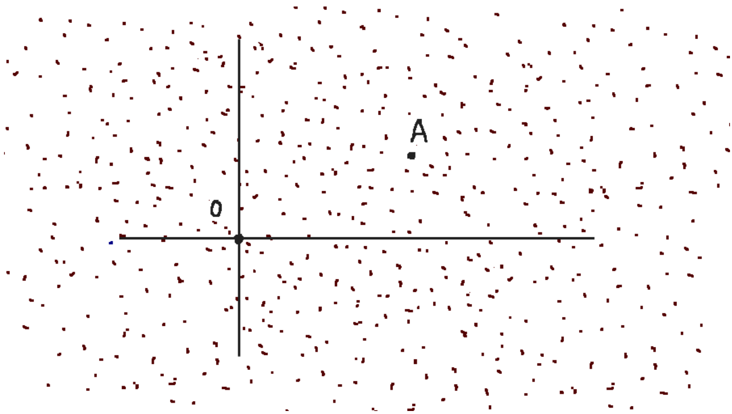

Figure 2 : Reference frame in a uniform universe. For convenience, we represent a two-dimensional world, but think of it in three dimensions. In which direction is A pulled?

We want to figure out the movement of any galaxy A. It looks like A should not go in any direction in particular. Should it stay where it is? The natural thing to guess, it seems, is that the universe should be just static. But that is wrong.

It is the objective of this first lesson to describe how a homogeneous universe must necessarily evolve with time. We are going to work out the actual Newtonian equations of cosmology.

You may have heard that the model of an expanding universe somehow fits together especially well, and that it wasn't really understood until Eintein's general relativity came along. However, the idea that it is general relativity which made it possible to construct the model of an expanding universe is simply false.

There is indeed a coincidence of dates : general relativity was built between 1907 and 1915 – see volume 4 of the collection *The Theoretical Minimum* devoted to it – and the model of an expanding universe was first proposed shortly afterwards in the nineteen twenties. But Newton could have built the model of the expanding universe.

Since Newton didn't do it, we are going to do it here the way Newton should have done it, if only he had pushed a little further the consequences of his theory of gravitation.

To begin with, we shall introduce a set of coordinates with the reference frame shown in figure 2. But there will be an astute twist. They won't be classical static Euclidean coordinates, corresponding to fixed distances on the axes as is usually done, which would enable us to follow the movement of galaxies through the evolution over time of their coordinates.

The coordinates which we introduce are evolving with time, in such a way that the positions of the galaxies, on average, don't move with respect to the coordinates.

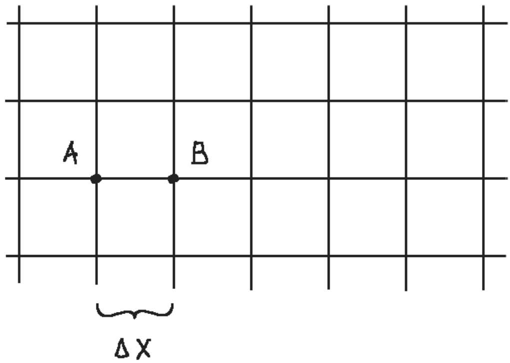

Figure 3 : Fictitious grid of coordinates following in unisson the expanding, or contracting, universe. The two galaxies shown remain at the same lattice points. Their grid-distance remains the same, but their real distance may evolve with time.

The lattice points always go through the same galaxies. In other words, the galaxies in the universe provide a grid. If the galaxies are moving relative to each other, perhaps away from each other or closer to each other, then the grid moves with them.

That this is possible is ultimately justified by observation. When we look unto the heavens, galaxies appear to move in a nice, uniform, coherent way. You can think of them as the raisins in an expanding, or contracting, cake in the owen. They move in unisson. Of course this is on average. Galaxies may move locally from their position on the moving grid, but observation show that, unlike the movement of molecules in a gas, it is a minor secondary movement compared to the general one. On average galaxies are moving very coherently exactly as if they were embedded in a grid, with, as said, the grid perhaps expanding, perhaps contracting – we will come to that – but with the whole grid being sort of frozen with respect to the galaxies.

We choose coordinates, X, Y and Z for our grid. Notice that they are not measuring length because the length of the grid cell may change with time. Although not strictly necessary until later in the reasoning – when we talk about Newton –, we may also think of an origin, for instance the position of our galaxy in the universe. As usual in our figures, for simplicity, we represent the universe with only two dimensions, but we mean three.

Thus we have labelled the galaxies by where they are in the grid. And now we can ask some more questions. Let's start with two points separated by a value ΔX on the grid, for instance the two points shown on figure 3. How far apart are they ?

