8. The Expanding Universe, Revisited

Now that we have learned something about Einstein's theory of gravity, we are ready to revisit what we have learned about the expansion of the universe from a more sophisticated point of view.

A note on numbering: If you seem to be missing section 7 of the course notes, it isn't a mistake. Section 7 was "Einstein vs. Newton," covered in Barbara Ryden's guest lecture. I haven't made separate notes for this section, but I did post a pdf version of Prof. Ryden's slides on the web page.

Reading: The reading for this section is Chapter 7. For now you can focus on sections 7.1-7.3.

Einstein vs. Newton: A Recap

Einstein's theory of gravity and Newton's theory of gravity are *conceptually* very different:

- Newton: Matter exerts gravitational force. Gravitational force causes masses to accelerate.
- Einstein: Matter and energy curve spacetime. Freely falling bodies follow "shortest paths" through curved spacetime.

"Freely falling" means not acted on by anything other than gravity.

Curved spacetime also affects the path of photons: gravity bends light.

We will often refer to Einstein's theory of gravity as General Relativity, or simply GR.

When
(a) velocities are much slower than c
(b) gravity is "weak" (light-bending angles are small)
Einstein's theory makes predictions that are almost identical to Newton's theory.

For example, it makes very similar predictions for the orbits of planets.

Gravity is almost always "weak" in the above sense (light-bending angles are much, much less than 1 radian) unless you are close to the event horizon of a black hole.

But when one measures precisely enough to tell the difference, Einstein's predictions work, and Newton's do not.

For example, the orbit of Mercury is not a perfectly closed ellipse: the long axis of the ellipse rotates each orbit, completing a full cycle every 3 million years (13 million Mercury orbits).

(If you want an illustration of what this looks like, Google images of "precession of Mercury.")

Other planets show similar, but smaller, deviations from Newtonian predictions, all of them well explained by Einstein's theory.

GR also predicts that gravity affects the flow of time, hence the speed of clocks.

For example, GPS satellites orbit at 20,000 km altitude, where the earth's gravity is weaker than it is on the surface ($R_{\rm earth} = 6400\,{\rm km}$.) Hence clocks on GPS satellites run faster than clocks on earth, by about one part in 10 billion.

If this effect were not taken into account, GPS locations would drift away by about 1 foot every 10 seconds — so they wouldn't be much use!

(In practice, there are several different relativity effects that are important at this level.)

In addition to small quantitative differences, Einstein's theory makes some qualitatively distinctive predictions:

- light bending, gravitational lenses
- gravity affects the flow of time
- existence of black holes, objects whose gravity is so strong that it traps light
- gravitational waves, ripples of curved spacetime that propagate at the speed of light

All of these predictions have been confirmed, directly or indirectly, by experiments or astronomical observations.

Particularly important to our discussion of cosmology: GR allows the space of the universe itself to be curved.

For a homogeneous universe (same density everywhere), there are three possibilities:

• positively curved, like the surface of a sphere, finite total volume but no edge

• flat, like a plane, infinite

• negatively curved, like the surface of a Pringle's potato chip (or horseback saddle), infinite

These cases can be distinguished, at least in principle, by measuring the angles of triangles. (See Ryden lecture and textbook figure 6.2.)

Energy, pressure, and repulsive (?) gravity

Another subtle but important difference between GR and Newtonian gravity: when we compute the gravity from (say) a collection of atoms, we must account for their pressure as well as their mass.

This is a consequence of Einstein's famous formula $E = mc^2$, which expresses an equivalence between mass and energy.

High pressure $\longrightarrow$ atoms moving fast, have lots of energy $\longrightarrow$ produce stronger gravity.

But if atoms are moving much slower than c, as they usually are, then this is a small effect.

Fields of energy can also have pressure.

For example, a powerful magnet creates a magnetic field that has both energy and pressure.

For an energy field (as opposed to a collection of atoms), the contribution of pressure to the gravitational effect can be important.

A field with strong *negative* pressure could in principle produce *repulsive* gravity according to GR.

What does "negative pressure" mean?

If something (a gas, or a liquid, or an energy field) has positive pressure, then it "wants to expand" and will do so unless held in check by something else.

For example, the high pressure gas in a balloon is held in by the tension of a balloon. The high pressure gas in the sun is held in by the gravity of the sun.

