The expanding universe

History

The discovery of the expanding universe is remarkably recent. **Slipher** began a programme of spectroscopy of galaxies in 1912, and for about 10 years had the field pretty well to himself. This was a period in which the true nature of galaxies as large stellar systems was not known. As you learned in the **great debate**, this was first proved in 1924, when **Hubble** detected the characteristic light curves of **Cepheid variable** stars in M31. By this time, Slipher had already established the expansion of the universe.

The Doppler effect

Slipher's tool was, of course, the **Doppler shift**. What he found was that almost all galaxies showed a shift of characteristic spectral lines (e.g. the Sodium line at 5892Å that makes street lights yellow) to longer wavelengths. We define the **redshift**, symbolized z, as the ratio of observed to emitted wavelength:

$$\boxed{1 + z \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm emit}} = \frac{\nu_{\rm emit}}{\nu_{\rm obs}} \simeq 1 + v/c.}$$

By this definition, z is negative for an object coming towards us (as indeed M31 is, at $300\,{\rm km\,s^{-1}}$). We then have a **blueshift**. Note the "$\simeq$" sign at the end; this Doppler formula only applies when $v \ll c$.

In cosmology, it is often necessary to consider large redshifts (the record galaxy has $z = 5.6$). In this case, it is better not to try to think about the redshift as being due to a recession velocity. A simpler and more powerful interpretation is given below.

Hubble's law

By the 1920s, Slipher had established that almost every galaxy showed a redshift, up to thousands of ${\rm km\,s^{-1}}$. Also, the fainter galaxies showed larger redshifts. By this time, **Einstein** had produced his theory of gravity (**general relativity**), which allowed the dynamics of the whole universe to be considered. Some workers had already suggested that galaxies should show a redshift proportional to distance. The ground was therefore well prepared for Hubble, who was able to resolve a number of galaxies into stars, as he had done with M31 in 1924, and so get their distances. Although his data were rather poor, this was the first evidence for a linear distance-redshift relation:

$$\boxed{v = H\,d.}$$

This is now known as **Hubble's law**.

The quantity H is called **Hubble's constant**, and it measures the rate of expansion. It is a bad name, because H in fact changes with time, as we will see. The **units** of H are $\mathrm{km\,s^{-1}Mpc^{-1}}$; 1 Mpc is 10^6 parsecs, where 1 pc is 3.09×10^{16} m or 3.26 light years. This odd length unit stands for a parallax of 1 arcsecond – i.e. a star at 1 pc moves on the sky by 1/3600 of a degree when the Earth is at opposite ends of its orbit. The Mpc is a convenient length unit for cosmology, because the average distance between galaxies is about 1 Mpc.

The best modern estimate for H is $65\,\mathrm{km\,s^{-1}Mpc^{-1}}$, but this is uncertain by about 10%. This is a pity, since many physical quantities of interest depend on H. For example, Hubble's law is normally used in reverse: we measure v, and so deduce the distance to an object (not easy otherwise) by $d = v/H$. The smaller H is, the larger the distance. Where a value of H is needed in these notes to obtain a physical number, I have assumed $65\,\mathrm{km\,s^{-1}Mpc^{-1}}$. However, books or articles you read may assume other values (a common convention is to scale results to the nice round number of $H = 100\,\mathrm{km\,s^{-1}Mpc^{-1}}$).

Isotropic expansion

The expanding universe seems to place us at the centre of things, as if a mighty explosion occurred at our location in the past. Suppose such a thing happened at $t = 0$: debris would fly off at varying speeds, and at time t a given piece would have reached radius $r = vt$ (if v stays constant for each piece of debris). This is Hubble's law, $v = r/t$, and it tells us that the time since the explosion is related to H, via the **Hubble time**:

$$\boxed{t_{\mathrm{H}} = H^{-1} = 15.0 \times 10^9\,\mathrm{yr}.}$$

Later, we will see that this does indeed give the correct characteristic time for an expanding universe: the entire universe began life 10 to 15 billion years ago. This origin of time is commonly called the **big bang**, although we will see that it is very different from a simple explosion.