We don't know yet how far apart they are. But we are going to postulate that the distance between them – the actual distance in meters or in some physical unit that we measure with a ruler, it could be a light year on a side, it could be a million light years on a side – is proportional to ΔX. It is ΔX times a parameter a which does not depend on the position on the grid. It is called the *scale parameter.*

$$D = a \, \Delta X \tag{1}$$

The scale parameter a, which by hypothesis is the same for the whole universe, may or may not be a constant with time too. If it were constant with time, then the distance between galaxies fixed in the grid would stay constant over time. But it may also be time dependent. So let's allow that :

$$D = a(t)\ \Delta X \tag{2}$$

Now let's write the formula for the actual distance D_{AB} between any two galaxies, figure 4. We apply Pythagoras theorem to the grid coordinates and multiply by the scale parameter.

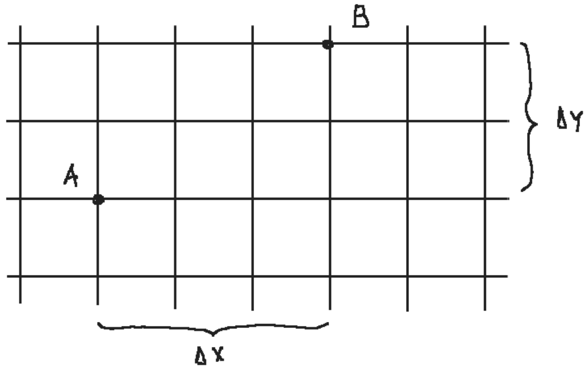

Figure 4 : Calculation of the distance D_{AB} between two galaxies.

This yields

$$D_{AB} = a(t)\ \sqrt{(\Delta X)^2 + (\Delta Y)^2 + (\Delta Z)^2} \tag{3}$$

In other words, we measure the distance along the grid in grid units, and then multiply it by $a(t)$ to find the actual physical distance between the two points.

Let's stress that *a priori* a(t) may or may not be constant in time. But of course in reality it is not. If it were constant in time that would mean literally that the galaxies were just frozen in space and did not move. But that is not what we observe. We see that they are moving apart from each other, in other words that the universe is expanding.

Let us calculate the velocity between galaxy A and galaxy B. We know the real distance between A and B. It is given by equation (3). For simplicity let's just work with the formula in one dimension, already seen in equation (2), which we rewrite below more explicitely

$$D_{AB} = a(t)\ \Delta_{AB}X \tag{4}$$

This way we don't have to worry about Pythagoras theorem. It does not really make any difference.

Equation (4) gives us the distance between A and B. What is the relative velocity between the two galaxies ? It is just the time derivative of the disance. And since $\Delta_{AB}X$ is fixed, the only element which varies in the formula is $a(t)$. Using the usual notation with a dot for the time derivative, we write

$$V_{AB} = \dot{a}(t)\ \Delta_{AB}X \tag{5}$$

Now we can compute the ratio of the relative velocity to the distance. The term $\Delta_{AB}X$ nicely cancels, and we get

$$\frac{V_{AB}}{D_{AB}} = \frac{\dot{a}(t)}{a(t)} \tag{6}$$

Observe that this ratio does not depend on which pair of galaxies A and B we choose. It is true for every pair, no matter how far apart they are, or how close, and no matter in which direction the segment that links them is oriented. The ratio of the velocity to the distance, $\dot{a}/a$, has a name : it is called the *Hubble constant*, denoted H.

$$H = \frac{\dot{a}(t)}{a(t)} \tag{7}$$

The term Hubble *constant* is a bit of a misnomer, because it has no particular reason to be a constant [2]. It depends on the parameter t. Only if the time cancelled out in equation (7) would H be a genuine constant. In fact, we deduce from observations that H is not a constant. But what is important to remember is that it is independent of $\Delta_{AB}X$. It does not matter where you are, which pair of galaxies you are talking about, whether they are close to each other or far apart.

A better name is *Hubble parameter*, or *Hubble function.* And a better notation is $H(t)$. In fact, when we speak of the Hubble constant, we usually mean its value today.

To summarize, the *Hubble law* can be written

$$V = H\ D \tag{8}$$

It is valid for any pair of galaxies in the universe. And H depends on time.

2. When a function depends a priori on two variables, to say that it is a constant or not is ambiguous. It may be constant in one and variable in the other. In fact H is constant in space but not in time.

This law is nothing more than a simple consequence of the initial observation that galaxies in the universe stay nicely homogeneously distributed and don't move much from their position *on a fictitious grid* – which itself may move, and in fact it does. As some witty person once remarked, you should not be surprised that the farthest horse goes the fastest.

Those are the facts that Hubble discovered in the nineteen twenties and nineteen thirties. And from them theoretical cosmologists had something to work with.

Let's say a few more things about this simple model of the universe.

What about the mass within a region ? Consider a region of size $\Delta X\ \Delta Y\ \Delta Z$, and let's take it big enough so that we can average over the small scale structure. How much mass is in there ?

