Contents

3 The Friedmann–Lemaître–Robertson–Walker model

3.1 Geometry

3.1.1 The Robertson–Walker metric

In cosmology the basic approximation is that spacetime has spatial hypersurfaces that are exactly *homogeneous* and *isotropic*. The coordinate t that is constant on these hypersurfaces, and labels them, is called the *cosmic time*.

There is good evidence that the universe is indeed *statistically* homogeneous (all places look the same) and isotropic (all directions look the same) on scales larger than about 100 Mpc. In particular, the CMB looks highly isotropic to us. If we accept the *Copernican principle* according to which we are not in a special location, typical observers should all see an almost isotropic CMB. Somewhat surprisingly, this does *not* prove that the universe would be almost homogeneous, though it does lend it support [1, 2]. Observations of the distribution of galaxies do give strong support for statistical homogeneity [3].

These observations of statistical homogeneity and isotropy do not prove that the universe would be well described by a model that is *exactly* homogeneous and isotropic, but it does motivate using it as a first approximation. (We will see that the approximation is quite good, and at early times it is excellent, as the universe was then more homogeneous and isotropic than today.) Statistical homogeneity and isotropy is called the *Cosmological Principle*. Sometimes this term is used to describe exact homogeneity and isotropy. In any case, nowadays statistical homogeneity and isotropy is not an independent principle. Theoretically it follows from cosmic inflation and it can be observationally tested. In the first part of the course we only consider spacetimes with exact spatial homogeneity and isotropy, in the second part we will look at perturbations around homogeneity and isotropy.

Since space is spatially homogeneous and isotropic, the spacetime curvature is the same at all points in space, but can vary in time. It can be shown that the metric can be written in spherical coordinates as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2 \left[\frac{\mathrm{d}r^2}{1 - Kr^2} + r^2 \mathrm{d}\theta^2 + r^2 \sin^2\theta \, \mathrm{d}\varphi^2 \right] . \tag{3.1}$$

An alternative form, in Cartesian coordinates, is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2 \frac{1}{\left(1 + \frac{K}{4} r^2\right)^2} (\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2) \ . \tag{3.2}$$

In either form, this is called the *Robertson–Walker* (RW) metric, sometimes the *Friedmann–Robertson–Walker* (FRW) metric or the *Friedmann–Lemaître–Robertson–Walker* (FLRW) metric. Nowadays the most common term is probably the FLRW metric, and we mostly stick to that. However, some authors prefer to make a distinction between the geometry (with the names Robertson and Walker attached) and the equations of motion (endowed with the name Friedmann and sometimes also Lemaître). Note that neither form of the metric has the same amount of symmetry as the spacetime itself: the metrics are isotropic, but not homogeneous. The full symmetry of the spacetime is usually not apparent in the metric itself, even though all physical quantities calculated from the metric display the symmetry. Here K is a constant, related to curvature of space (not spacetime) and $a(t)$ is a function of time, called the *scale factor*, that tells how the universe expands (or contracts). We will need the Einstein equation to solve $a(t)$. From the geometrical point of view, it is just an arbitrary function of the time coordinate t. The 2-dimensional surfaces where both t and r are constant have the metric of a sphere with radius ar, for any value of K.

The Robertson–Walker metric has two associated length scales, both of which in general evolve in time. The first is the *curvature radius* of space,

$$R_{\mathrm{curv}} \equiv a(t)/\sqrt{|K|} \ . \tag{3.3}$$

The second is the time scale of the expansion, the *Hubble time*, $t_H \equiv H^{-1}$, where $H \equiv \dot{a}/a$ is the *Hubble parameter*, and dot denotes derivative with respect to t. The Hubble time multiplied by the speed of light, $c = 1$, is the *Hubble length*, $\ell_H \equiv ct_H = H^{-1}$. In the case $K = 0$ the Hubble length is the only length scale.

The coordinates (t, r, θ, φ) of the Robertson–Walker metric are called *comoving* coordinates. This means that the coordinate system follows the expansion of space, so the spatial coordinates of objects that do not move with respect to the homogeneous and isotropic frame remain the same. The homogeneity of the universe fixes a special frame of reference, the *cosmic rest frame* given by the above coordinate system, so (unlike in the empty Minkowski space) the concept "does not move" can be defined in a physically meaningful way. The coordinate distance between two such objects stays the same, but their physical, or *proper*, distance grows with time as space expands. The time coordinate t, the *cosmic time*, gives the proper time measured by such an observer, at $(r, \theta, \varphi) = \text{constant}$.

We have the freedom to rescale the radial coordinate r. For example, nothing changes if we multiply all values of r by 2 and also divide a by 2 and K byf 4. The spacetime geometry stays the same, the meaning of the coordinate r has just changed: the point that had a given value of r has now twice that value in the rescaled coordinate system. There are two choices of how to use the rescaling to simplify notation. If $K \neq 0$, we can rescale r to make K equal to ± 1. In this case K is usually denoted k. Then r is dimensionless, and $a(t)$ has the dimension of distance. The other possibility is to set the scale factor today to unity[1], $a(t_0) \equiv a_0 = 1$. We will use this latter convention in this course, unless otherwise noted. In this case $a(t)$ is dimensionless, and r and $K^{-1/2}$ have the dimension of distance.

If $K = 0$, the spatial part (t = constant) of the Robertson–Walker metric is flat, i.e. that of ordinary Euclidean space, with the radial distance given by ar. The *spacetime* is, however, curved, since $a(t)$ depends on time, describing the expansion or contraction of space. It is often said that the "universe is flat" in this case, though if the universe is understood as the four-dimensional spacetime (as opposed to a spatial slice), "spatially flat" is more correct.

If $K > 0$, the coordinate system is singular at $r = r_K \equiv 1/\sqrt{K}$. (Recall the discussion of the 2-sphere in the previous chapter.) With the coordinate transformation $r = r_K \sin\chi$ the metric becomes

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)K^{-1}\left[\mathrm{d}\chi^2 + \sin^2\chi(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)\right]. \tag{3.4}$$

The spatial part has the metric of a 3D *hypersphere*, a sphere with one extra dimension. There is a new angular coordinate χ, whose values range from 0 to π, just like θ. The singularity at $r = 1/\sqrt{K}$ disappears in this coordinate transformation, showing that it was just a coordinate artifact, not a physical singularity. The original coordinates covered only half of the hypersphere, as the coordinate singularity $r = 1/\sqrt{K}$ divides the hypersphere into two halves. The case $K > 0$ corresponds to a *closed* universe, with positive spatial curvature.[2] This is a finite universe, with circumference $2\pi a r_K = 2\pi R_{\mathrm{curv}}$ and volume $2\pi^2 a^3 r_K^3 = 2\pi^2 R_{\mathrm{curv}}^3$, and R_{curv} is the radius of the hypersphere.

If $K < 0$, there is no coordinate singularity, and r ranges from 0 to ∞. The substitution $r = |K|^{-1/2}\sinh\chi$ is, however, often useful in calculations. The case $K < 0$ corresponds to an *open* universe, with negative spatial curvature. The metric is

$$ds^2 = -dt^2 + a^2(t)|K|^{-1}\left[\mathrm{d}\chi^2 + \sinh^2\chi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right]. \tag{3.5}$$

This universe is infinite, just like in the case $K = 0$[3].

