Contents

Part I

Global Cosmology

1 Geometry

Describing the global properties of the Universe seems like a formidable problem. In physical sciences, whenever we are faced with difficult problems we ask if we can simplify it. In the case of global cosmology a very important simplification is indicated by observations: homogeneity and isotropy. Deep pictures of the sky are very uniform, in any wavelength, for example X-ray background and Cosmic Microwave Background. Distribution of discrete sources is also uniform across the sky, for example, optical galaxies, Active Galactic Nuclei, Gamma-ray bursts, radio sources. Hubble recession law applies equally well to any direction in the sky (on cosmological scales). Are we at the center of things? We cannot directly prove that that's not the case, but we can learn from past mistakes. Starting from very early on in our history every time we assumed we were at the center we were proven wrong: Earth at the center, Solar system at the center... So we probably should not assume we are at the center. Or anybody else. This is the cosmological principle (CP). In fact, the only type of galaxy recession law that is compatible with global homogeneity and isotropy is a linear law, i.e. Hubble recession law. It will be observed at every point. So the fact that we observe linear recession law supports homogeneity and isotropy.

Why not go a step further and declare that any point in time is equivalent to any other time? This goes by the name of perfect cosmological principle. There is evidence against perfect CP. For example, one of the simpler solutions to the Olbers paradox (darkness of the night sky) is that the Universe has a finite age, or at least that Universe looked a lot different in the past. Further evidence that Universe is not static in time is that the properties of galaxy populations have changed over time. So perfect CP does not represent reality.

A starting point for any geometry is a metric, and ours must be the same everywhere at any given cosmic time, i.e. it must be maximally symmetric. Given a coordinate system the metric describes how small increments along coordinate directions add up to the incremental distance between two nearby points. Metric describing our world must reduce to the Minskowski metric (pseudo-Euclidean geometry) for very small patches, i.e. locally a small patch of the Universe should look flat. In other words, it must be of Reimannian type: squared differential distance between two neighboring points must be expressed as a homogeneous sum of quadratic differentials in the surface coordinates. The metric we will now derive will have that form.

1.1 The metric: spatial part

First, spatial part of the metric. Adopting the cosmological principle allows us to assume that on the largest scales the Universe is very smooth. In fact, let's assume that it is perfectly smooth, which will make the metric easy to formulate. Let us also consider the Universe at one given fixed time. Our 3D space can be curved due to the presence of matter, etc, uniformly distributed throughout the Universe. We erect a Cartesian coordinate system with coordinates x, y, z, w. The first three, x, y, z are coordinates that we can measure in our spatially 3D world; the fourth coordinate w is fictitious, and will be eliminated shortly. Every point in our world satisfies $R^2 = x^2 + y^2 + z^2 + w^2$, where R is radius of curvature, and is constant, but can be real or imaginary.

Squared distance between two neighboring points is $dl^2 = dx^2 + dy^2 + dz^2 + dw^2$, which is the starting point for our metric.

The extra dimension w is fictitious, not observable to us, so we want to get rid of w, using $w^2 = R^2 - x^2 - y^2 - z^2$ and hence $dw = rdr/(R^2 - r^2)^{1/2}$, so then

$$dl^2 = dx^2 + dy^2 + dz^2 + \frac{r^2 dr^2}{R^2 - r^2}\ . \tag{1}$$

Now, instead of Cartesian system let us use the spherical coordinate system which is more appropriate in a world where any direction away from us is as good as any other direction. By analogy with a 1-sphere (circle) and a 2-sphere (balloon surface) let us parameterize our 3-sphere as

$$z = r\cos\theta, \quad x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \tag{2}$$

where the angles θ and ϕ refer to the usual spherical coordinate system centered on us: θ (goes from 0 to π) and is the angle between the positive z axis and the direction we are looking in, and ϕ (goes from 0 to 2π) is the angle measured in the $x - y$ plane, from the positive x-axis. When written out as differentials, i.e $dz = dr\cos\theta - r\sin\theta d\theta$, etc, and substituted for $dx^2 + dy^2 + dz^2$ in eq. 1, the metric becomes

$$dl^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) + \frac{r^2 dr^2}{R^2 - r^2} = \frac{dr^2}{1 - r^2/R^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{3}$$

Quantity $r^2(d\theta^2 + \sin^2\theta d\phi^2)$ is the square length of the line segment along the surface of the sphere centered on us and at coordinate radius r. R is the radius of the 3-sphere; when R is infinitely large dl^2 reduces to the Euclidean flat-space metric. In a general case R^2 is finite and can be positive, zero, or negative. These three cases correspond to positively curved, flat, and negatively curved spatial universe models, respectively.

Let us make another change of variables. Remember that $R^2 = r^2 + w^2$, so we can write $r = R\sin\chi$ and $w = R\cos\chi$ for the positive curvature case, and $r = R\sin(i\chi) = iR\sinh\chi$ and $w = R\cos(i\chi) = R\cosh\chi$ for the negative curvature case. In contrast to ϕ and θ, angle χ is not observable. The metric now becomes,

$$dl^2 = R^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)] \qquad \text{[closed]} \tag{4}$$

$$dl^2 = -R^2[d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta d\phi^2)] \qquad \text{[open]} \tag{5}$$

In the case of a spatially flat Universe we start with eq. 3 and notice that $(1 - r^2/R^2)$ is 1. Setting $\chi = r$, we get

$$dl^2 = [d\chi^2 + \chi^2(d\theta^2 + \sin^2\theta d\phi^2)] \qquad \text{[flat]} \tag{6}$$

In eq. 3 one sometimes uses $k = 1/R^2$, where k is called the curvature, or the curvature index:

$$dl^2 = \frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{7}$$

Eqs. 3-6 are maximally symmetric metrics in the sense that all spatial dimensions are treated the same, and there is no special dimension or direction, in accordance with the cosmological principle. This metric often goes by the name of Friedman-Robertson-Walker, or FRW (or sometimes FLRW, adding the name of Lemaitre).