We readily see that the amount of mass that is in there is proportional to $\Delta X\ \Delta Y\ \Delta Z$. The bigger the region, the more mass. And even though the volume changes with time, the amount of mass in it does not. Let's introduce the measure ν = the amount of mass per unit volume of the grid, that is the volume *not being measured in cubic metres*, but being measured in X, Y and Z, which are the labels on the grid. Thus in a given region of volume $\Delta X\ \Delta Y\ \Delta Z$, we can write the mass as

$$M = \nu\ \Delta X\ \Delta Y\ \Delta Z \tag{9}$$

On the other hand, what is the actual volume V of that region ? It is not $\Delta X\ \Delta Y\ \Delta Z$ because the three sides vary with time. We have to take into account the scale parameter $a(t)$ three times, once for each dimension :

$$V = a^3\ \Delta X\ \Delta Y\ \Delta Z \tag{10}$$

Now we can write a formula for the density of mass. We mean the actual physical density : how much mass there is per cubic kilometer, or cubic light-year or whatever units we use. We haven't specified units yet. Later on we will specify some. For the time being, the International System of Units, meters, kilograms and seconds, is fine. What is the density ? It is the number of kilograms per cubic meters, that is the ratio of the mass to the volume. The standard notation for density is ρ. From equations (9) and (10) we readily calculate that

$$\rho = \frac{\nu}{a^3} \tag{11}$$

Let's repeat that the amount of mass in each grid-cell in figure 4 stays fixed. Why ? Because the galaxies move with the grid. So the amount of mass of a given region of the grid stays the same. It is just something we called ν (the Greek letter nu) times the "grid-volume" of the cell, which is 1 if $\Delta X = \Delta Y = \Delta Z = 1$.

Therefore, if a changes with time, so does ρ. For instance if $a(t)$ increases then $\rho(t)$ decreases. Equation (11) is an important formula which we will use from time to time.

So far we haven't done anything that the Greeks themselves couldn't have done. Euclid could have done those calculations. We did not need Newton yet. But now enters Isaac

Newton.

Introducing Newton's gravitation

We shall look at the effect of Newton's gravitation on the universe we just modeled. As the reader knows, Newton was a very self-centered person, so it is natural that he choose the coordinates X, Y and Z of the grid, so that he be at the center of the universe, in other words at the origin of the grid, figure 5.

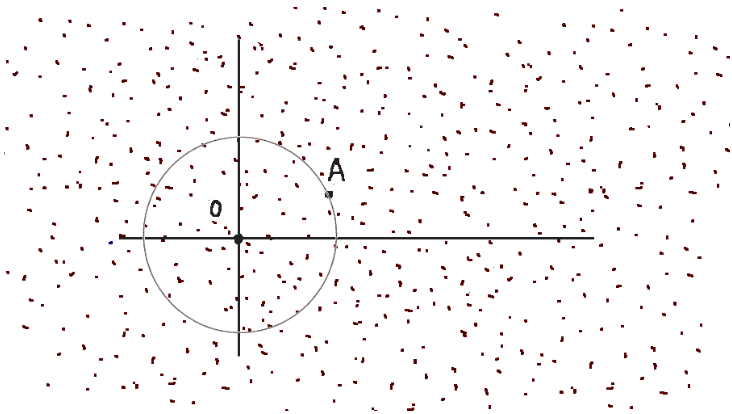

Figure 5 : Calculation of the effect of gravity on galaxy A.

Actually we know, and Newton knew, that we would get the same equations of motion wherever we place the origin and orient the axes[3]. But there is nothing wrong with placing ourselves at the center of the grid.

Moreover, Newton would also say : I'm not moving, I'm standing still. So, for mathematical purposes, Newton is at rest at the center of the universe, at point O in figure 5.

Newton looks at a distant galaxy A. And he wants to know how that galaxy moves. Well, that galaxy moves under the assumptions of Newton's equations. They say that everything gravitates with everything else. But there is something special about Newton's equations. There is a very useful theorem. In fact it is due to Newton too.

Newton's theorem says this : in a frame of reference where everything is isotropic with respect to the origin – it doesn't even have to be homogeneous –, if we want to know what is the gravitational force exerted on a particle A of mass m, then draw a sphere centered at the origin and going through the particle as shown in figure 5. Then take all of the mass M within the sphere and pretend that it is just sitting at the origin. And ignore all the mass in the universe farther away than A from the origin because its net force on A is null. Then the force exerted on A can be calculated as the force due to a unique point mass M located at O.

It is thanks to this fact that we can sit where we are and be only subject to the pulling force of the Earth as if it was all concentrated at its center, 6000 km below us, and not feel

3. What is more, according to Galileo's or Newton's relativiy principle two coordinate systems may move with constant velocity with respect to each other and remain suitable. However they cannot rotate with an angular velocity with respect to each other. And, anyway, here we want the location where we are to be the origin.

at all the immense masses, much more important than that of the Earth, that are out there in the universe and pulling at us too, but with a null net effect. Such calculations of course are frame dependent. We took a frame of reference with the center of the Earth as the origin. We could place the origin elsewhere, at us for instance, but then we would have to be careful that things are not isotropic around us. And a slight variant of Newton's theorem would then have to be used.