It can be shown that expansion causes the motion of an object in free fall to slow down with respect to the comoving coordinate system. For nonrelativistic physical

[1] In some discussions of the early universe, it is more convenient to set a to unity at some early time instead.

[2] Positive (negative) curvature means that the sum of angles of any triangle is greater than (less than) 180° and that the area of a sphere with radius r is less than (greater than) $4\pi r^2$.

[3] The terminology of open vs. closed refers to the simplest possible choice of topology for the space. The $K > 0$ models are always finite, but it is also possible for the $K = 0$ and $K < 0$ models to be finite in a compact space with non-trivial topology. We will not discuss this possibility. (The mathematically oriented reader will note that the terms "open" and "closed" do not have the same meaning in cosmology as they do in topology.)

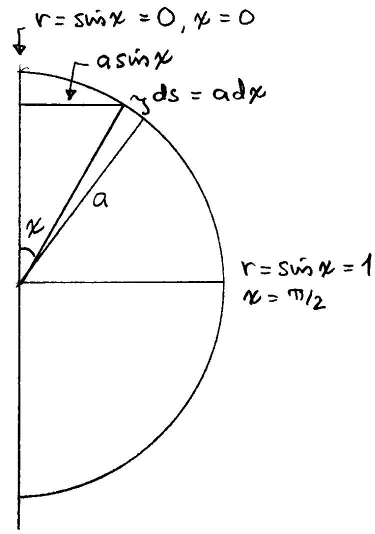

Figure 1: The hypersphere. This figure is for $K = k = 1$. Consider the semicircle in the figure. It corresponds to χ ranging from 0 to π. You get the (2-dimensional) sphere by rotating this semicircle off the paper around the vertical axis by an angle $\Delta\varphi = 2\pi$. You get the (3-dimensional) hypersphere by rotating it twice, in two extra dimensions, by $\Delta\theta = \pi$ and by $\Delta\varphi = 2\pi$, so that each point makes a sphere. Thus each point in the semicircle corresponds to a full sphere with coordinates θ and φ, and radius $(a/\sqrt{K})\sin\chi$.

velocities we have

$$v(t_2) = \frac{a(t_1)}{a(t_2)} v(t_1) \ . \tag{3.6}$$

The velocity of a galaxy with respect to the background is called *peculiar velocity.*

3.1.2 Conformal time

If we want the metric to remain isotropic, we cannot make coordinate transformations that mix the time coordinate with the spatial coordinates. However, just as we redefined the radial coordinate, we can make redefinitions that involve only the time coordinate. In the comoving coordinates used above, the spatial part of the coordinate system is expanding with the expansion of the universe. It is often practical to change the time coordinate so that the unit of time (i.e. separation of time coordinate surfaces) also increases in time. The *conformal time* η is defined as

$$\mathrm{d}\eta \equiv \frac{1}{a(t)}\mathrm{d}t, \qquad \text{or} \qquad \eta = \int^t \frac{\mathrm{d}t'}{a(t')}. \tag{3.7}$$

The metric can be written in a convenient form as

$$\begin{aligned} \mathrm{d}s^2 &= a(\eta)^2 \left[-\mathrm{d}\eta^2 + \mathrm{d}\chi^2 + \frac{1}{-K}\sinh^2\left[\sqrt{-K}\chi\right]\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right] &(3.8)\\ &= a(\eta)^2[-\mathrm{d}\eta^2 + \mathrm{d}\chi^2 + \left\{\begin{matrix}\sin^2\chi \\ \chi^2 \\ \sinh^2\chi\end{matrix}\right\}\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)] \ . &(3.9)\end{aligned}$$

In the first equality $\sinh\left(\sqrt{-K}\chi\right)$ is to be understood as the analytic continuation in the case $K > 0$ (so we get sin with a real argument), and as the limit of $K \to 0$ in the case $K = 0$. Note that here we have redefined (for $K \neq 0$) $\chi \to \sqrt{|K|}\chi$ compared to the previous expressions, to shift the factors of $\sqrt{|K|}$ away from the $\mathrm{d}\chi^2$ term to the angular part. We have also denoted $r \equiv \chi$ in the case $K = 0$. In the second equality we have chosen the normalisation of $a(t)$ so that $K = k = \pm 1, 0$, and the cases in the curly brackets correspond, from top to bottom, to $K = +1, 0, -1$. This form of the metric makes it easy to calculate the proper spatial distance along a straight line: if we choose the coordinates so that the line is radial, it is simply $a(\eta)\Delta\chi$, where $\Delta\chi$ is the difference between the values of χ at the end and the beginning. This form is particularly suited for light propagation. Since light travels in straight line in space (due to toe symmetry), we can choose the light propagation direction as the radial direction, so $\mathrm{d}\theta = \mathrm{d}\phi = 0$, and the remaining part of the metric is conformal to the 1+1-dimensional Minkowski metric in terms of η and χ. The condition $\mathrm{d}s^2 = 0$ then leads simply to $\mathrm{d}\eta = \pm\mathrm{d}\chi$, and light rays travel in 45° angles in the coordinates (η, χ).

3.1.3 Redshift

As mentioned in chapter 1, redshift is one of the most important cosmological observables. Let us find how it is related to the spacetime geometry in the case of the FLRW metric. Consider galaxy A. Light leaves the galaxy at time t_1 with wavelength λ_1 and arrives at galaxy O at time t_2 with wavelength λ_2. It takes a time $\delta t_1 = \lambda_1/c = 1/f_1$ to send one wavelength and a time $\delta t_2 = \lambda_2/c = 1/f_2$ to receive one wavelength. Follow now the two light rays sent at times t_1 and $t_1 + \delta t_1$ (see figure 2). Since all directions are equivalent, we can choose the direction of propagation to be radial, so that $\mathrm{d}\theta = \mathrm{d}\varphi = 0$. Light follows *lightlike* geodesics for which

$$\mathrm{d}s^2 = 0 \ . \tag{3.10}$$

We thus have

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\frac{\mathrm{d}r^2}{1 - Kr^2} = 0 \tag{3.11}$$

$$\Rightarrow \quad \frac{\mathrm{d}t}{a(t)} = -\frac{\mathrm{d}r}{\sqrt{1 - Kr^2}} \ . \tag{3.12}$$

Integrating this, we get for the first light ray

$$\int_{t_1}^{t_2} \frac{\mathrm{d}t}{a(t)} = \int_0^{r_\mathrm{A}} \frac{\mathrm{d}r}{\sqrt{1 - Kr^2}} \ . \tag{3.13}$$

For the second light ray we get

$$\int_{t_1+\delta t_1}^{t_2+\delta t_2} \frac{\mathrm{d}t}{a(t)} = \int_0^{r_\mathrm{A}} \frac{\mathrm{d}r}{\sqrt{1 - Kr^2}} \ . \tag{3.14}$$