1.2 The full metric

Next the full metric, i.e. the expression for an infinitesimal distance between two events in space-time, which for the negatively curved model is

$$d\tau^2 = -c^2dt^2 + dl^2 = -c^2dt^2 + a(t)^2|R^2|[d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta d\phi^2)]. \tag{8}$$

Notice that scale factor $a(t)$ now multiplies dl. This is done because with the hindsight of the observed Hubble recession law we want to allow for the possibility of Universe that undergoes global expansion/contraction. This scale factor of the Universe will be discussed in detail a little later. Eq. 8 applies to a negative spatial curvature Universe, but flat and closed cases can be written down too, using appropriate dl^2 from the previous section. Time t is the cosmic time, as measured by fundamental observers sprinkled throughout the Universe, and whose identical clocks were synchronized at a sufficiently early epoch, say, the Big Bang.

There are two different ways of treating separations between objects in cosmology. One is comoving, the other is physical. Comoving, or coordinate separation between points is given by the spatial part of the metric without the scale factor. For points that stay put and do not move under their own power and are not being pushed by some external force, comoving coordinates will stay the same. The actual physical, or proper separation will change, increase, as the Universe expands. On the other hand, if there is a physical object that is held together by electrostatic forces or is strongly gravitationally bound (like in galaxy), then its physical/proper size will stay the same as the Universe expands, but its comoving size will constantly shrink.

Comoving distances are usually calibrated such that at the present time the scale factor $a(t_0) = a_0$ is set to 1; t_0 usually represents present time. It is common in cosmology to quote results in comoving coordinates; so if you have a length Δl_{phys} at some time t other than the present, then its comoving length is expressed by $\Delta l_{phys} = a(t) \cdot \Delta l_{comov}$. This rescaling of distances is done to 'take out' the universal expansion; it is similar to taking out the inflation in prices of things. For example, even though a can of Coke used to sell for 5 cents in the 1930's (its 'proper', or 'physical' price), economists would say its cost was 50 cents in today's dollars, which would correspond to its 'comoving', or 'coordinate' price. (I might have gotten the actual prices wrong...)

2 Dynamics

In this section we will discuss the properties and time behavior of the scale factor $a(t)$.

2.1 Relativistic dynamics: Einstein's field equations

Having derived the metric for the smooth Universe, we know the metric elements, the $g_{\mu\nu}$'s, and hence can write down the Christoffel symbols, the Riemann curvature tensor elements, the Ricci tensor and the Ricci scalar. The last two combine to give us the Einstein tensor, which is the LHS of the field equations. To the left hand side Einstein added another term, $g_{\mu\nu}\Lambda$, an 'integration constant', which is allowed because the (covariant) derivative of the metric tensor is zero, so any multiple of it is allowed. The RHS is the stress-energy tensor, whose diagonal elements are $(\rho c^2, p, p, p)$. Thus, the Einstein field equations are four equations: the 00, i.e. the time-time component is,

$$3\frac{\dot{a}^2}{a^2} + 3\frac{kc^2}{a^2} - \Lambda c^2 = 8\pi G\rho \tag{9}$$

This is known as the Friedman equation, and it governs the expansion rate of the Universe. The expansion rate is determined by the contributions from curvature (k-term), cosmological constant

(Λ-term), and matter (ρ-term). The three spatial components, which are all identical by symmetry dictated by the cosmological principle, are:

$$-2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{kc^2}{a^2} + \Lambda c^2 = 8\pi G p/c^2 \tag{10}$$

Subtracting eq. 9 from eq. 10, and then dividing the resulting equation by (-2), we get

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p/c^2) + \frac{\Lambda c^2}{3} \tag{11}$$

which is an equation of motion and looks very similar to its Newtonian analog (see § 2.3). Another algebraic manipulation of eqs. 9 and 10 results in the continuity equation: multiply eq. 9 by a^3, then differentiate this new equation w.r.t. t, then divide by $3a^2\dot{a}$. Now what you have as the LHS of your last equation is minus the LHS of eq. 10. Next, equate the RHS's of the two equations, and with some further manipulation you will get:

$$\dot{\rho} + 3(\dot{a}/a)(\rho + p) = 0. \tag{12}$$

where we have assumed $c = 1$.

To sum up, the field equations convey the physical content of the three key classical equations: continuity, Poisson and equation of motion. The job of these equations is to tell us how a system evolves in time, given its constituents. That's exactly what the equations in this subsection tell us, how the scale radius of the Universe evolves given its matter contents. Note that only two of the above equations are independent, for example, eq. 9 and eq. 10. But the equations that are usually used in cosmology are some combination of eq. 9, 11, and 12.

