

General relativity, 7

The expanding universe

The fact that the vast majority of galaxies have a spectral redshift can be interpreted as implying that the universe is expanding. This interpretation stems from the Doppler effect in which the relative motion of an emitter and a detector produces a frequency shift of the detected light with respect to the emitted light. The frequency shift formula can be written as a

wavelength shift formula: $\lambda_d = \left(\frac{1+\beta}{1-\beta}\right)^{1/2} \lambda_e$. Now, the galactic spectral shift factor $z = \frac{\lambda_d - \lambda_e}{\lambda_e}$

can be written in terms of relative speed as $z = \left(\frac{1+\beta}{1-\beta}\right)^{1/2} - 1$. For $\beta \ll 1$, the square root on the

right hand side can be evaluated using the binomial expansion (what else?), with the result being $z \approx \beta$. In 1931, Hubble published observed redshift data for 20 or so galaxies containing visible Cepheid variables, which could be used to determine the galactic distances (D) from Earth. The z -values he measured ranged up to about 0.05 and the corresponding distances ranged to about 100 Mly. Hubble found a remarkably linear relation: $z \approx KD$, where K was a constant. If the galactic spectral shift can be interpreted as due to motion, then, except for a very small number of close-by galaxies that seem to be moving randomly relative to the Milky Way, most galaxies seem to be moving away from us—the farther, the faster. Assuming that there is nothing physically unique about the Milky Way, this galactic recession should be observable from any galaxy. Thus, Hubble's z - D relation appears to imply the universe is expanding more-or-less equally in all directions.

Since Hubble's original work, many more galactic redshifts have been measured—using, appropriately enough, the Hubble Space Telescope. In addition, a new kind of standard candle—namely, light from “supernovae type Ia”—has greatly expanded our distance measuring capability. At present, the Hubble z - D rule is well established for z up to about 0.1. It is conventional to write $z \approx KD$ as $v \approx KcD = H_0 D$, where H_0 is known as the “Hubble constant.” The best-measured value of H_0 , based on redshifts of about 200 galaxies, is 72 ± 7 km/s/Mpc. One Mpc (mega-parsec) = 3.262×10^6 ly; thus, $z = v/c = 0.1$ corresponds to $D \approx 4 \times 10^8$ ly. More recently, data from the WMAP satellite (more about WMAP later) using a different, indirect measurement puts the value of H_0 at 70.0 ± 2.2 km/s/Mpc. (There are now other slightly different values based on still different measurements.) Notice that the dimensions of H_0 are $LT^{-1}/L = T^{-1}$, so $1/H_0$ is a time.

Example: $1 \text{ ly} = (c) \times (1 \text{ y}) = 3 \times 10^5 \text{ km-y/s}$. $1 \text{ Mpc} = 3.26 \times 10^6 \times 3 \times 10^5 \text{ km-y/s} = 9.78 \times 10^{11} \text{ km-y/s}$. Thus, $1/H_0 = 1 \text{ Mpc}/70 \text{ km/s} = 9.78 \times 10^{11} \text{ km-y/s}/70 \text{ km/s} = 14.0 \times 10^9 \text{ y}$. As we will see this is an extremely important value.

Friedmann-Lemaître-Robertson-Walker (FLRW) universe

We would like to have a theoretical framework for understanding Hubble's observations as well as the other central facets of modern cosmography, in particular, the cosmic microwave background and the observed cosmic hydrogen-helium-deuterium-lithium abundances. Because Einstein's general relativity is so successful at joining relativity and gravity for simple

systems of masses, it is natural to ask whether GR can also provide a usefully predictive description of gravity at the cosmological level. The starting point for such a model is the assumption that there are, on average, no special places in the universe (on average). Thus, the conventional cosmological model assumes that the universe is spherically symmetric and uniform everywhere. (That approximation is more believable on large length scales where clumpiness becomes less obvious. That might be 10s or 100s of millions of light years.) The goal is to find a solution to Einstein's equations that is valid *inside* the uniform mass-energy distribution of such a universe.