The right hand sides of the two equations are the same, since neither the sender nor the receiver have moved (they stay at $r = r_A$ and $r = 0$). Thus

$$0 = \int_{t_1+\delta t_1}^{t_2+\delta t_2} \frac{\mathrm{d}t}{a(t)} - \int_{t_1}^{t_2} \frac{\mathrm{d}t}{a(t)} = \int_{t_2}^{t_2+\delta t_2} \frac{\mathrm{d}t}{a(t)} - \int_{t_1}^{t_1+\delta t_1} \frac{\mathrm{d}t}{a(t)} = \frac{\delta t_2}{a(t_2)} - \frac{\delta t_1}{a(t_1)}, \tag{3.15}$$

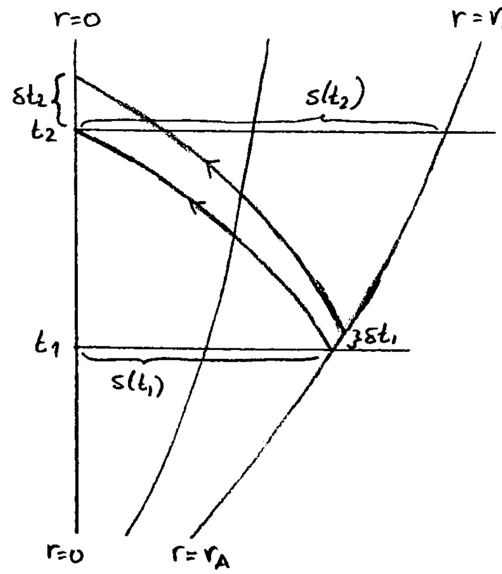

Figure 2: The two light rays to establish the redshift.

and the time to receive one wavelength is

$$\delta t_2 = \frac{a(t_2)}{a(t_1)}\,\delta t_1. \tag{3.16}$$

This derivation is even simpler when using the coordinates discussed in (3.1.2): it is then obvious that $\delta\eta_2 = \delta\eta_1$, from which the result (3.16) follows immediately.

As is clear from the derivation, this *cosmological time dilation* effect applies to observing any event taking place in galaxy A. We see everything happening in galaxy A in slow motion, slowed down by the factor $a(t_2)/a(t_1)$, which is the factor by which the universe has expanded since the light (or any electromagnetic signal) left the galaxy. This effect has been observed e.g. in the light curves (luminosity as a function of time) of supernovae.

For the redshift we have the result

$$1 + z \equiv \frac{\lambda_2}{\lambda_1} = \frac{\delta t_2}{\delta t_1} = \frac{a(t_2)}{a(t_1)}\ . \tag{3.17}$$

The result is simple: the wavelength expands with the universe. So the redshift tells us how much smaller the universe was when the light left the galaxy.

3.1.4 Age-redshift relation

If we see a source at redshift z, how old was the universe when the light left the source? In the FLRW universe we have

$$\mathrm{d}t = \frac{\mathrm{d}a}{\dot{a}} = \frac{\mathrm{d}a}{a}\frac{1}{H} = -\frac{\mathrm{d}z}{1+z}\frac{1}{H}\ , \tag{3.18}$$

where we have assumed that $\dot{a} \neq 0$, i.e. that $a(t)$ is monotonic. The age of the universe at redshift z is then

$$t(z) = \int_0^t \mathrm{d}t' = \int_z^\infty \frac{\mathrm{d}z'}{1+z'}\frac{1}{H(z')}\ , \tag{3.19}$$

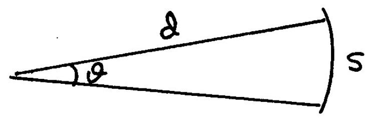

Figure 3: Defining the angular diameter distance.

where we have already used the knowledge that for realistic cosmological models, the universe has a finite age and that at the beginning $a = 0$ (i.e. $z = \infty$), and chosen the beginning of time as $t = 0$. Putting $z = 0$ gives the present age of the universe,

$$t_0 = \int_0^{\infty} \frac{\mathrm{d}z'}{1+z'} \frac{1}{H(z')} \ . \tag{3.20}$$

Subtracting the two tells us for how long the photons have travelled to arrive at our detectors:

$$\Delta t \equiv t_0 - t(z) = \int_0^{z} \frac{\mathrm{d}z'}{1+z'} \frac{1}{H(z')} \ . \tag{3.21}$$

Note that whereas time t is a coordinate whose origin is in the past (usually chosen to be at the beginning of the universe), the origin of the redshift is set to be today. Conceptually, t is just like Newtonian time, so it is simple to use. For example, if we discuss two different cosmological models, it is straightforward to compare them when the universe has the same age (assuming both have a beginning of time). In contrast, comparing them at the same redshift doesn't make sense unless we specify by which criteria you select "today" in the two models. Observationally, however, it is difficult to directly determine the age of the universe, while measuring the redshift is easy. The redshift is useful when it is related to quantities which are easier to measure than time, such as distances, to which we now turn.

3.1.5 Angular diameter distance

Almost all cosmological observations are made along the past lightcone, and important observable quantities include, in addition to the redshift, angular diameter and luminosity. We want to use the FLRW model to relate these observable quantities to the parameters of the model, so that we can constrain the geometry of the universe with observations. (As we will soon see, we can then also constrain the matter content of the universe.)

In Euclidean space, an object with proper size $\mathrm{d}s$ distance D away is seen at an angle (when $D \gg \mathrm{d}s$)

$$\mathrm{d}\theta = \frac{\mathrm{d}s}{D} \ . \tag{3.22}$$

In general relativity, we therefore *define* the *angular diameter distance* of an object with proper size $\mathrm{d}s$ and angular size $\mathrm{d}\theta$ to be

$$D_A \equiv \frac{\mathrm{d}s}{\mathrm{d}\theta} \ . \tag{3.23}$$

The reasoning of the Euclidean situation is here reversed. Objects do not look smaller because they are further away, *they are further away because they look smaller*. In the case of curved spacetime this can lead to behaviour at odds with intuition from Euclidean geometry; we will encounter one example in the next section. In order to determine the angular diameter distance, we need to know the proper spatial size of the object we are observing. Suppose we have a set of *standard rulers*, objects that we know are all the same small size $\mathrm{d}s$, observed at different redshifts. Their observed angular sizes $\mathrm{d}\theta(z)$ then give us the angular diameter distance as $D_A(z) = \mathrm{d}s/\mathrm{d}\theta(z)$. This can then be compared to the *theoretical* $D_A(z)$ for the FLRW universe to find parameter values which give the best fit between observation and theory. In cosmology, we don't have reliable knowledge of the precise size of individual objects on the sky, so spatial scales involved in statistical distributions are used instead. The two most notable cases of such distributions are the anisotropy pattern of the CMB and the pattern of galaxies on the sky, in particular baryonic acoustic oscillations.