Finally, note that the curvature does not contribute to the acceleration equation (of motion), eq 11, and that the total 'gravitating density', $(\rho + 3p)$, now also includes (relativistic) pressure. Cosmological constant Λ is a constant contribution to $\ddot{a}/a$ and $\dot{a}/a$ equations.

2.2 Thermodynamics of the expanding fluid

Here we will discuss eq. 12.

The material inside the sphere is assumed to be free of viscosity, conductivity, etc. When the material has negligible relativistic pressure, it is called 'dust' in cosmology. For example, Hydrogen gas is considered dust because its pressure due to kinetic motions of atoms is negligible compared to its mass energy (recall mc^2). The net energy within the sphere is $U = \rho\, V$, where ρ and V are total energy density and volume. As our sphere expands with the Universal expansion, the material inside does work, $dW = p\, dV$. Since there cannot be any net influx or outflow of heat into or from the sphere (if there was, that would violate homogeneity and isotropy assumptions), and there are no sources/sinks of heat, the first law of thermodynamics states that $dU + p\, dV = 0$. This says that the work done by pressure, $p\, dV$, happens at the expense of the internal energy, which decreases by $dU = \rho\, dV + d\rho\, V$. If the work done is positive, then the total internal energy decreases, but if pressure work done is negative—as in the case of the cosmological constant—then there will be an increase in the total internal energy. Assuming that the change in volume happens in a short time interval dt, we have

$$\rho\, \dot{V} + V\, \dot{\rho} + p\, \dot{V} = 0. \tag{13}$$

Using $V \propto a^3$, and so $dV/V = 3\,(da/a)$, we get

$$\dot{\rho} = -3(\rho + p)(\dot{a}/a) = -3H(\rho + p), \tag{14}$$

where $H = \dot{a}/a$ is the Hubble parameter. This is the continuity, or the conservation of mass-energy equation. It can be compared to the continuity equation that we derived for pressure-less fluid, $\dot{\rho} = -\vec{\nabla} \cdot (\rho\ \vec{v})$. For a uniform density fluid this reduces to $\dot{\rho} = -\rho \vec{\nabla} \cdot \vec{v} = -3\rho(\dot{a}/a)$, because the term $\vec{\nabla}\rho$ is zero since there are no density gradients at $t =$const in a universe that obeys the cosmological principle. So eq. 14 is equivalent to the continuity eq., except the former also includes pressure. Eq. 14 says that in a small cube of fixed proper, or physical size, the total mass-energy density will be changing in time, and there are two contributing reasons for the change: the term $3H\rho$ represents the reduction in energy density due to the Hubble expansion, and $3Hp$ term is due to the thermodynamic work done by the pressure during expansion.

Let us assume that the equation of state of the relativistic fluid is simple, $p = w\rho$, i.e. sound speed in the fluid is $c_s = \sqrt{dp/d\rho} = w$. Substituting into eq. 14 we get,

$$\frac{\dot{\rho}}{\rho} = -3\frac{\dot{a}}{a}(1 + w), \tag{15}$$

whose solution is

$$\rho = \rho_0 a^{-3(1+w)}, \tag{16}$$

with ρ_0 being the present day density of the fluid component whose equation of state is w. The Universe can contain several different fluid components, each specified by its own equation of state:
- ordinary matter ('dust'): $w = 0$, $\rho \propto a^{-3}$;
- radiation: $w = 1/3$, $\rho \propto a^{-4}$;
- cosmological constant: $w = -1$, $\rho =$ const;
- 'quintessence', etc.: other values of w, possibly $w(t)$.

2.3 Newtonian dynamics

Non-relativistic versions of eqs. 11 and 12 can be derived as follows.

Consider a small spherically symmetric volume some place in the Universe and a test particle sitting on the surface of its bounding sphere. According to the cosmological principle this region is equivalent to any other region. According to one of the Newton's theorems the outside mass does not contribute anything to the force on a particle inside the sphere (the contributed potential is constant inside, so its gradient is zero), and only the interior mass matters determines the motion of the particle with respect to the center of the sphere. According to Birkoff's theorem this Newtonian result can be extend to the relativistic (cosmological) case, hence we can proceed with the Newton's 2nd law and write down, $\ddot{a} = -GM/a^2$, where a is the distance from the center of the sphere to its bounding surface. Notice that we have used a above, same as the scale factor we had at the end of the previous section. In equations that we eventually end up with, i.e. those that can be compared to observations, the normalization factor, $a_0 = a(t_{\rm now})$ drops out. Therefore, the actual value of a is irrelevant, only its scaling with time is important.

Assume a uniform density ρ inside the sphere, then equation of motion can be written as $\ddot{a} = -(4G\pi/3)\rho a$. Multiply this by $2\dot{a}$ to get a perfect differential, which we integrate to get the Friedman equation:

$$(\dot{a})^2 = \frac{8\pi G\rho_0}{3} \cdot \frac{1}{a} + k'. \tag{17}$$

This looks like a classical energy equation, with the kinetic and potential terms, and the constant of integration, k', is just the total energy in the Newtonian interpretation. Having already seen the corresponding GR equation, we know that $k' = -kc^2 = -c^2/R^2$, where k was a part of eqs. 7, 9 and 10.

The interpretation of $\dot{a}$ in eq. 17 is as follows.

$\star$ If $k' < 0$ ($k > 0$), then $\dot{a}^2$ reaches 0 at finite a, so the Universe stops expanding at some finite time from now, and will begin to recollapse. This would correspond to a closed, or positively curved Universe, with $R^2 > 0$.