Let's call D the actual distance between A and the origin O. If the grid-coordinates of A are X, Y and Z, then

$$D = a(t)\ \sqrt{X^2 + Y^2 + Z^2} \tag{12}$$

To make formulas lighter, let define R as $\sqrt{X^2 + Y^2 + Z^2}$. R is not measured in meters, it just comes from Pythagoras theorem applied to "grid-distances".

Newton's equations are about forces and accelerations. So we want to express the acceleration of the galaxy A relative to the origin. First of all, the velocity is

$$\dot{D} = \dot{a}(t)\ R \tag{13}$$

Then the acceleration is

$$\ddot{D} = \ddot{a}(t)\ R \tag{14}$$

We need not worry about the derivatives of R with respect to time, because R is fixed. That is the nice thing about the fictitious moving grid and the scale parameter $a(t)$. The scale parameter takes care of all the time variability of actual distances. The "grid-distances" don't change.

Now we can write Newton's equation relating force, mass and acceleration, using Newton's law of gravitation for the force

$$m\ \ddot{D} = -\frac{mMG}{D^2} \tag{15}$$

G is Newton's constant, equal to $6,67$ x $10^{11}\ m^3/kg\ s^2$. And the minus sign means that the force is attractive, pulling A toward the origin. That is the convention : force pulling in is counted as negative ; force pushing out is counted as positive. To obtain the acceleration due to gravity, just drop the factor m on each side.

$$\ddot{D} = -\frac{MG}{D^2} \tag{16}$$

This is, in the frame of reference we chose, the acceleration of galaxy A due to gravity. It had better be equal to that which we obtained in equation (14). So we reach

$$-\frac{MG}{D^2} = \ddot{a}(t)\ R \tag{16}$$

We are just pushing the equations. God knows where they will take us. We are following our nose doing the maths.

That is always how we physicists work : we start out with some physical principles, we write down the equations, then we blindly follow them for a while using our math toolbox, until we need to pause and think again.

So at present we are on autopilot, just doing equations.

Let's rewrite equation (16), plugging-in the formula we got for D from equation (12) :

$$\ddot{a}(t)\ R = -\frac{MG}{a(t)^2\ R^2} \tag{17}$$

At some point we might actually reach something that looks interesting. At the moment, we keep trudging through equations.

Let's drop the t parameter in the notations, keeping in mind that a depends on time. Furthermore, let's divide both sides by a and by R

$$\frac{\ddot{a}}{a} = -\frac{MG}{a^3\ R^3} \tag{18}$$

Of course, I secretely know where I'm going. You may have guessed too : a^3R^3 is related to the volume of the sphere. So we want to make it appear in the equation, hoping that we will reach a nice formula, easy to interpret and remember.

The volume of a sphere of radius aR is $\frac{4}{3}\pi a^3 R^3$. Therefore on the right hand side of equation (18) we multiply upstairs and downstairs by $\frac{4}{3}\pi$. If we denote by V the volume of the sphere of radius aR, we get

$$\frac{\ddot{a}}{a} = -\ \frac{4\pi}{3}\ \frac{MG}{V} \tag{19}$$

Now we have M/V in the formula. What is M/V ? It is the density, the quantity of mass per actual volume, which we called ρ. Hence we have

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}\, G\, \rho \tag{20}$$

This is a nice equation. Notice that it does not depend on R anymore. If we know what the density of the universe is – and the density of the universe does not depend on where we are – we can use this equation anywhere. Equation (20) is true for the entire universe, for any region, any galaxy no matter how far away. Had we considered another galaxy or another origin, in figure 5, for our reasoning and calculations, we would have gotten the same equation.

That equation (20) does not depend on R is of course a good thing, because if we want to think of $a(t)$ as a parameter which doesn't depend on where we are in the universe, then R had better drop out.

So Newton confirms what we might have expected, that the equation for $a(t)$ is a universal equation for all galaxies.

Notice that to do the calculations we have been using one frame of reference, but then it turned out, satisfactorily, that the choice of which frame of reference to use was irrelevant [4]. Secondly, and more importantly, all the calculations and the results rest heavily on the fact that the universe is assumed to be homogeneous. The density ρ does change

4. We chose to place ourselves at the center of the frame of reference because it simplified the calculations. Similarly, when studying the movement of the solar planets, we choose a reference frame with the Sun at the origin. That way we obtain simple equations of ellipses. We could use another frame of reference to solve the problem. The calculations and formulas would be quite a bit more intricate, but the end results, that is the movements, would be the same.

with time, but it does not change with space.

Equation (20) is a central fundamental equation of cosmology. It is a differential equation for how $a(t)$ changes with time.