From the FLRW metric, the proper distance corresponding to angle θ is, from $\mathrm{d}s^2 = a^2(t)r^2\mathrm{d}\theta^2 \quad \Rightarrow \quad \mathrm{d}s = a(t)r\mathrm{d}\theta$. We thus have

$$D_A = a(t)r = \frac{1}{1+z}r \ . \tag{3.24}$$

Now we have to relate the radial coordinate r to the observed redshift. As light travels on null geodesics, we have

$$\begin{aligned} ds^2 &= -\mathrm{d}t^2 + a(t)^2\frac{\mathrm{d}r^2}{1-Kr^2} = 0 \\ \Rightarrow \mathrm{d}t &= -a(t)\frac{\mathrm{d}r}{\sqrt{1-Kr^2}} \ . \end{aligned} \tag{3.25}$$

Since we place the observer at the center, the radial coordinate for incoming light rays decreases as time increases, hence the minus sign. Integrating, we obtain

$$\begin{aligned} \int_{t_1}^{t_0}\frac{\mathrm{d}t}{a(t)} &= \int_0^r \frac{\mathrm{d}r}{\sqrt{1-Kr^2}} \\ &= (-K)^{-1/2}\,\mathrm{arsinh}[(-K)^{1/2}r] \\ &= \begin{cases} K^{-1/2}\arcsin(K^{1/2}r) & K>0 \\ r & K=0 \\ |K|^{-1/2}\,\mathrm{arsinh}(|K|^{1/2}r) & K<0 \ . \end{cases} \end{aligned} \tag{3.26}$$

To facilitate handling all three cases simultaneously, we define the function

$$S_K(x) \equiv \frac{1}{\sqrt{-K}}\sinh(\sqrt{-K}x) = \begin{cases} K^{-1/2}\sin(K^{1/2}x) & K>0 \\ x & K=0 \\ |K|^{-1/2}\sinh(|K|^{1/2}x) & K<0 \ . \end{cases} \tag{3.27}$$

In other words, $\frac{1}{\sqrt{-K}}\sinh(\sqrt{-K}x)$ is understood as the analytical continuation when K is positive, and as the limit of small K when $K=0$. The inverse of this function is denoted by $S_K^{-1}(x)$. Putting together (3.26) and (3.27), we have

$$\begin{aligned} r &= \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K} \int_{t_1}^{t_0} \frac{\mathrm{d}t}{a(t)} \right) \\ &= \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K} \int_0^z \frac{\mathrm{d}z'}{H(z')} \right) , \end{aligned} \tag{3.28}$$

where we have on the second line used the relation (3.18).

Inserting (3.28) into (3.24), we finally obtain the angular diameter distance as a function of redshift:

$$\begin{aligned} D_A(z) &= (1+z)^{-1} S_K\left(\int_0^z \frac{\mathrm{d}z'}{H(z')} \right) \\ &= (1+z)^{-1} \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K} \int_0^z \frac{\mathrm{d}z'}{H(z')} \right) . \end{aligned} \tag{3.29}$$

In the spatially flat case this reduces to

$$D_A(z) = (1+z)^{-1} \int_0^z \frac{\mathrm{d}z'}{H(z')} . \tag{3.30}$$

This relation tells us how distance scales in the FLRW universe change because of the expansion of the universe. For a general FLRW metric, the angular diameter distance depends only on the redshift, the coordinate curvature radius $1/\sqrt{-K}$ and the integral over the inverse Hubble parameter. Note that if the universe expands rapidly in the past (as is the case in the real universe), the contribution to the distance from early times is small at late times, because the length scales at early times were much smaller.

Because of the factor $(1+z)^{-1}$, the angular diameter distance is not necessarily monotonic in redshift, i.e. the distance to objects may *decrease* with growing redshift. This curious feature is present in realistic cosmological models, and can occur even if the spatial geometry is Euclidean (i.e. $K = 0$). It is related to the fact that the angular diameter distance is defined along a lightlike direction in the non-Euclidean spacetime, not along a spatial slice (which may or may not be non-Euclidean).

3.1.6 Luminosity distance

The luminosity distance is defined in a similar manner. In Euclidean space, if an object radiates isotropically with *absolute luminosity* L (this is the radiated energy per unit time measured next to the object), an observer at distance D sees the flux (energy per unit time per unit area)

$$F = \frac{L}{4\pi D^2} . \tag{3.31}$$

In general relativity, the *luminosity distance* D_L is defined as

$$D_L \equiv \sqrt{\frac{L}{4\pi F}} . \tag{3.32}$$

As with the angular diameter distance, objects in curved spacetime are further away because they look fainter, not the other way around. (However, at least in homogeneous and isotropic models, the luminosity distance always grows with increasing spatial distance, which is not the case for D_A.)

Consider the situation in the FLRW universe. The absolute luminosity can be expressed as:

$$L = \frac{\text{number of photons emitted}}{\text{time}} \times \text{their average energy} = \frac{N_\gamma E_{\text{em}}}{t_{\text{em}}}\,. \tag{3.33}$$

If the observer is at a coordinate distance r from the source, the photons have at that distance spread over the area (recall that $a(t_0) = 1$)

$$A = 4\pi r^2\,. \tag{3.34}$$

The flux can be expressed as:

$$F = \frac{\text{number of photons observed}}{\text{area} \cdot \text{time}} \times \text{their average energy} = \frac{N_\gamma E_{\text{obs}}}{t_{\text{obs}} A}\,. \tag{3.35}$$

The number of photons N_γ is conserved, but their energy is redshifted, $E_{\text{obs}} = E_{\text{em}}/(1+z)$. Also, if the source is at redshift z, it takes a factor $1+z$ longer to receive the photons $\Rightarrow t_{\text{obs}} = (1+z)t_{\text{em}}$. Thus,

$$F = \frac{N_\gamma E_{\text{obs}}}{t_{\text{obs}} A} = \frac{N_\gamma E_{\text{em}}}{t_{\text{em}}} \frac{1}{(1+z)^2} \frac{1}{4\pi r^2}\,. \tag{3.36}$$

We thus have

$$\begin{aligned} D_L &= \sqrt{\frac{L}{4\pi F}} = (1+z) r \\ &= (1+z)^2 D_A(z) \\ &= (1+z) \frac{1}{\sqrt{-K}} \sinh\left(\sqrt{-K} \int_0^z \frac{\mathrm{d}z'}{H(z')} \right)\,, \end{aligned} \tag{3.37}$$

where we have used (3.24) and (3.29). Compared to the angular diameter distance $D_A(z)$, there are two extra factors of $1+z$. One-half comes from the redshift of photon energy, one-half from cosmological time dilation in receiving the emitted photons, and one from the change in the area element. The relation $D_L = (1+z)^2 D_A$ holds in any spacetime, not just in the FLRW case, so there is only one independent observational cosmological distance[4]

3.1.7 Proper distance

The only cosmological distances we can directly measure are those defined along lightlike curves. However, spacelike distances are still theoretically interesting. In particular, the proper length of an object is a useful quantity: it is the size of an object (more generally, distance between two points) in the rest frame of that object (more generally, at a surface of constant cosmic time). Deviations from the mean flow are small, $v < 10^{-3}$ (recall that $c = 1$, so $10^{-3} = 3000$ km/s), so effects like Lorentz contraction and time dilation are small.