"Something else" can include pressure of the surroundings — gas only expands if it is *higher* pressure than the gas around it.

Something with negative pressure "wants to contract" — negative pressure is basically another word for "tension."

Caution: Note the confusing way this works. If you have a gas or field with positive pressure, the pressure tends to make it expand, but the *gravitational* effect of the pressure is attractive.

If you have a field with negative pressure (tension), the pressure tends to make it contract, but the *gravitational* effect of the pressure is repulsive.

All the normal materials and fluids and gases and fields that we know of have positive pressure, or if they have tension then it is too small to produce a noticeable gravitational effect.

I'll refer to all of this normal stuff collectively as "attractive" matter and energy.

An "exotic" form of energy with strong negative pressure could produce repulsive gravity instead of attractive gravity.

A static universe?

Einstein completed his theory of General Relativity in 1915.

He showed that it successfully explained the orbit of Mercury, which convinced him that it was correct.

In 1917 he applied it to cosmology.

He assumed (correctly) that if he wanted to describe the universe on large scales (i.e., much larger than the distances between galaxies) he could approximate it as homogeneous, filled with matter of constant density ρ.

He proposed that space is curved like the surface of a sphere, so that the universe is finite but has no edge.

He assumed (incorrectly) that the universe is static, neither expanding nor contracting.

He found that GR didn't allow such a universe: gravity should make it contract.

Einstein responded by adding a new term, a "cosmological constant," to the equations of GR.

In effect, he introduced exotic energy with negative pressure, pervading all of space, whose repulsive gravity was exactly what was needed to balance the attractive gravity of matter.

Because the exotic energy is present everywhere, tension cannot make it collapse.

Expanding universe: Friedmann and Lemaitre

In the 1920s, Alexander Friedmann (Russian) and Georges Lemaitre (French) developed cosmological models based on GR, assuming a homogeneous universe, but dropping the assumption that it was static.

With no cosmological constant, they found that the universe must expand or contract, and that if one tracks backward in time it must begin from infinite density.

When Hubble discovered the expansion of the universe in 1929, Einstein concluded that Friedmann and Lemaitre were right and that he was wrong to introduce the cosmological constant in the first place.

(Some historians of science argue that Lemaitre actually discovered the expansion of the universe two years before Hubble.)

The scale factor $a(t)$

It is usually most useful to think of the expansion of the universe as an expansion of space itself, carrying galaxies along like corks in the current of a river.

Galaxies that are "at rest" in the expanding space are carried away from each other.

If two locations in the expanding space (e.g., marked by two galaxies) are separated by a distance x today (time $t = t_0$), then at some earlier time $t < t_0$ they were separated by a smaller distance $a(t)x$.

The notation $a(t)$ means that a is a function of time t.

The *scale factor* increases from $a(t) = 0$ at $t = 0$ (the beginning of the universe) to $a(t_0) = 1$ today.

The *same* scale factor applies to all separations in the universe. When $a(t) = 0.5$, all separations were a factor of two smaller; when $a(t) = 0.1$ they were all ten times smaller.

At any time t, when the Hubble constant is $H(t)$, the *Hubble time* $t_H = H^{-1}$ is the time it would take $a(t)$ to double if space kept expanding at a constant rate.

For example, with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, the universe would double in size in 14 billion years.

The Friedmann equation

Friedmann used Einstein's GR equations to come up with an equation (known as the Friedmann equation) for $a(t)$.

Because GR gives the same results as Newtonian gravity for velocities $v \ll c$ and weak gravity, we know the result must be the same as the one we got earlier from considering a homogeneous expanding sphere.

The acceleration at the edge of the sphere is

$$\mathrm{acc} = \frac{-GM}{r^2} = -G \times \frac{4}{3}\pi r^3 \bar{\rho} \times \frac{1}{r^2} = -\frac{4\pi G}{3} r \bar{\rho(t)}.$$

The negative sign means that the expansion is slowing down instead of speeding up.

This can be converted to an equation for $a(t)$ by using $r = a(t)x_{\rm sph}$, where $x_{\rm sph}$ is the sphere's radius today, and $\bar{\rho}(t) = \bar{\rho}_0/[a(t)]^3$.