There is a number of consequences to look at. But the first interesting thing to observe is that it is impossible to have a universe which is static. Only if $\rho(t) = 0$ for all t, that is only if the universe is empty, can the time derivative of a and its second time derivative be zero, and can the universe be static. But of course ρ is never zero, therefore the universe must have a time evolution. We derived the fact that the universe is not static.

Friedmann's equation

Our next goal is to figure out the average movement of galaxies in the universe, that is to solve equation (20). A first thing we can do is to replace ρ by an expression involving ν. Unlike ρ, the density ν is literally a constant in space *and* time. Remember that ν is the mass of galaxies per unit of "grid-volume" – a volume with one unit of grid coordinate on each side –, which does not change with time because the galaxies are frozen in the moving grid. While ρ is the actual density, that is the mass per actual volume. The actual volume of a cubic region of grid side one is a^3.

$$\rho = \frac{\nu}{a^3}$$

Plugging this in equation (20), we get

$$\frac{\ddot{a}}{a} = - \frac{4\pi \; G \; \nu}{3 \; a^3} \tag{21}$$

Notice that the presence of $\ddot{a}$ on the left-hand side is not surprising, because Newton's equations are about acceleration. And everything on the right-hand side of equation (21) is a constant except $1/a^3$. It is a differential equation for the time evolution of the scale factor a, or $a(t)$.

It was discovered by Friedmann [5] in the context of his work on the general theory of relativity. The one that usually bears his name however is a variant of it that we will derive below, equation (30).

Equation (21) is consistent with general relativity. Einstein could have derived it, and in fact should have done so. But there is nothing in it that is not just classical Newtonian mechanics.

The equation discovered by Friedmann does not tell us if the universe is expanding or contracting because it does say anything about $\dot{a}$. That depends on the initial conditions, just like Newton's equation for the movement of a stone in the air does not tell us whether it is going upward or downward.

5. Alexander Friedmann (1888 – 1925), Russian physicist and mathematician

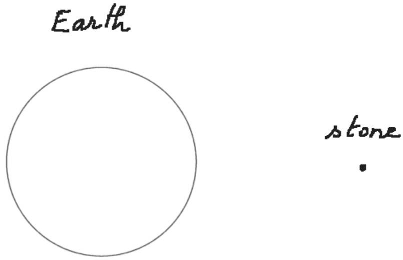

Figure 6 : Movement of a stone in the air.

Before pursuing our study of the movement of galaxies in the universe, let's examine for a moment the movement of a stone in the vicinity of the Earth, figure 6. If we call x the height of the stone above the Earth, it satisfies the following equation

$$\ddot{x} = -\frac{MG}{x^2} \tag{22}$$

This equation tells us that the stone is accelerating toward the Earth. It is the consequence of the minus sign. But whether it is moving away from the Earth or toward the Earth is a question of velocity not acceleration. Is the velocity to the right or to the left ?

We can imagine the beginning of the experiment – that is the initial conditions. Someone located at height x_0 grabs the stone and throws it away from the Earth. Then it will have a positive velocity. We can also imagine the same person throwing the stone in the other direction, or just releasing it. In those cases x will immediately be decreasing.

But the acceleration will be the same. In either case it is the Earth gravity.

There are three possibilities :

a) The stone begins to go up, but after a while turns around and then falls toward the Earth.
b) The stone begins to fall right away.
c) At the outset the stone is thrown away from the Earth so forcefully that it is given the escape velocity. It starts going up, and never returns.

The formula for the escape velocity is given in equation (25) below. It depends on the initial distance from the Earth center. At the surface of the Earth its value is 40 270 km/h, or 11.2 meters per second.

The same phenomenon happens with the movement of galaxies. Friedmann's equation does not tell us whether the universe is expanding or contracting, but it tells us that the second derivative of a is negative. So it means – in this simple model – that even if the universe is expanding, it tends to slow down. If it is contracting, it tends to speed up its contraction.

There is also an analog here of whether we are above or below the escape velocity. We will come to it.

But, first of all, let's stress that we are investigating a classical model built in the nineteen twenties, and that Newton himself could have built. It is what all cosmologists thought was the right thing to do until about the end of the 20th century. It could have been called the "standard model" of cosmololy, or close to it.

Since then astronomers discovered that the expansion of the universe is accelerating, whereas equation (21) describes a decelerating universe. So there must be some other terms in Friedmann's equation. And indeed we will introduce several more terms. Some parts will have to do with Einstein.

Let's go again to just particles, rocks, stones thrown upward from the surface of the Earth. As we saw, the equations are very similar. We shall examine them for a while and take home a couple of lessons. We represent again the Earth, figure 7, and we might as well think of it as a point because Newton proved the theorem that says that, when looking at its gravitational effect on things not inside the Earth, we could think of it as a point.

Figure 7 : The Earth, as a point mass, and a particle outside the Earth. They are separated by a distance x function of time t.