Alexander Friedmann produced the first solution. His 1922 paper on the subject annoyed Einstein, because the "size" of Friedmann's universe had to be dynamically changing—something Einstein "knew" was wrong. Remember, this was before Hubble discovered other galaxies and their associated red shifts; at that time, common wisdom dictated that the universe must be static, always as it now appeared, and infinite in extent. (Such a universe poses a dilemma: every direction in the sky should intersect a star, so the night sky should be bright in every direction—which it clearly is not. This is known historically as *Olbers' Paradox*. As we will see, Hubble's expansion result fixes this problem.) In fact, Einstein already knew that his original 1915 version of general relativity had to describe a dynamically evolving universe, and to "stop" it, he proposed a modification that included a "cosmological constant" term—a kind of fudge factor that made his equations time-independent. Einstein wrote a letter to the journal in which Friedmann's paper appeared saying that Friedmann had made a mathematical error. Friedmann was sure that he hadn't and for two years tried to get Einstein to retract his criticism. Eventually, Einstein recanted, but still maintained that Friedmann's calculation was physically irrelevant. Friedmann died shortly after, without seeing the vindication of his work that would result from Hubble's.

In 1927, Georges Lemaître, trained as both a physicist and a Catholic priest, reinvented Friedmann's solution, without having read Friedmann's paper. He sent a manuscript to Einstein detailing his expanding universe, which elicited, in essence, the response: "Yes, I know this. Friedmann has already showed it to me. His physics was wrong and so is yours." In a subsequent paper, Lemaître went even further. He proposed that if the universe is now expanding, earlier it must have been denser and hotter, and that all of the matter in the universe might have originated from the decay of one giant "radioactive atom." Of course, once Hubble's findings became known, Einstein had to reconsider his objections. Famously, he called introducing his "cosmological constant" the biggest mistake of his life (as we will see, getting rid of it was even bigger!) and warmly and publicly embraced Lemaître's "primeval atom" idea ("this is the most beautiful and satisfactory explanation of creation to which I have ever listened"). (Lemaître was inducted into the Pontifical Academy of Science in 1936 by Pope Pius XI for proposing a plausible scientific rationale for the creation story in Genesis, though Lemaître never claimed such a connection himself.)

Finally, in 1935 Howard Robertson and Arthur Walker proved that the solution of Einstein's gravitational equations by Friedmann and Lemaître is the *only one* possible inside a spherically symmetric, uniform mass distribution. Curiously, (in America) the solution is now often referred to as the "Robertson-Walker metric," but more properly should be attributed to the whole cast of characters. In the literature, this attribution is usually abbreviated "FLRW." Though not the usual convention, it is more poetic and hopeful to invert the R and W, as we do subsequently, giving "FLWR"—the "flower" universe. In the FLWR universe, proper time is given by

$$(d\tau)^2 = (dT)^2 - [a(T)]^2 \left[\frac{(dr)^2}{1 - k(r/r_0)^2} + r^2 (d\theta)^2 + r^2 \sin^2 \theta (d\phi)^2 \right]. \quad (1)$$

Note that in (1) time is the same everywhere, so all clocks are always synchronized. The parameter r_0 is a proper length scale (e.g., the radius of the visible universe) at some reference time (such as the present). The space part of (1), without the a -multiplier, marks off *coordinate* distances, physical distances at the reference time. In this s-t, “free-falling” objects are assumed to be at fixed coordinates (r, θ, ϕ) and physical distances between them change in time according to the scale factor $a(T)$, which is dimensionless. The parameter k can be 0, ± 1 . If $k = +1$, the spatial part of the corresponding s-t is closed (i.e., is *finite* in extent) and *positively curved*, much like the surface of a sphere (although here the surface has three dimensions). If $k = -1$, the space part is *infinite* in extent and *negatively curved*. If $k = 0$, the space part is *infinite in extent and flat*. The FLWR s-t has a *constant spatial curvature* equal to k/r_0^2 . In the FLWR s-t, galaxy clusters are like the proverbial raisins in a loaf of raisin bread that is rising: as the dough expands it causes the raisins to grow increasingly farther apart. In the FLWR s-t, the expanding dough is space itself; the “raisins,” i.e., the galaxies, are going along for the ride.