$\star$ If $k' > 0$ ($k < 0$), then $\dot{a}^2$ never becomes 0, and so the Universe continues to expand forever, with a becoming arbitrarily large. This is a negatively curved, or open Universe, with $R^2 < 0$.

Thus curvature/geometry and dynamical evolution are linked in this case. However, in full GR, eq. 9 and 10 also contain Λ, and if this is non-zero, the relation between the sign of curvature k and the rate of change of expansion rate is not as simple.

2.4 A different form for the Friedman equation

Let us rewrite eq. 9 as follows,

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 = \frac{8\pi G}{3}\rho_0 a^{-3(1+w)} - \frac{kc^2}{a^2} \tag{18}$$

where we have used eq. 16. Depending on whether $w = 0$, or $w = -1$, the density term will become $8\pi G\rho/3$, or $\Lambda c^2/3$, respectively, if we define $\Lambda = 8\pi G\rho_\Lambda$. We see that the density term can represent any fluid component, depending on w, or a sum of them.

Eq. 18 also defines the Hubble parameter, and tells us how it evolves with a, depending on what types of mass-energy w components make up the Universe. Eq. 18 is called the Friedman equation, or energy equation, and is one of the main equations in cosmology.

The contribution of different components in the Universe is usually expressed in terms of Ω_w's. Consider a fiducial model with only the ordinary matter; $\Lambda = 0$, and $k = 0$. This is called Einstein-de Sitter model. Then, at the present epoch, $t = t_0$ eq. 18 reduces to $H_0^2 = 8\pi G\rho_0/3$. We use this to define the critical mass density, or energy density if you multiply it by c^2:

$$\rho_{crit} = \frac{3H_0^2}{8\pi G} = 1.88 \times 10^{-29}\ h^2\ \mathrm{gm/cm^3} \tag{19}$$

where the dimensionless h is H_0 expressed in units of 100 km/sec/Mpc, and $H_0^{-1} = 9.78h^{-1}$ Gyr. From observations, values of h hover around 0.7±0.2. With this, the contribution of each component is described in units of ρ_{crit}:

$$\Omega_m = \frac{\rho_m(t = t_0)}{\rho_{crit}}, \quad \Omega_\Lambda = \frac{\rho_\Lambda(t = t_0)}{\rho_{crit}}, \quad \text{etc.} \tag{20}$$

For the today's Universe, $t = t_0$ we rewrite eq. 18:

$$H_0^2 = \frac{8\pi G}{3}\sum_w \rho_{0,w}\ a_0^{-3(1+w)} - \frac{kc^2}{a_0^2} = H_0^2 \sum_w \Omega_w - \frac{kc^2}{a_0^2}, \tag{21}$$

where the first term is a sum over all the matter/radiation/cosmological constant, etc. components and the second is the curvature term. Dividing eq. 21 by H_0^2,

$$1 = \frac{\rho_{0,m}}{\rho_{crit}} + \frac{\rho_{0,r}}{\rho_{crit}} + \frac{\rho_{0,\Lambda}}{\rho_{crit}} - \frac{k}{a_0^2\ H_0^2} = \Omega_m + \Omega_r + \Omega_\Lambda + (1 - \sum_w \Omega_w) \tag{22}$$

$$(1 - \sum_w \Omega_w) = -\frac{kc^2}{a_0^2\ H_0^2}, \quad \text{or} \quad R^2 = -\Big[(1 - \sum_w \Omega_w)\ \frac{a_0^2\ H_0^2}{c^2}\Big]^{-1} \tag{23}$$

The Ω's in eq. 22 are present day values, and so are constant (not a function of time) for a given cosmological model. Some authors define Ω's at some time other than the present, in which case it is customary to state so. Eq. 22 and 23 are different ways of writing the Friedman equation.

Using the above, the Friedman equation can also be rewritten as

$$H = H_0 \Big[\sum_w \Omega_w \Big(\frac{a}{a_0}\Big)^{-3(1+w)} + (1 - \sum_w \Omega_w)\Big(\frac{a}{a_0}\Big)^{-2} \Big]^{1/2} = H_0 \, E(z, \Omega_w). \tag{24}$$

and a_0 is taken to be 1 today. We have introduced parameter E, which encapsulates the evolution of the Hubble parameter. We've written that E is a function of z because of this monotonic relation: $a = (1+z)^{-1}$, which we will address in §2.5.

As cosmic time progresses the energy density of radiation drops off fastest, $\rho \propto a^{-4}$, while that of Λ the slowest, $\rho \propto a^0$ =const. So the early Universe was radiation dominated. Matter, and even more so, cosmological constant were insignificant. Towards the end of times, Λ, if it exists, will dominate the dynamics. If cosmological constant is zero then curvature will dominate.

Consider a simplified case where a single component dominates the Universe. Then $(\dot{a}/a)^2 \propto a^{-3(1+w)}$. The solution to this is $a \propto t^{2/[3(1+w)]}$. So for ordinary matter, $a \propto t^{2/3}$, and for radiation, $a \propto t^{1/2}$. The Universe expands slower in the latter case because of the contribution of pressure to the gravitating "mass". Also note that the Universe's expansion will be accelerating (i.e. $\ddot{a} > 0$) if $w < -1/3$.