The equation of motion of the particle is Newton's equation $F = m\ddot{x}$. But there is actually a more useful version of Newton's equation which is equivalent. It is the equation expressing energy conservation. Let's write down the energy of the particle on figure 7. It is its kinetic energy plus its potential energy.

The kinetic energy is

$$\frac{1}{2}m\dot{x}^2 \tag{23}$$

And the potential energy is

$$-\frac{mMG}{x} \tag{24}$$

That may come as a surprise, but the sum of the two terms can be positive, null or negative. The total energy does not have to be positive. Only the kinetic energy does. Remember that the potential energy is defined up to an additive constant. We can make it zero wherever we choose, at the surface of the Earth, or at infinity for instance, see volume 1 of the collection *The Theoretical Minimum*. In the above formula (24), the potential energy is zero at infinity. At any finite distance from the Earth it is negative.

Then suppose that the particle shown on figure 7 is at rest at some time t_0. We don't know how it got there. It does not concern us. It is an initial condition : $t = t_0$, $x = x_0$ and $\dot{x}_0 = 0$. At time t_0 its kinetic energy is zero, its potential energy is negative, therefore in that case its total energy is negative. And as we know it will stay so, because in this system the total energy is constant.

The total energy can also be positive. Suppose we now take the same particle at time t_0 at the same position but impart it an initial velocity. If the velocity is big enough then it can outweigh the potential energy and the total energy be positive. Again, as said, since it is a conserved quantity, it will stay positive.

If at the beginning the particle is thrown away from the Earth strongly enough and its total energy is positive, then it cannot turn around. Why ? Because if it turned around, at the turning point the kinetic energy would be zero, and the potential energy would be negative, therefore the total energy would be negative. But it started positive and it is a conserved quantity.

Therefore if the energy is positive the particle doesn't turn around. Conversely, if it turns around the particle's total energy is negative. The case energy = 0 is some sort of edge of the parameter space.

If the total energy is positive the particle just keeps going and going without ever stopping, which means it escapes. If the energy is zero that is exactly the escape velocity. We will ask later whether it escapes or not if the total energy is exactly at zero.

What is the escape velocity ? It is the positive value of v satisfying the following equation

$$\frac{1}{2}mv^2 - \frac{mMG}{x} = 0$$

or equivalently

$$v = \sqrt{\frac{2MG}{x}} \tag{25}$$

It depends on x. And if the particle is given the escape velocity at time t_0 and position x_0, it will move away from the Earth in such a way that at any x it will be at the escape velocity for that x.

In exactly the same manner, the universe can be above the escape velocity, below the escape velocity, or at the escape velocity. What it means is this :

a) If the universe is above escape velocity, initially at some point the outward expansion was large enough that it will never turn around.

b) If it is below the escape velocity, and it is at present expanding, then at some future time the universe will turn around and begin to contract.

c) If it is at the escape velocity, we are again at a kind of edge point of the parameter space.

The escape velocity is also the velocity at which the total energy is equal to 0.

Let's now return to the motion of all the galaxies, and concentrate on one of them, for instance galaxy A shown in figure 5. As usual we think of it as a particle. Even though it is one of the myriad galaxies of the universe, for all practical purposes all this particle knows is that it is moving in the gravitational field of a point mass M at the center O.

Thus, for all practical purposes too, the problem of the expanding universe can be replaced by the problem we just analyzed of a particle moving in the gravitational field of the Earth, figures 6 and 7. It is exactly the same problem.

Let's work out the energetics again : the kinetic and the potential energy of the galaxy, and keep in mind that the sum conserved. Equations (23), (24) and (25), that we wrote before, still hold, except that the distance between galaxy A

and O is now aR instead of x. And v is $\dot{a}R$. Remember that R is fixed because it is the grid-coordinate of A and the galaxies and the grid move together. Said another way, they don't move with respect to each other ; they expand or contract together.

In Newton's frame of reference, the kinetic energy of galaxy A is analogous to formula (23). It is now

$$\frac{1}{2}\, m\, \dot{a}^2 R^2 \tag{26}$$

And the potential energy is analogous to formula (24) :

$$-\frac{mMG}{aR} \tag{27}$$

The sum of these two terms is the total energy corresponding to A.

Let's do the case where the total energy is exactly equal to 0. That is let's find out what is the function $a(t)$ in that case. The other cases are just as easy. They are left as exercises for the reader.

In the case we are considering, the universe is just on the edge. A priori it is not clear yet whether it will turn around and go back or it will keep going, as we saw for the stone thrown away from the Earth.