If we include only attractive matter and energy, then the density of the universe determines whether $a(t)$ will grow forever or eventually stop and begin to decrease.

The Einstein/Friedmann approach also gives us a way to calculate the impact of exotic energy, if it exists.

Specifically, the acceleration equation including pressure becomes

$$\text{acc} = -\frac{4\pi G}{3} r \left(\bar{\rho} + \frac{3p}{c^2}\right),$$

where p is the pressure.

Strong negative pressure could thus lead to acceleration instead of deceleration.

Cosmological Redshift

The expansion of space also stretches the wavelengths of photons, in proportion to $a(t)$.

(You can show this using the equations of GR.)

This gives us a different way to think about redshifts in the expanding universe.

Light from a distant object left it when the universe is smaller than it is today, so its wavelength has been stretched en route by the expansion of the universe.

Light emitted with wavelength λ_e is observed with wavelength λ_o, where

$$\frac{\lambda_o}{\lambda_e} = \frac{a(t_o)}{a(t_e)}.$$

We define the redshift

$$z = \frac{\lambda_o}{\lambda_e} - 1.$$

The corresponding velocity from the Doppler shift formula is just $v = c \times z$.

If we look at objects close enough that the redshift is small (roughly $z < 0.1$, $v < 0.1c$), then this gives us the same answer as Doppler shifts and Hubble's law.

[It takes some calculation to demonstrate that the above statement is true.]

We therefore don't have to abandon our previous way of thinking about redshifts and Hubble's law, but connecting to the expansion factor $a(t)$ gives us a more general way of thinking about redshift, which still works when the redshift isn't small.

The most distant galaxies and quasars that have been discovered so far have redshifts around $z = 8$, so their light was emitted when the universe was 9 times smaller than it is now!

If we also know the history of $a(t)$, then we can figure out how long the light has been traveling to get to us.

The expanding balloon analogy

We can now return to the balloon analogy for cosmic expansion with new appreciation.

It describes a 2-dimensional universe on the surface of the balloon, and the radius of the balloon is proportional to $a(t)$.

As the surface of the balloon expands, objects on the surface are carried away from each other, even if they aren't moving around on the surface.

A perfectly homogeneous universe would correspond to a perfectly spherical balloon, with the same curvature everywhere.

Local structure (galaxies, clusters of galaxies, etc.) produces dimples of curvature on the balloon, like the dimples on a golf ball.

Peculiar velocities correspond to motions of galaxies on the surface of the expanding balloon, caused by these local dimples.

Light emitted from a star or galaxy or quasar travels across the surface of the balloon, at the speed of light.

As the light travels, the universe continues to expand, and the wavelength of the light stretches in proportion to $a(t)$.

By the time light gets to an observer, it has been redshifted by the factor $a(t_o)/a(t_e)$.

Einstein's theory of gravity allows 3-dimensional space to be curved like the surface of a sphere (though we don't have a 4-dimensional perspective from which to visualize it).

However, our universe could also have flat space, in which case the analogy would be an expanding rubber sheet, or negatively curved space, in which case we should think of an expanding rubber sheet that is curved like a potato chip.

The critical density, again

According to GR, a dense universe will have positively curved space, like a sphere, while a low density universe will have negatively curved space, like a potato chip.

We previously identified the critical density as the boundary between a gravitationally bound universe ($v < v_{\rm esc}$), which will eventually recollapse, and an unbound universe ($v > v_{\rm esc}$), which will expand forever.

In GR, it turns out that the *same* critical density is the boundary between a universe that is positively curved and one that is negatively curved.

If $\rho = \rho_{\rm crit}$ exactly, then space is flat.

If $\rho > \rho_{\rm crit}$, then space is positively curved. However, if ρ is just very slightly above $\rho_{\rm crit}$, then the radius of curvature is very large — much larger than the speed of light times the age of the universe.

If $\rho < \rho_{\rm crit}$, then space is negatively curved.

The possible existence of exotic energy introduces a complicated twist.

Attractive energy/matter and exotic energy both contribute to curvature, so the geometry of space depends on the *sum* of the densities of attractive energy/matter and exotic energy.

Attractive energy/matter and exotic energy have *opposite* gravitational effect. Whether the universe is bound or unbound therefore depends on the density of attractive energy/matter *minus* the density of exotic energy.