FLWR kinematics: galactic redshift and energy densities

Note that (1) is a *kinematic* statement: it harbors certain qualitative consequences for how light and matter move about in the universe. Equation (1) does not tell us, however, anything directly about the values of a or k , or where they come from. The situation is a little like the statement, “near the surface of Earth all objects fall along parabolic trajectories.” From this information one can derive certain consequences, for example, that the acceleration each body experiences is independent of position and time. You could also predict how things like times-of-flight and maximum ranges would depend on initial conditions, and from these you could determine g_E if you had enough trajectory data. You could not, however, use that information to determine Newton’s law of gravity—that is, g_E ’s origin. In the same sense, to determine where a and k come from we need dynamics, namely, Einstein’s equations in which correct values of cosmic masses and energies are inserted.

By itself, (1) is useful for describing cosmological kinematics, primarily associated with how light propagates through a universe with a varying scale factor. Matter in the FLWR s-t is continuously distributed, like a fluid or a dust cloud. We can imagine that every point in s-t has a bit of that fluid attached to it. The FLWR s-t deals with enormous distances and times; we can therefore interpret these bits of fluid as galaxies—or better, clusters of galaxies. In the FLWR metric all places in the universe are identical, so all can equally serve as the spatial “origin.” We might as well, therefore, take $r = 0$ to be at the Milky Way (we are at the center of the FLWR universe).

Suppose a crest of a light wave is emitted from a galaxy at coordinate position $r = R$ relative to the Milky Way at time T_e . The next crest is emitted at $T_e + \Delta T_e$. The first crest is detected by us at $r = 0$ at time T_d (“now”); the second reaches us a little later at time $T_d + \Delta T_d$. During the time it takes light to travel to us, the scale factor of the universe changes, so the time between crests at detection is not the same as at emission: $\Delta T_d \neq \Delta T_e$. As a result, there’s a spectral shift due to the change in a . To determine the effect of this, recall that light travels

through s-t with zero elapsed proper time. Set $d\tau$ and the angular terms in (1) equal to zero.

Then $\int_{T_e}^{T_d} \frac{dT}{a(T)} = \int_R^0 \frac{-dr}{\sqrt{1-k(r/r_0)^2}} = \int_{T_e+\Delta T_e}^{T_d+\Delta T_d} \frac{dT}{a(T)}$, because the *coordinate* (not physical) distance

traveled (the middle integral) is the same for both crests. The time between crests of light is much, much shorter than the time for light to go from one galaxy to another, so the last integral

can be approximated by $\int_{T_e+\Delta T_e}^{T_d+\Delta T_d} \frac{dT}{a(T)} \approx \int_{T_e}^{T_d} \frac{dT}{a(T)} + \frac{\Delta T_d}{a(T_d)} - \frac{\Delta T_e}{a(T_e)}$. In other words, $\frac{\Delta T_d}{\Delta T_e} = \frac{a(T_d)}{a(T_e)}$, the

ratio of the period of detected light to that of the emitted light is the same as the ratio of the scale factor at detection to the scale factor at emission. Since wavelength of light is directly

proportional to period we can write $\frac{\lambda_d}{\lambda_e} = \frac{a(T_d)}{a(T_e)}$. More importantly, we can express the galactic

spectral shift as

$$z = \frac{a(T_d)}{a(T_e)} - 1. \quad (2)$$

Equation (2) is a direct result of the kinematics of the FLWR s-t. To evaluate z for any particular signal requires knowing the two a s. On the other hand, we know from Hubble's and others' work that z for all sufficiently distant galaxies is > 0 , so it must be that the ratio of the a s is > 1 . The scale of the universe is apparently greater (at least at this moment in the universe's history) as time goes on. In other words, the FLWR model s-t tells us that positive z implies **cosmic length scale expansion**. In this model, the galaxies are all receding, not because they were all once together at the same point and a mighty explosion hurled them out into space, but because, like the raisins in the rising dough, the distance scale of space is increasing.