We start from the equation stating that the sum of the kinetic energy and the potential energy of A is equal to zero. And we are going to use the various things we know.

$$\frac{1}{2}\, m\, \dot{a}^2 R^2 - \frac{mMG}{aR} = 0 \tag{28}$$

Now we do a little bit of algebra on autopilot as before. First, we get rid of m and multiply by 2

$$\dot{a}^2 R^2 - \frac{2MG}{aR} = 0$$

Then, as already done, we want to make the volume of the sphere centered at O and going through A appear in the equation, because we want to work with a density.

$$\dot{a}^2 - \frac{2MG}{aR^3} = 0$$

$$\frac{\dot{a}^2}{a^2} - \frac{2MG}{a^3 R^3} = 0$$

$$\frac{\dot{a}^2}{a^2} - \frac{8\pi}{3} \frac{MG}{\frac{4\pi}{3} a^3 R^3} = 0 \tag{29}$$

Notice that $\dot{a}/a$ is the Hubble constant [6] H, and that

$$\frac{M}{\frac{4\pi}{3} a^3 R^3}$$

is the mass density that we have denoted ρ. So the above equation (29) can be rewritten

6. Despite its name, Hubble constant, H is constant in space, but not in time.

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho \tag{30}$$

That is *Friedmann's equation* in the form it is usually written. It is equivalent to equation (20) in the case of total energy equal to zero.

Equation (20) is Newton's equation, and equation (30) is the energy conservation equation. But they are equivalent just as, for a particle, Newton's equation of motion and energy conservation are equivalent.

Friedmann's equation is more useful. It is not completely general though because we did set the energy to 0. We are just exactly at the critical escape velocity.

This universe is not going to re-collapse. To get a physical feel for the reason why is it so, let's think again at what happens to a projectile if we shoot it at exactly the escape velocity away from the Earth ? What happens as time goes on ? It slows to zero at infinity. It goes slower and slower, and its velocity asymptotically goes to zero, but it never turns around. For the same reason, in our model, the expansion of the universe will grow slower and slower asymptotically, but it will never turn around.

Next step is to solve Friedmann's equation. Again we have to express its right-hand side differently, because ρ depends on time. We need to write explicitely the time dependence of ρ. Recall that it is the simple function

$$\rho = \frac{\nu}{a^3}$$

where only a depends on time, and ν is the density with respect to the grid. By this we mean the quantity of mass per unit grid-volume, that is for instance a cubic volume of evolving side length equal to one fictitious coordinate unit. With the astute grid, the parameter ν is constant in space and time. Incidentally, we can choose our units so that the numerical value of ν is any number we like. Hence the basic equation we have to solve is simply

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{k}{a^3} \tag{31}$$

where k is the constant $8\pi G\nu/3$. Since we may choose our units to make k equal to 1, we may rewrite the equation as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{a^3} \tag{32}$$

If we can solve equation (32), we can solve equation (31). It is straightforward to go from one to the other.

How to solve the differential equation (32) ? Notice first of all that the right-hand side is always positive. In fact it never quite goes to zero, no matter how big $a(t)$ gets. As $a(t)$ gets really big, the right-hand side gets smaller and smaller but never zero. That tells us that $\dot{a}/a$ never becomes zero either. The equality $\dot{a} = 0$ would correspond to the time when the universe would be turning around.

So the Hubble constant never goes to zero. Therefore it never changes sign. But it does slow down. It gets smaller and smaller with time. It is as if the universe just grew tired of expanding but it never got tired enough to stop.

To solve equation (32), in this first lesson we will take the easy way. We will just look for a particular type of solution. But we will come back to this kind of equation because it is absolutely central to all of cosmology. And we can solve them quite easily.

Let's look if we could find a solution of the form

$$a(t) = c\ t^p \tag{33}$$

where c is some constant, and p is some real power. We don't know if there is a solution of this form, but we can try. Since we know that $a(t)$ slows down, p ought to be less than one. Let's use equation (32) to find constraints that c and p must satisfy. First we can write for $\dot{a}(t)$

$$\dot{a} = c\ p\ t^{p-1}$$

Therefore

$$\frac{\dot{a}}{a} = \frac{p}{t}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{p^2}{t^2}$$

Equation (32) tells us that this must be equal to $1/a^3$, that is to $1/c^3 t^{3p}$. So we can write

$$\frac{p^2}{t^2} = \frac{1}{c^3\ t^{3p}}$$

This is possible only if the powers of t are the same. Therefore we first get a constraint on p

$$3p = 2$$

Or $p = 2/3$.

Then to find c we equate the constants.

$$\left(\frac{2}{3}\right)^2 = \frac{1}{c^3}$$

This yields

$$c = \left(\frac{3}{2}\right)^{\frac{2}{3}}$$

The constant c plays no important role. What is interesting is that $a(t)$ expands like t to the power two-thirds.

$$a(t) = c\ t^{\frac{2}{3}} \tag{34}$$

That is the way a Newtonian universe would expand if it was right at the critical escape velocity. It would expand with a such a scale factor $a(t)$, and everything, all galaxies, would separate from each other over time as t to the two-thirds power. That is quite a remarkable derivation!