The equation of state for cosmological constant is $p = -\rho$ ($w = -1$), and $\dot{\rho} = 0$, so the density stays constant as the Universe expands. The solution to the Friedman equation when cosmological constant dominates the energy budget is, $a = e^{H_\Lambda t}$ with $H_\Lambda = (\Lambda/3)^{1/2}$; this is called the de Sitter solution. This is the same cosmological constant that was added to the Einstein field equations, as the $\Lambda g_{\mu\nu}$ term. One of the possible physical interpretations of the cosmological constant is vacuum energy density, even though a simple (simple-minded?) evaluation of what the vacuum energy density should be gives a value of $\rho_\Lambda \sim 10^{120} \rho_{crit}$, which is obviously wrong!

2.5 Redshift and the scale factor

Imagine an emitter and an observer, both at $\phi = 0$ and $\theta = 0$, because we can draw our coordinate system that way, and at r_e and r_o respectively. From the geodesic equation of motion for photons, $d\tau^2 = 0$, we have,

$$\int_{t_e}^{t_o} \frac{dt}{a(t)} = \int_{r_e}^{r_o} \frac{dr}{\sqrt{1 - r^2/R^2}}, \tag{25}$$

where a given crest of a wave is emitted at r_e at t_e and absorbed (or detected) at r_0 at t_0. The next wave crest satisfies,

$$\int_{t_e+\delta t_e}^{t_o+\delta t_o} \frac{dt}{a(t)} = \int_{r_e}^{r_o} \frac{dr}{\sqrt{1 - r^2/R^2}} \tag{26}$$

Both the RHS are the same because the emitter and observer have not moved from their initial comoving coordinates. Therefore we can equate the LHS of the above equations. We also assume that the scale factor did not change between t_e and $t_e + \delta t_e$, and stayed at $a(t_e)$. And also, the scale factor stayed the same between t_o and $t_o + \delta t_o$, and was $a(t_o)$. In other words, we assume that the separation between emitter and observer is cosmological, whereas the wavelength of radiation is tiny compared to cosmological scales. So,

$$\int_{t_e+\delta t_e}^{t_o+\delta t_o} \frac{dt}{a(t)} = \int_{t_e}^{t_o} \frac{dt}{a(t)} + \frac{\delta t_o}{a(t_o)} - \frac{\delta t_e}{a(t_e)} = \int_{t_e}^{t_o} \frac{dt}{a(t)} \tag{27}$$

Therefore,

$$\frac{\delta t_o}{a(t_o)} = \frac{\delta t_e}{a(t_e)}, \quad \text{or} \quad \frac{\lambda_o}{a(t_o)} = \frac{\lambda_e}{a(t_e)} \tag{28}$$

The result depends only on the ratio of scale factors at the time of emission and absorption/detection.

$$\frac{\lambda_o}{\lambda_e} = \frac{a(t_o)}{a(t_e)} = 1 + z_e, \quad \text{or,} \quad a = (1+z)^{-1}. \tag{29}$$

This defines the redshift of the emitting object as seen by the observer, and because of the ease of obtaining λ_o/λ_e from spectral lines of distant emitting objects, redshift is widely used in cosmology. Furthermore, redshift is a measure of cosmological distance that is completely assumption-free; it does not depend on the mass-energy content of the Universe, or the rate expansion of the Universe. Converting redshift to, say proper distance, or angular diameter distance, relies on the specifics of a cosmological model.

From the preceding derivation we see that, strictly speaking, a measured redshift tells us how big the scale factor of the Universe was at the time the radiation was emitted compared to how big it is today. Redshift does not say that the object whose z you measured is moving away from us at some velocity. That contribution to the redshift, i.e. the proper motion, would be in addition to the cosmological redshift, and can be shown to be negligible for objects at z above 0.1 or so. There is no sense in interpreting cosmological redshift as velocity of recession.

2.6 Nearby cosmology

In the nearby Universe, $z \ll 1$, relations between redshift, time, distance, etc. simplify considerably. Take, for example, the scale factor. For nearby objects (cosmologically speaking) it can be expanded in a Taylor series around the present value,

$$a = a_0 + \dot{a}(t - t_0) + \frac{1}{2}\ddot{a}(t - t_0)^2 + ... \tag{30}$$

Remembering that H_0 is the present day value of $\dot{a}/a$ we have,

$$a(t) \approx a_0\Big[1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2 (t - t_0)^2 + ...\Big], \tag{31}$$

where q_0 is the dimensionless deceleration parameter and is given by

$$q_0 = -\frac{\ddot{a}(t_0)a(t_0)}{\dot{a}(t_0)^2}, \tag{32}$$

where all the quantities are evaluated at the present time. q_0 can be expressed in terms of Ω_m. If ordinary matter is all there is, we can use the expression for $\ddot{a} = -(4\pi G/3)\rho\, a$ and the definition of ρ_{crit} and Ω_m to derive: $q_0 = \Omega_m/2$.

Since the redshift is simply related to the scale factor, $1 + z = a_0/a$, the redshift of nearby objects is

$$z \approx -H_0(t - t_0) + ... = H_0 \Delta t = H_0(d/c) = v/c, \tag{33}$$

where Δt is the time it would take a light signal to travel from some nearby galaxy to us. For such a galaxy Δt is the ratio of its distance to c. The last two expressions are often used to convert between distance, redshift, and “velocity of recession” of nearby galaxies. Since z is related to the fractional change in observed wavelength of radiation, z can be wrongly interpreted (using classical Doppler) as the velocity of recession of galaxies.