Proper distance is defined as the physical distance measured on a slice of constant time. If we consider the proper distance between galaxy O and galaxy A, we can

[4] The *parallax distance*, related to the change of angular position of objects on the sky due to the movement of the observer, does not reduce to the angular diameter distance, but at present it has not been measured on cosmological scales.

without loss of generality choose the direction between them to be radial and set O to be at the origin $r = 0$; let us choose A to be at radial coordinate r. The distance interval is given by

$$ds^2 = a(t)^2 \frac{dr^2}{1 - Kr^2} , \tag{3.38}$$

so the proper distance is

$$\begin{aligned} D_P &= \int_O^A ds \\ &= a(t) \int_0^r \frac{dr}{\sqrt{1 - Kr^2}} \\ &= a(t) S_K^{-1}(r) . \end{aligned} \tag{3.39}$$

The distance between two points which are fixed in the comoving coordinates grows proportionally to the scale factor as the universe expands, as we would expect. Note that in terms of the χ-coordinates (3.8), the proper distance is simply $D_P = a(t)\chi$. (In that case the non-Euclidean aspects of the spatial geometry are shuffled to the angular part, whereas in the r-coordinates they are put into the radial part of the metric.)

In cosmology, it is common to use the *comoving distance*, which just means the physical distance to redshift z scaled by the difference between the scale factor then and now. So if we have some distance measure $D(z)$, the corresponding comoving distance, denoted $D^c(z)$, is $D^c(z) = (1 + z)D(z)$. The idea is that it is easier to compare objects from different eras if we discuss them in terms of the distance they would now span. For example, the sound horizon of the photon-baryon plasma at the time of last scattering when the universe was about 380 000 years old is $r_s \approx 0.13$ Mpc, whereas the comoving sound horizon is $(1 + z_*)r_s \approx 140$ Mpc, where $1 + z_* =$ is the redshift of the last scattering surface. This is especially convenient for the comoving proper distance, which remains constant in time,

$$D_P^c = (1 + z)D_P = S_K^{-1}(r) . \tag{3.40}$$

The relation (3.39) shows how the coordinate r is related to the physical distance D_P,

$$r = S_K(D_P/a) = S_K(D_P^c) . \tag{3.41}$$

The radial coordinate r does not give the physical distance, but nevertheless has a clear physical interpretation. The physical distance to an object at coordinate r is D_P, the length of the circle with physical radius $D_P(t, r)$ is $2\pi a(t)r$ and its surface area is $4\pi a(t)^2 r^2$, as can be immediately verified from the FLRW metric (3.1).

The functions S_K and S_K^{-1} convert between two natural length measures of a FLRW universe: the proper distance measured along the radial line (i.e. the *proper radius*) and the area distance measured along the surface of a sphere. The fact that these quantities do not agree is a reflection of the fact that the space is non-Euclidean. In the flat case with $K = 0$, we have simply $S_K(x) = x$, as the space is Euclidean. In this case the only relativistic effect is the stretching of space.

In addition to the straightforward issue of proper distance as a function of time as measured on the spacelike slice, we can ask the following slightly more involved

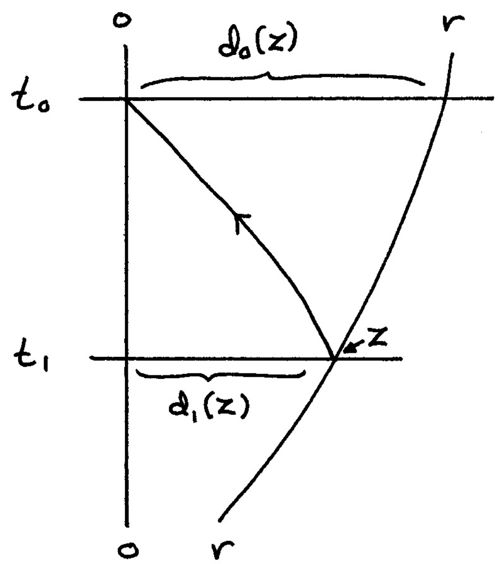

Figure 4: Calculation of the proper distance.

question: if we see (along a null geodesic) a galaxy at redshift z, what is the proper distance (along the spacelike slice) to the galaxy today? Here we assume that the galaxy is at rest in the comoving frame (i.e. we neglect peculiar velocities) and still exists today. (In fact, we cannot know what has happened to the galaxy since the light left it.)

From (3.24), (3.29) and (3.39) we have for the proper distance to object that emitted light at time t_1, as measured at time t:

$$D_P(t_1, t) = a(t) S_K^{-1}(D_A/a) = a(t) \int_{t_1}^{t} \frac{\mathrm{d}t'}{a(t')} = (1+z)^{-1} \int_{z}^{z_1} \frac{\mathrm{d}z'}{H(z')} , \tag{3.42}$$

Note that this result is independent of spatial curvature. So the proper distance to redshift z today is

$$D_P(z) = \int_0^z \frac{dz'}{H(z')} . \tag{3.43}$$

As the distance in (3.43) is defined today, it makes no difference whether it is comoving or not. The longest distance (as measured along the spatial slice) from which it has been possible to receive signals at time t is called the *horizon distance* $D_{\rm hor}$ at time t. We get it by putting $t_1 = 0$, or equivalently $z = \infty$ (as $a(0) = 0$) in (3.42),

$$D_{\rm hor}(t) = a(t) \int_0^t \frac{\mathrm{d}t'}{a(t')} = (1+z)^{-1} \int_z^\infty \frac{dz'}{H(z')}$$

$$\Rightarrow \quad D^c_{\rm hor}(t) = \int_0^t \frac{\mathrm{d}t'}{a(t')} = \int_z^\infty \frac{dz'}{H(z')} . \tag{3.44}$$

We get the horizon distance today by putting $t = t_0$ or equivalently $z = 0$.

There are actually a few different concepts in cosmology called the horizon. The one given above is the *particle horizon*, and it indicates the maximum distance from which we can in principle have received any information up to now. Another horizon concept is the *event horizon*, which is related to how far the light can travel in the future. (More precisely, the event horizon is the boundary of the region, if any, from which the observer can never receive any signals, even infinitely far into the future.) The *Hubble distance* H^{-1} is also often referred to as the horizon (especially when one talks about *subhorizon* and *superhorizon* distance scales.). This terminology is somewhat confusing, although widely used. For realistic cosmological models without inflation, the particle horizon and the Hubble distance are of the same order of magnitude, they differ only by a factor of order unity. Inflation changes this dramatically, as we will see later.

3.1.8 The Hubble law

In chapter 2 we discussed the Hubble law, which is a redshift-distance relationship is linear for small redshifts, $z = H_0 D$. Given the different measures of distance (and we can define new distance measures simply by multiplying D_A by any power of $(1+z)$), the question arises: what is the distance that appears in the Hubble law?

The answer is that for small redshifts, all of the above distance measures agree. From (3.29), (3.37) and (3.42) we get

$$D_A \simeq D_L \simeq D_P \simeq H_0^{-1} z \tag{3.45}$$

for $z \ll 1$, where $H(z=0) = H_0$ is the Hubble constant. For redshifts that are not small, the relation between the distance and the redshift is more complicated, as shown by (3.29), (3.37) and (3.42). We then need to know not just the present value H_0, but the function $H(z)$ all the way to the redshift of the source (in the case of the angular diameter distance and the luminosity distance, we also need the spatial curvature). The function $H(z)$ is determined by the matter content according to the dynamics of general relativity, to which we now turn.