Newton should have done it. It is somehow annoying that he did not do it. He speculated a lot about the evolution of a homogeneous universe, but stopped right on the threshold of doing this calculation. One of the reasons he did not do it may be that he was a believer in the literal truth of the

Bible. In mid-XVII-th century, shortly after the birth of Newton, an Irish bishop by the name of James Ussher had meticulously calculated from biblical data that the creation of the universe had taken place on 22 October 4004 BC in early evening – just after tea time. It was difficult to make it jibe with the elements of cosmological knowledge that were already known at that time and with the above model [7].

Let us stress that the model of universe we built and analysed, for which we established the rate of expansion given by equation (34), is a Newtonian universe, furthermore with zero total energy, that is at exactly escape velocity. It is a pure 3D Euclidean universe, infinite and spatially flat, plus a straightforward time dimension, without any interesting Einsteinian geometry.

We did it first of all because it is simple and a good illustration of the way physicists work. Secondly because it contains, in a simple form, a lot of the physics that we are going to be dealing with in this course. It gives us a first model universe with a scale factor that increases like the two-thirds power of the time.

When the universe is below or above escape velocity there

7. Every great scientist's biography is usually interesting, but Newton's is fascinating. He developed his most important mathematical and physical ideas in his twenties, around the time of the Great Plague of London in 1665. During his lifetime he wrote much more about religion and alchemy than about science. Despite his superior intellect he got financially wiped out by the Tulip Bubble. He never married. After the age of fifty he became Master of the Mint of England, and for the last 30 years of his life made important contributions to the management of gold and silver money. Some authors credit him with inadvertently creating the gold standard, see Peter L. Bernstein's book *The Power of Gold*, John Wiley, 2000.

is another term on the right-hand side of equation (34). We will study it in the next lesson. We will examine the three possibilities : less than escape velocity, at escape velocity, and above escape velocity. Recall that it is analogous to the initial velocity given to a stone thrown away from the Earth.

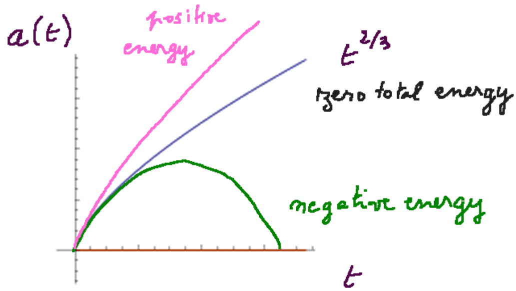

Figure 8 : Scale parameter as a function of time in the three configurations of total energy.

Figure 8 is a familiar diagram representing the curve $a(t) = t^{2/3}$, that is the rate of expansion of the universe at exactly the escape velocity, which corresponds to total energy equal zero, and the curves for a faster expanding universe (positive energy), or a contracting universe (negative energy).

In all three cases, the tendency is to bend over because the expansion speed is slowing down. In the third case, of course, this is before the turning point, after which the collapse speeds up again.

The real universe however, as was discovered at the end of the XXth century, doesn't quite look like that. It started following the middle curve, but then bent upward.

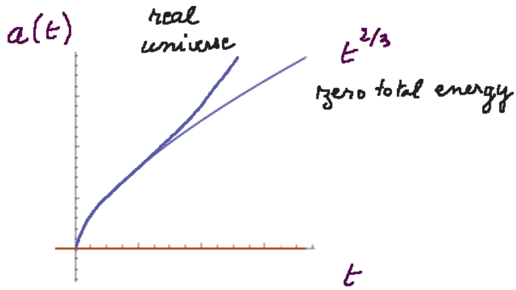

Figure 9 : Real universe expansion, starting like the zero energy universe, then accelerating.

One last comment on the expanding universe : the description of this grand movement is *on average*. To the largest observable distances the universe appears to be expanding. And we modeled it as a homogeneous, *matter dominated* universe – what is meant by that will be explained in the forthcoming lessons. Hubble law is not exactly true for all possible distances. It becomes more and more accurate as distances get larger.

There are myriad relatively small regions in the universe where things are contracting, just like in the room the air presssure, measured on a small scale, is not exactly the same everywhere. Even when observing things at a higher level than the molecular scale, there are fluctuations, places where the air is more dense, places where it is less dense.

Hubble law is certainly not accurate for things which are bound together by gravity or any other force that may pull them together. Here and there we find galaxies which have a *peculiar motion*. It is the technical term for things which display a movement away from the average expansion.

It is the case of Andromeda and the Milky Way. The Andromeda galaxy is not receeding away from ours but moving toward it. Whatever way it was formed, it happened in a pocket which was dense enough, slightly out of the ordinary, so that these two galaxies have enough mass to overcome the effect of global expansion. It is a fluctuation away from the norm.

Averaged over a large enough volume, however, everything is moving away from everything else. We think of galaxies as embedded in a grid, and the grid is expanding with a scale parameter $a(t)$ satisfying Hubble law.