3 The metric, Friedman equation, and the observable world

3.1 Expansion age of the Universe

A proper time interval dt at redshift z (as measured by an observer with a clock, at that redshift) is related to the redshift interval dz (as measured by us at redshift 0) is given by

$$dt = \frac{c}{H}\frac{da}{a} = -\frac{1}{H_0}\cdot\frac{dz}{1+z}\cdot\frac{1}{E(z)} = -\frac{1}{H_0}\cdot\frac{dz}{1+z}\Big[\sum_w \Omega_m(1+z)^{3(1+w)}+[1-\sum_w \Omega_w](1+z)^2\Big]^{-1/2} \quad (34)$$

where we used the definition of $E(z)$ from eq. 24. For a flat $\Omega_m = 1$ Einstein-de Sitter model the above reduces to a simple form: $dt/dz = -{H_0}^{-1}(1+z)^{-5/2}$, which can be integrated to obtain the expansion age of the Universe,

$$t = -\int_{\infty}^{0} {H_0}^{-1}(1+z)^{-5/2}dz = (2/3)H_0^{-1}(1+z)^{-3/2}|_{\infty}^{0} = (2/3)H_0^{-1}, \quad (35)$$

where we assumed that the earliest epoch of the Universe has $z = \infty$. Eq. 35 is an important result. It says that if you can measure the current value of H_0 you can calculate the expansion age of the Universe, provided we live in a flat $\Omega_m = 1$ Universe. But even in a more general case the knowledge of H_0 can be translated in to the age, using eq. 34. Notice that in general the age is smaller than H_0^{-1} because the expansion rate H was higher in the past, and the presence of matter decelerates the expansion.

The last expression can be converted to one in terms of density, using $\rho_{crit} = 3H_0^2/8\pi G$: $t_0 = (6\pi G\rho_{crit})^{-1/2}$. With appropriate limits, eq. 35 can be used to obtain age of object at any z.

Eq. 35 and similar expressions for other Universe models can be used as tests of global cosmology. Consider a galaxy whose redshift z we have measured. We want to know how much time has passed from the Big Bang until the light that we see from the galaxy was emitted by that galaxy at z. This is the age of the Universe at that time (at z), and so also the maximum possible age of that galaxy in a given cosmology. This suggests a cosmological test. Suppose the cosmologically estimated age at z is t_c. Also suppose you can estimate the age of that galaxy by entirely different means, for example by studying the ages of stars in the galaxy, or considering some other chemical evolution process; this age is t_s. The latter age, t_s, better be less than t_c, otherwise your Universe model is in trouble. This type of argument is used to place constraints on global cosmological parameters using very nearby and very distant objects: globular clusters and high redshift galaxies. Results strongly rule out Einstein-de Sitter Universe: its age is just too short!

3.2 Angular diameter distance

Unlike the situation in Euclidean geometry, in cosmology we have a different distance for every application—there is angular diameter distance, luminosity distance, parallax distance, proper motion distance, etc. These differ from each other because the intrinsic spatial geometry need not be flat, and because the Universe scale factor evolves in time.

Consider a galaxy at some z: we measure a certain angular diameter, $d\Theta$ and would like to know what physical size that corresponds to in the galaxy itself.

Using the FRW metric we express the transverse physical (proper) size of a galaxy by realizing that $d\chi$ can be set to 0 because the galaxy is oriented perpendicular to our line of sight, and dt can be set to 0 because all of the galaxy is observed at the same time:

$$dl_{gal} = a(t)|R^2|^{1/2}\sin\chi(d\theta^2 + \sin^2\theta d\phi^2)^{1/2} = a(t)|R^2|^{1/2}\sin\chi d\Theta, \quad (36)$$

and sinh would be used in the open case. $d\Theta$ is the angular size of the galaxy as seen by us. The latter is determined at the time that the light (that will eventually reach us) is emitted from the galaxy. This angle does not change as the light rays from the two ends of the galaxy travel towards us because the angle between two null geodesics stays the same as the Universe expands.

In Euclidean geometry the angular size, physical size and distance are related by the relation size=distance×$d\Theta$. In cosmology we would like to preserve this relation. The price we pay is that the notion of distance has to be redefined; we call it angular diameter distance, D_A, and it is given by

$$D_A = a(t)|R^2|^{1/2}\sin\chi; \quad a(t)|R^2|^{1/2}\sinh\chi; \quad a(t)\chi \tag{37}$$

for spatially closed, open and flat models respectively. t is the cosmic time when light was emitted from the galaxy. R can be eliminated using eq. 23 (setting $a_0 = 1$),

$$|R^2|^{1/2} = (\frac{H_0}{c}|1-\sum_w \Omega_w|^{1/2})^{-1}, \tag{38}$$

Using the absolute value of $(1-\sum_w \Omega_w)$ makes this expression applicable in the case of positively and negatively curved spatial geometries.