3.2 Dynamics

3.2.1 The Friedmann–Lemaître equations

The considerations thus far have been purely geometrical and kinematical. In order to find how the scale factor $a(t)$ evolves, we need to consider the Einstein equation. In general, the Einstein equation is a non-linear system of ten partial differential equations. In the case of the FLRW universe, it reduces to two ordinary non-linear differential equations,

$$3\frac{\dot{a}^2}{a^2} + 3\frac{K}{a^2} = 8\pi G_{\rm N}\rho + \Lambda \tag{3.46}$$

$$3\frac{\ddot{a}}{a} = -4\pi G_{\rm N}(\rho + 3p) + \Lambda \ , \tag{3.47}$$

where ρ is the energy density, p is the pressure, and Λ is the *cosmological constant*. Homogeneity implies that they only depend on time, $\rho = \rho(t)$, $p = p(t)$. These are called the *Friedmann equations*, or sometimes the *Friedmann–Lemaître equations*, *FRW equations*, or *FLRW equations*. We will mostly use the term Friedmann equations. The expression *Friedmann equation* in the singular refers to (3.46).

From the Einstein equation also follows the *continuity equation*, which is related to conservation (or not) of energy and momentum. In the FLRW case it reduces to

$$\dot{\rho} = -3(\rho + p)\frac{\dot{a}}{a} \ . \tag{3.48}$$

(**Exercise**: Derive (3.48) from the Friedmann equations.) The equation (3.48) shows how the energy density evolves as the universe expands. We can rewrite (3.48) as

$$\begin{aligned} p &= -\frac{1}{3H}\frac{1}{a^3}\frac{\mathrm{d}(a^3\rho)}{\mathrm{d}t} \\ &= -\frac{\mathrm{d}(a^3\rho)}{\mathrm{d}(a^3)} \ . \end{aligned} \tag{3.49}$$

If the pressure is zero, the energy in some volume remains constant as the universe expands or contracts. If the pressure is positive, the total amount of energy decreases with the expansion of the universe (and increases if the universe contracts). If the pressure is negative, the opposite happens: then the energy of an expanding universe increases. We can compare (3.49) with the first law of thermodynamics,

$$T\mathrm{d}S = \mathrm{d}U + p\mathrm{d}V - \sum_i \mu_i \mathrm{d}N_i \ , \tag{3.50}$$

where T is the temperature, S is the entropy, U is the internal energy, V is volume and μ_i and N_i are chemical potential and particle number for particle species i. We see that the energy density in a FLRW universe changes like the energy density of a gas which expands or contracts adiabatically and has constant particle number. However, while pressure has a kinematical interpretation in the statistical physics of a gas of particles, the quantity p appearing in (3.49) is more general. The pressure of matter which consists of a gas of (almost) free particles is always positive, but other forms of matter (such as coherent scalar fields or topological defects) can have negative pressure.

3.2.2 Critical density

The Hubble parameter $H(t)$ gives the expansion rate of the universe. The dimension of H is 1/time, or equivalently 1/distance. The Friedmann equation (3.46) connects the density ρ, the cosmological constant Λ, the spatial curvature K/a^2, and the expansion rate H:

$$3H^2 = 8\pi G_{\mathrm{N}}\rho + \Lambda - 3\frac{K}{a^2} \ . \tag{3.51}$$

Dividing by $3H^2$, we have

$$\begin{aligned} 1 &= \frac{8\pi G_{\mathrm{N}}\rho}{3H^2} + \frac{\Lambda}{3H^2} - \frac{K}{(aH)^2} \\ &\equiv \Omega + \Omega_K \ , \end{aligned} \tag{3.52}$$

where we have defined the *density parameters*

$$\begin{aligned} \Omega(t) &\equiv \frac{8\pi G_{\mathrm{N}}\rho + \Lambda}{3H^2} \\ \Omega_K(t) &\equiv -\frac{K}{(aH)^2} \ . \end{aligned} \tag{3.53}$$

Another often used quantity is the *critical density*, defined as

$$\rho_{\mathrm{c}}(t) \equiv \frac{3H(t)^2}{8\pi G_{\mathrm{N}}} \ . \tag{3.54}$$

The critical density is the energy density of a spatially flat universe that expands with the rate $H(t)$. So the critical density changes as the universe evolves and the Hubble parameter changes. Often in cosmology the word critical density is used to refer just to the present value. For clarity, we always use the subscript 0 when referring to the present value:

$$\rho_{\mathrm{c0}} \equiv \rho_{\mathrm{c}}(t_0) = \frac{3H_0^2}{8\pi G_{\mathrm{N}}} \ . \tag{3.55}$$

We have

$$\Omega(t) \equiv \frac{\rho(t)}{\rho_{\mathrm{c}}(t)} \ , \tag{3.56}$$

where we have now included the contribution of the cosmological constant in ρ, as is conventionally done.

Positive curvature contributes to the Hubble rate with a negative sign and negative curvature with a positive sign, as (3.46) shows. In other words, if we measure that the density of the universe is ρ and the critical density is ρ_{c} (i.e. the Hubble parameter is H), we can make the following conclusion about the spatial curvature:

$$\rho < \rho_{\mathrm{c}} \quad \Leftrightarrow \quad \Omega < 1 \Leftrightarrow \quad K < 0 \tag{3.57}$$

$$\rho = \rho_{\mathrm{c}} \quad \Leftrightarrow \quad \Omega = 1 \Leftrightarrow \quad K = 0 \tag{3.58}$$

$$\rho > \rho_{\mathrm{c}} \quad \Leftrightarrow \quad \Omega > 1 \Leftrightarrow \quad K > 0 \ . \tag{3.59}$$

Thus $\Omega = 1$ implies that the universe is spatially flat, $\Omega < 1$ implies that spatial curvature is negative and $\Omega > 1$ that spatial curvature is positive. The Friedmann equation can be written as

$$\Omega(t) = 1 + \frac{K}{a(t)^2 H(t)^2} = 1 + \left(\frac{\ell_H}{R_{\mathrm{curv}}}\right)^2 \ , \tag{3.60}$$

where ℓ_H is the Hubble length and R_{curv} is the curvature radius. If $\Omega < 1$ (or > 1) at some instant of time, it will stay that way (since K is constant). And if $\Omega = 1$, it will stay constant, $\Omega = \Omega_0 = 1$. Observations indicate that the density of the universe today is close to critical, $\Omega_0 - 1 = -0.001 \pm 0.002$ [4].

3.2.3 Matter components

In the two Friedmann equations (3.46) and (3.47), there are three unknowns, $a(t)$, $\rho(t)$ and $p(t)$. We can also consider a system of three equations, with (3.48) added to the mix, but in that case only two are independent. The system is underdetermined, reflecting the fact that different matter components affect the expansion rate differently: we have to specify which kind of matter there is in the universe. In order to close the system, it is enough to give the relation between pressure and the energy density: we can then solve for the energy density from (3.48) or (3.49) and insert the solution into (3.46) and integrate.