Of course, the **Copernican principle** says that we shouldn't expect to occupy such a special place. Is it possible that all observers in the universe think that they are at the centre? The answer is yes, as is easily shown if you are happy with vector algebra (non-examinable). Write Hubble's law about us as $\mathbf{v} = H\,\mathbf{r}$ (velocity vector points in the same direction as radius vector for all galaxies). Suppose you pick galaxy a and write the particular form of Hubble's law for it: $\mathbf{v}_a = H\,\mathbf{r}_a$. If you subtract this from the general law, you get $\mathbf{v} - \mathbf{v}_a = H\,(\mathbf{r} - \mathbf{r}_a)$. On the left, you have the velocity relative to a; on the right, you have H times the position vector of any galaxy with a taken as the origin: a thinks it is the centre of the universe too. If you aren't very familiar with vector algebra, the same point can be demonstrated with a diagram when we discuss the cosmic scale factor below.

In this construction, we might end up in general with more galaxies on one side of the sky than another, so there would still be a special centre. The **Cosmological principle** guesses that the universe is in fact a uniform and symmetric place, and that the distribution of galaxies is **isotropic** about each observer. It is then easy to show that the density must be a constant everywhere at a given time. Of course, the real universe can't be like this, otherwise galaxies would be forbidden. However, we can consider reality as being a perturbed version of the ideal case.

Curved space

What happens at infinite distance in an isotropic expanding universe?. The cosmological principle guarantees that any galaxy, however distant, is surrounded by an isotropic distribution of neighbours – so the universe has no edge. This means that the question 'what does the universe expand into?' is misleadingly worded: we are not dealing with a single point explosion that ejects debris into some surrounding void. All that happens is that every galaxy progressively gets further away from every other galaxy.

It sounds like this must mean that the universe is infinite, with galaxies continuing forever in every direction. This may be true, but it is not the only possibility. General relativity introduced the concept of curved space, and the universe could be the three-dimensional analogue of the surface of a sphere. In other words, there may only be a finite size to the universe, and if you travel far enough in a straight line, you return to your starting point. This would be a space of constant positive curvature, but the curvature can also be negative (impossible to visualize); there are thus two possible kinds of universe:

Closed universe :	positive curvature; finite volume
Open universe :	negative curvature; infinite volume

One of the big questions in cosmology is to decide which kind of universe we live in.

Cosmic scale factor and redshift

The uniform Hubble velocity field arises automatically if we consider a distribution of galaxies in which all separations increase with time in proportion to a **scale factor**, $R(t)$:

$$\boxed{\text{separation} \propto R(t).}$$

For the explosion model, where debris travels at unchanging velocity v, the separation is just vt, so $R(t) \propto t$. In general, gravity causes the expansion to decelerate, so that the variation of scale factor with time is not linear.

The scale factor is obviously related to H, since the more rapidly R changes, the faster galaxies will move and the higher H will be. A more direct measure of R comes from the cosmic redshift. We have seen that this increases the wavelength of radiation from distant objects. These are the ones for which the radiation has also been travelling the longest, so we see that radiation travelling through an expanding universe tends to stretch. Suppose we make a small universe by trapping some radiation in a box with silvered walls, which then expands. The standing waves in the box will keep their form, so that wavelengths stretch exactly in proportion to $R(t)$. The redshift therefore tells us how much the redshift has expanded since the objects we see emitted their light:

$$\boxed{1 + z = R(\text{now}) \,/\, R(\text{when light was emitted}).}$$

An object at $z = 1$ is therefore seen now as it was when the universe was half its present size. This is the best interpretation of the redshift when z is large.

It is worth noting that we have no direct proof of the expansion: distances are too great to measure directly. Over the years, a number of people have tried to construct models in which the universe is static, but something else causes the redshift – such as the **tired light** theory, which suggests that photons spontaneously lose energy over very long distances. The problem with this is that it is rather easy to cook up some new effect that only operates on cosmological scales. The rules of the game are therefore that we try to explain the universe using only the laboratory laws of physics; only if we find a contradiction will there be any justification for claiming that there is something missing from the standard toolkit.