Looking at eq. 37 we see that we now need to evaluate χ. Using the expression for the metric, eq. 8, we recall that photons travel along null geodesics: $d\tau^2 = 0$, and so $dt/a(t) = |R^2|^{1/2}d\chi$

$$d\chi = \frac{cdt}{a|R^2|^{1/2}} = \frac{cda}{|R^2|^{1/2}Ha^2} = \frac{c}{H_0}\frac{1}{|R^2|^{1/2}E(z)}dz = \frac{dz}{E(z)}|1-\sum_w \Omega_w|^{1/2} \tag{39}$$

So angular diameter distance for the case of a negatively curved Universe becomes,

$$\begin{aligned} D_A(z) = \frac{c}{H_0}\frac{1}{(1+z)|1-\sum_w \Omega_w|^{1/2}} \times \\ \sinh\left[|1-\sum_w \Omega_w|^{1/2}\int_0^z \frac{dz}{[\sum_w \Omega_w(1+z)^{3(1+w)} + [1-\sum_w \Omega_w](1+z)^2]^{1/2}}.\right] \end{aligned} \tag{40}$$

Popular cosmological models these days have $\Omega_m + \Omega_\Lambda = 1$, i.e. curvature is zero (and the contribution of radiation is negligible at the present cosmic epoch). For these models angular diameter distance reduces to a simpler expression,

$$D_A(z) = \frac{c}{H_0}\frac{1}{(1+z)}\int_0^z \frac{dz}{[\Omega_m(1+z)^3 + \Omega_\Lambda]^{1/2}} \tag{41}$$

Recall that in a negatively curved non-evolving universe the angular diameter distance increases monotonically for objects further away. In a positively curved non-evolving universe the angular diameter distance has a maximum at some finite coordinate distance away from the observer. In an *evolving* Universe, some models with flat or negatively curved spatial geometry have angular diameter distances that attain a maximum. This is because the Universe was smaller in the past when the light from the objects started out on its way to us. Back then, objects of same proper size occupied a larger coordinate size, and so would have subtended a bigger angle. For example, for $\Omega_m = 1$, the angular diameter distance reaches a maximum at $z \sim 1.0$ i.e. a galaxy of a fixed physical/proper size will appear to get bigger in the sky if it is moved from $z \sim 1$ to higher redshifts.

3.3 Luminosity distance

Now we will relate a galaxy's intrinsic luminosity L (energy emitted per unit time) to the flux f (energy per unit time per unit area of the detector). For now let us consider bolometric luminosities, i.e. we assume that we collect radiation of *all* wavelengths. Luminosity from the source gets spread over a sphere of proper area $A = 4\pi r^2 = 4\pi\ |R^2| \sinh^2 \chi$. When we receive the radiation we are sitting on the surface of this sphere, and the scale factor $a_0 = 1$. The energy of each photon is reduced by a factor of $a_0/a(t)$ (wavelength is increased, i.e. photons are cosmologically redshifted), and reduced by another factor of $a_0/a(t)$ because the photons arrive at the observer at a slower rate (cosmological time dilation). So the luminosity distance D_L is given by

$$f = \frac{L}{4\pi\ |R^2| \sinh^2 \chi\ (1+z)^2} = \frac{L}{4\pi D_L^2}, \quad \text{where} \quad D_L = |R^2|^{1/2} \sinh \chi\ (1+z) = D_A (1+z)^2. \quad (42)$$

Here again we insisted on defining the relation between intrinsic luminosity and received flux of an object the same way we define it in Euclidean geometry, i.e. $f = L/4\pi D_L^2$. This forced us to redefine D_L. In the equation above we have assumed a geometrically negatively curved model. D_L and D_A are different, but are related through the proper emitter-observer distance at the time of the observation, $D_P = |R^2|^{1/2} \sinh \chi$.

3.4 Surface brightness

SB is defined as the energy received per unit time per unit solid angle per unit area of the detector, therefore it is flux per unit solid angle. Total proper surface area of a galaxy (perpendicular to our line of sight) is related to the solid angle it subtends at the detector and the angular diameter distance to it: $\Delta A = D_A^2 \Delta\Theta$.

$$\mathrm{SB} = \frac{f}{\Delta\Theta} = \frac{L}{4\pi D_L^2 \Delta\Theta} = \frac{L}{4\pi D_A^2 (1+z)^4} \frac{D_A^2}{\Delta A} = \frac{L}{4\pi (1+z)^4 \Delta A} = \frac{\mathrm{SB}_{\mathrm{intrinsic}}}{(1+z)^4} \quad (43)$$

where $\mathrm{SB}_{\mathrm{intrinsic}}$ is the intrinsic surface brightness of the galaxy that you would measure if you were sitting very close to it. Thus the surface brightness of an object decreases as a function of redshift only; there is no dependence on any cosmological parameters! The relation was first proposed as a test of FRW-type cosmologies by Tolman.

3.5 Redshift dependence of non-bolometric luminosities: K-corrections

Let us now consider a more realistic case of how luminosities of galaxies are measured: Instead of measuring bolometric luminosities, we can only measure luminosity through a given filter, which filters out all but some wavelengths. If observations of a galaxy at z are made using a spectral filter centered on frequency ν with a width $d\nu$ as measured in our frame, then the radiation received came from a frequency range $(1+z)d\nu$ centered on $(1+z)\nu$. The flux is then

$$f d\nu = \frac{[1+z] d\nu L([1+z]\nu)}{4\pi D_L^2} = \frac{L(\nu)}{4\pi D_L^2} \left\{ \frac{[1+z] L([1+z]\nu)}{L(\nu)} \right\}. \quad (44)$$

The term in curly brackets is the so-called K-correction. It is customary to express it in terms of magnitudes, i.e. as $-2.5 \log\{....\}$. In general, the intrinsic shape of galaxies' spectral energy distribution is such that K-corrections work to make the galaxy appear even fainter. However, in some cases, specifically galaxies that contain a large amount of warm dust, the K-corrections in the far-IR, micron and sub-mm parts of the spectrum make galaxies brighter than they would appear otherwise.