The relation between the pressure and the energy density is called the *equation of state*. In cosmology, this term refers specifically to the combination p/ρ. The simplest equations of state are *barotropic*, which means that the pressure is a function of the energy density, $p(\rho)$. (Scalar fields, for which the equation of state is not barotropic, will be important when we discuss inflation in the second part of the course.) The simplest possibilities are the following:

- **Matter.** The term "matter" refers to a form of matter for which the pressure is zero $p = 0$, or at least negligible, $|p| \ll \rho$. Such a form of matter is also called "dust". (The name "dust" is more common in a pure general relativity context than in cosmology.) This is the case for a gas of free non-relativistic particles, where the energy density is dominated by the mass. The relation (3.49) shows that $\mathrm{d}(\rho a^3)/\mathrm{d}t = 0$, or $\rho \propto a^{-3}$.

- **Radiation**. The term "radiation" refers to matter for which the pressure is (exactly or very closely) 1/3 of the energy density, $p = \frac{1}{3}\rho$. This is the case for a gas of free ultrarelativistic particles, for which the energy density is dominated by the kinetic energy (i.e. the momentum is much bigger than the mass). In particular, this always holds for massless particles such as photons. From (3.49), we get $\mathrm{d}(\rho a^4)/\mathrm{d}t = 0$, in other words $\rho \propto a^{-4}$.

- **Vacuum energy.** For vacuum energy the energy density does not change in time, $\rho =$ constant. From (3.48) it follows that the pressure is very negative, $p = -\rho$. (This type of matter is, a bit misleadingly, also called the cosmological constant; see section 3.2.4 below.) Thus, positive vacuum energy corresponds to negative vacuum pressure. The total amount of energy increases proportional to the volume of space because there is a constant amount of energy per unit volume.

The universe contains non-relativistic matter in the form or ordinary, baryonic matter (i.e. atoms, ions and electrons) as well as (most probably) *dark matter*, which is (practically) pressureless, weakly interacting and extremely cold. Dark matter is usually thought to consist of a gas of a new heavy particles. We will discuss dark matter in more detail in chapter 6. There is also radiation, most importantly in the form of the CMB, which is a remnant of the radiation that used to dominate the expansion of the universe. In addition, there are neutrinos, which behaved like radiation in the early universe but now behave like matter. This happens for all particles that are not strictly massless: the kinetic energy falls with the expansion of the universe, so that at some point the mass starts dominating the particle energy. In chapter 4 we discuss in detail how different particle species behave as radiation in the early universe when it is very hot, but as the universe cools, the massive particles change from being ultrarelativistic (radiation) to being nonrelativistic (matter). During the transition period the pressure due to that particle species falls from $p = \rho/3$ to $p \approx 0$. In this chapter, we focus on the late universe, when it is sufficient to divide matter into dust ($p \approx 0$) and radiation ($p \approx \rho/3$), without worrying about the transitions. (Neutrinos may undergo the transition quite late –the neutrino masses are not precisely known– but their contribution to the total energy budget is negligible at late times, so we can skip this detail.)

We have mentioned that the present observational data cannot be explained in terms of known particles (or hypothetical particles with similar properties), general relativity and the FLRW metric. One of the three assumptions –known forms of matter, general relativity and the approximation of homogeneity and isotropy– is then wrong. Sticking to the FLRW metric and general relativity, the observations indicate that the expansion of the universe has accelerated during the past few billion years. From (3.47) we see that this requires an energy component with negative pressure, *dark energy*. It is called *dark* since it has not been observed to emit or absorb light, and *energy*, since the name "dark matter" was already taken (though "dark pressure" might be more appropriate). The simplest possibility for dark energy is just the cosmological constant (vacuum energy), which generates repulsive gravity, leading to accelerated expansion. This explanation has explained and predicted a wide varity if data successfully for over 25 years. Therefore we shall carry on our discussion assuming that there are three components: matter, radiation, and vacuum energy. We will later comment on how much observations constrain the equation of state of dark energy, if it is not vacuum energy.

If the universe contains only these three components, we can arrange (3.46) into the form

$$3\frac{\dot{a}^2}{a^2} = 8\pi G_{\rm N}\rho_{r0}a^{-4} + 8\pi G_{\rm N}\rho_{m0}a^{-3} - 3Ka^{-2} + \Lambda, \qquad (3.61)$$

where $\rho_{r0}, \rho_{m0}, a_0$, K, and Λ are constants.[5] The four terms on the right hand side are due to radiation, matter, spatial curvature, and vacuum energy, respectively. As the universe expands (a grows), different components on the right hand side become important at different times. The universe was first radiation-dominated up to about 50 000 years, then the expansion was dominated by matter until a few billion years ago, when vacuum energy started to dominate. (There has apparently never been an era where the spatial curvature would have been the largest term.)

The radiation component is insignificant at late times. If we consider the universe after the first few million years, we can neglect it in (3.61). In most models of inflation, the expansion was dominated by an energy density that changes very slowly in time, and that behaves almost like a cosmological constant in the very early universe (during the first small fraction of the first second).

We divide the density into matter, radiation, and vacuum components $\rho = \rho_m + \rho_r + \rho_{\rm vac}$, so we have

$$\Omega = \Omega_{\rm m} + \Omega_{\rm r} + \Omega_\Lambda \ , \qquad (3.62)$$

where $\Omega_{\rm m} \equiv \rho_m/\rho_{\rm c}$, $\Omega_{\rm r} \equiv \rho_r/\rho_{\rm c}$, and $\Omega_\Lambda \equiv \rho_{\rm vac}/\rho_{\rm c} \equiv \Lambda/3H^2$. The density parameters $\Omega_{\rm m}$, $\Omega_{\rm r}$, and Ω_Λ are functions of time (although $\rho_{\rm vac}$ is constant, $\rho_{\rm c}(t)$ is not).

Even more so than in the case of the critical density, the symbols $\Omega_{\rm m}$, $\Omega_{\rm r}$, Ω_Λ and Ω_K are often used to denote the present values of these quantities. In this course, to

[5] We ignore transfer of energy between the components. Such transfer is important only in the early universe, before the decoupling of the different particle species, or when particle species go from being relativistic to non-relativistic. In chapter 4 we return to this issue in some detail.

avoid confusion, we always use the subscript 0 when referring to the present values, $\Omega_{m0} \equiv \Omega_m(t_0)$, $\Omega_{r0} \equiv \Omega_r(t_0)$, $\Omega_{\Lambda 0} \equiv \Omega_\Lambda(t_0)$, $\Omega_{K0} \equiv \Omega_K(t_0)$. The present radiation density is relatively small, $\Omega_{r0} \sim 10^{-4}$ (we will calculate the precise number in chapter 4). So we usually write just

$$\Omega_0 = \Omega_{m0} + \Omega_{\Lambda 0} \ . \tag{3.63}$$

In addition to being small today, the radiation density is also known very accurately from the temperature of the cosmic microwave background, and therefore Ω_{r0} is not usually considered as a cosmological parameter (in the sense of a poorly known quantity that we are trying to determine from observations). This simple FLRW cosmological model is thus defined by present values of three cosmological parameters, which we can take to be $H_0 = h100\text{km/s/Mpc}$, Ω_{m0}, and $\Omega_{\Lambda 0}$.