Equation (2) leads directly to Hubble's law $z \approx H_0 D/c$, for small values of z . Small z means that $a(T_e)$ can be approximated by the first two terms of a Taylor expansion:

$a(T_e) \approx a(T_d) - \left(\frac{da}{dT}\right)_d \Delta T$, where $\Delta T = T_d - T_e$. Inserting this into (2) produces

$z = \left(\frac{1}{a} \frac{da}{dT}\right)_d \Delta T$. The quantity ΔT is approximately the distance, D , light travels between

emission and detection (at least for nearby galaxies) and the coefficient $\left(\frac{1}{a} \frac{da}{dT}\right)_d = \frac{H_0}{c}$ is the

Hubble constant (divided by c). It is customary to set $a(T_d = \text{now}) = 1$. When that's done

Equation (2) can be expressed as

$$z = \frac{1}{a(T_e)} - 1 \quad \text{or as} \quad a(T_e) = \frac{1}{(1+z)}. \quad (3)$$

The **larger is z , the smaller is a** , and the **older must be the light** we observe for that source now.

Measured values of D , augmenting Cepheid distances with distances inferred from supernovae type Ia, show that Hubble's $z \approx H_0 D/c$ holds pretty well for z up to about 0.1, but that z increases faster than D for larger values. A direct interpretation of this phenomenon is

that a changed from its current behavior to a different one sometime in the past. We will see in a bit what the contemporary understanding of this shift in a was due to.

Before turning to how a is linked to gravitational *dynamics*, it is important to note another purely kinematic consequence of the FLWR s-t. The universe is filled (in the uniform “cosmic” sense) with matter/energy in various forms, all of which contribute to the cosmic gravity. Some of the matter/energy is highly relativistic—e.g., photons, for sure, and possibly also neutrinos. Such entities travel at or near the speed of light and their contribution to the total gravity is associated with their kinetic energies. Some forms of matter/energy are non-relativistic, like atoms and dark matter, for example, and their contribution is fundamentally from rest energy. As the scale factor of the universe changes the relativistic and non-relativistic energy densities change also, but differently. To a very good approximation, the relativistic particles carry so little energy now that they don’t interact with non-relativistic matter anymore. In that event, the number of photons, neutrinos, electrons, and nucleons are all fixed. The energy density in non-relativistic matter is approximately just the rest energy of matter per unit volume. Volume is proportional to the scale factor, a , raised to the three-power. Thus, we can

write the energy density of non-relativistic matter as $\rho_m(T) = \rho_{m0} \left(\frac{1}{a(T)} \right)^3$, where the subscript

“0” means, “value measured at the present time.” The rest energy (if any) of a highly relativistic particle is negligible and its total energy can be expressed in terms of its de Broglie wavelength as hc/λ . As a changes, the wavelength changes also. This leads to a different energy density

dependence, namely, $\rho_r(T) = \rho_{r0} \left(\frac{1}{a(T)} \right)^4$. Though today the density of non-relativistic

matter/energy is estimated to be about 5500 times larger than that of relativistic matter/energy, at some point in the past the two must have been equal. That would have occurred at an a -value of $a_R = \rho_{r0} / \rho_{m0} = 1 / 5500$, provided the number of photons then is the same as today. In fact, the number of photons changed before this a -value and there’s good evidence that a_R is actually more like 1/3200. For times earlier than when $a_R = 1 / 3200$ the universe must have been dominated by relativistic particles – photons, neutrinos, and other particles whose kinetic energies were much greater than their rest energies. That early epoch of the universe is called “radiation dominated.”