3.6 Volume element

Suppose you used a telescope to obtain an image of a small portion of the sky. Say the limiting magnitude of your survey is such that it allows you to see a typical galaxy up to a maximum redshift z_f. So you would like to know what volume of the Universe you have captured. Obviously, the region you have imaged represents the same angular size at any redshift. At any given redshift the angular size $d\Theta$ of your observed field translates into a transverse distance of $D_A d\Theta$. If the field is square its area is $[D_A d\Theta]^2$, so the total volume is the volume element integrated form 0 to z_f:

$$V\left(< z_f\right) = \left(\frac{c}{H_0}\right)^2 d\Theta^2 \int_0^{z_f} D_A(z)^2 \, \frac{cdt}{dz} dz. \tag{45}$$

Because geometry enters through $D_A(z)$ and cdt/dz the volume upto z is a function of global cosmological parameters. Between any two redshifts the volume is generally larger for k-dominated or Λ-dominated models compared to the Einstein-de Sitter case. In principle, this dependence on cosmological parameters can be exploited in volume tests of the universe models.

4 Cosmological tests

4.1 Tests of global geometry

At the beginning of this series of notes we started out with an assumption of homogeneity and isotropy for the Universe (the cosmological principle), constructed the corresponding maximally symmetric metric, added it to the relativistic field equations of gravity, and ended up with equations for the dynamical evolution of the Universe. These equations have only a handful of parameters, namely, the various Ω_w's, and H_0. We have then derived equations that tell us how to relate the observable information about cosmic objects, such as angular sizes and luminosities of galaxies, to those few cosmological parameters. This allows us to test various models—i.e. various sets of cosmological parameters—through observations.

The simplest (in principle!) standard tests are divided into two classes, those that use standard candles and standard rulers, respectively. For each type of test we seek an astronomical object that retains its intrinsic luminosity or physical size regardless of its redshift. Since objects in the Universe are known to evolve in time, this is a tall order. Yet a few classes of astronomical objects have been shown to meet the requirement.

With standard candles one basically measures the luminosity distance to objects at different redshifts. Of the cosmologically distant sources, supernova Type Ia are a good example. Brightest cluster galaxies (BCG) and Gamma-Ray bursts are also sometimes considered to have a relatively constant intrinsic luminosity.

Standard rulers are harder to find. Examples of standard rulers are radio galaxies. Faranov-Riley Class II objects have radio lobes with bright spots (edges) at the ends of the lobes; the distance between these stays resolvable at cosmological distances. It is still not entirely clear whether these objects are really standard rulers or not.

Volume tests are not as reliable. The Universe is sprinkled with galaxies and QSOs. If these were randomly distributed and did not evolve in time, then we could easily determine the volume of a given segment of the Universe, between, say, z_1 and z_2, with an angular size of θ on the side, by simply counting the galaxies. Of course such an idealized situation does not exist because galaxies and QSOs evolve in time, so the simplest version on the volume test using these objects cannot work. Downgrading our expectations a bit, we can place limits on the amount of evolution, and hence, maybe rule out certain cosmological models.

Let us illustrate this with an example. Suppose we have an observed sample of galaxies and their redshifts. A simple theoretical model can be described with (1) cosmological model, i.e. Ω_w's and H_0;(2) local galaxy luminosity function, i.e. the distribution of galaxies in luminosity; (3) evolution of galaxy luminosity function as a function of luminosity, or mass; (4) spectral energy distribution of galaxies as a function of redshift. Since galaxies consist of stars, we know that there will be a certain amount of 'passive evolution', i.e. just due to stars aging. There can be other evolution on top of that, for example, ongoing star formation, or star bursts, galaxy mergers, etc. These effects can be parameterized. Models thus constructed can be compared against observations.

Note that the global geometry of the Universe can not be discerned in our cosmological neighborhood, at redshifts lower than 1 or so because the scale factor a has not changed much since $z \sim 1$. So any standard ruler/candle we have in this regime will only be able to tell us about the Hubble parameter, which is nearly constant in our cosmic neighborhood.

4.2 Other tests

There is a host of other tests for global cosmological parameters that do no not attempt to measure global geometry, but rather look at the properties of the objects that populate the Universe, such as galaxies and clusters of galaxies. These methods rely on much more than just that the simple geometrical relations we discussed in this section. These include, for example, rate of growth of structure, like galaxy clusters, gravitational lensing, and the microwave background, etc.

4.3 The Hubble parameter

Aside from the astronomical distance ladder there are three other ways of measuring H_0 all applied over cosmological distances: (1) fluctuation spectrum of the Cosmic Microwave Background, which we will discuss later, (2) test using galaxy clusters, which requires observations of Sunyev-Zeldovich temperature decrement and X-ray surface brightness from the same galaxy cluster, (3) time delay measurements of different images of the same gravitationally lensed multiply-imaged QSO. All three methods rely on different physics, and have their own drawbacks and biases, but amazingly enough, results agree to better than 20%, and give $h = 0.7 \pm 0.07$.