It is often useful to define the "physical" or "reduced" density parameters by multiplying away the dependence on the value of h: $\omega_m \equiv \Omega_{m0}h^2$, $\omega_r \equiv \Omega_{r0}h^2$. (The corresponding quantities ω_Λ and ω_K are not useful.) Note that the ω parameters are not defined as a function of time, they are constants defined with respect to present-day density only.

Two cosmological models have been particularly important. The first is the *Einstein–de Sitter model*, which contains only matter and is spatially flat, $\Omega_m = 1$, $\Omega_r = \Omega_K = \Omega_\Lambda = 0$. This model (with radiation added at early times, and with a specific spectrum of perturbations around homogeneity and isotropy – we will discuss this in the second part of the course) was known as the Standard CDM (SCDM) model from the 1980s onwards. The abbreviation CDM stands for cold dark matter.

At the end of the 1990s the SCDM model was supplanted by the ΛCDM model, which is identical except for the addition of vacuum energy (like SCDM, it is spatially flat). It is also known as the *Concordance Model* or the *standard model of cosmology* due to the fact that it has been able to fit a large number of independent observations, and has become the reference point for all other models. Comparing to observations, the parameters of the model turn out to be $h \approx 0.7$, $\Omega_{m0} \approx 0.3$ and $\Omega_{\Lambda 0} = 1 - \Omega_{m0} \approx 0.7$. The precise values depend on the datasets one fits to and the assumptions made in the analysis. CMB data from the Planck satellite gives the values [4]

$$\Omega_{m0} = 0.315 \pm 0.007 \ , \quad h = 0.674 \pm 0.005 \ . \tag{3.64}$$

3.2.4 Vacuum energy

Before proceeding into more details of the expansion history and the distance-redshift relationship for different matter contents, let us say a few words about the cosmological constant. It was originally introduced by Einstein because he thought the universe should be static. A look at (3.47) shows that this requires either negative pressure matter or a positive cosmological constant. Introducing a cosmological constant makes it possible to balance the gravitational attraction of matter against the repulsion of a positive cosmological constant. This model is called the *Einstein static universe*. (It is unstable to small perturbations and is thus not a viable model.)

While the cosmological constant is a geometrical term (a contribution to the left-hand side of the Einstein equation), we can add an identical term to the matter (the right-hand side of the Einstein equation), and this is called vacuum energy.

In quantum field theory, the fundamental physical objects are fields, particles are just quanta of the field oscillations. Vacuum refers to the ground state of the system, where fields have values which correspond to minimum energy. In quantum field theory, there is no reason why this minimum energy should be zero. The vacuum energy density is analogous to the zero-point energy of a harmonic oscillator in quantum mechanics. Vacuum energy is constant in time and space, so it is equivalent to the cosmological constant with the value

$$\Lambda = 8\pi G_{\mathrm{N}}\rho_{\mathrm{vac}} \ . \tag{3.65}$$

Vacuum energy is observationally indistinguishable from a cosmological constant, though conceptually they are different, because the former is a new matter component, and the latter is a modification of the law of gravity. We will sometimes drop Λ and instead include the vacuum in the energy budget.

A problem with vacuum energy is that the expected scale of vacuum fluctuations is huge, of the order of particle physics scales (perhaps the Planck scale of 10^{18} GeV, and at least 100 GeV), but observations restrict it to a much smaller value – if vacuum energy is responsible for acceleration at late times, the energy density is of the scale $(\mathrm{meV})^4$. Equivalently, the value of the cosmological constant $(\sim 10^{-33}\ \mathrm{eV})^2$. However, our present understanding of quantum theory does not allow us to calculate what the value of the vacuum energy is, so there is no conflict between theory and observation, just unmet expectations. Possibly there is some unknown principle which sets the vacuum energy to be zero, or at least prevents it from interacting gravitationally – or almost so. The cosmological constant problem was considered to be one of the most important issues in cosmology and particle physics already before the observation of late time acceleration.

3.2.5 Expansion and the big bang

Let us now solve the Friedmann equation in the case when it is dominated by a term with a constant equation of state, $\omega \equiv p/\rho =$ constant. From (3.48) we get

$$\rho \propto a^{-3(1+\omega)} \ . \tag{3.66}$$

As far as the expansion history is concerned, spatial curvature is equivalent to a fluid with the equation of state $\omega = -1/3$ and a positive (negative) energy density corresponding to negative (positive) spatial curvature, respectively. (However, the spatial curvature also changes the relation between the expansion rate and the distance, as we have discussed above; a fluid with the same equation of state has no such effect.) Inserting (3.66) into the Friedmann equation (3.46) and putting $K = 0$, we get

$$\frac{\dot{a}^2}{a^2} \propto a^{-3(1+\omega)} \ . \tag{3.67}$$

We assume that $\omega > -1$; the vacuum energy case $\omega = -1$ and the case $\omega < -1$ need to be treated separately. Integrating, we obtain

$$a \propto (t - t_i)^{\frac{2}{3(1+\omega)}} \ . \tag{3.68}$$

At a finite time in the past, the scale factor becomes zero; without loss of generality, we choose the origin of the time coordinate to be there, $t_i = 0$. At this time the energy density is correspondingly infinite, and the spacetime is infinitely curved. This singularity is called the *big bang*, and it is a general feature not only of FLRW models but also of realistic cosmological models which include inhomogeneities. Space and time do not continue beyond this event. However, at the big bang (or more properly, as we come near its vicinity) general relativity does not apply anymore, so we cannot make any definite statements about what happens very near the beginning. Also, we cannot really expect matter to behave in this simple way in the early universe; when we discuss inflation we will see one possibility of how the early universe can behave differently. (But inflation does not save us from the cosmological singularity.)

In particular, we have the three cases

$$w = 1/3 \qquad \text{radiation-dominated} \qquad a \propto t^{1/2} \tag{3.69}$$

$$w = 0 \qquad \text{matter-dominated} \qquad a \propto t^{2/3} \tag{3.70}$$

$$w = -1/3 \qquad \text{curvature-dominated } (K < 0) \qquad a \propto t\ . \tag{3.71}$$

In the case of vacuum energy domination, $\omega = -1$, the Hubble parameter H is constant, so the universe expands exponentially, $a \propto e^{\sqrt{\frac{\Lambda}{3}}t}$.

3.2.6 Age of universe

Now that we have a parametrised form of the expansion function $H(z)$, we can return to the age of the universe discussed in section 3.1.4. The Friedmann equation (3.46) reads

$$3H^2 = 8\pi G_{\rm N}\rho_{r0}a^{-4} + 8\pi G_{\rm N}\rho_{m0}a^{-3} - 3Ka^{-2} + \Lambda\ . \tag{3.72}$$

Dividing by $3H_0^2$, we get

$$\begin{aligned} \frac{H}{H_0} &= \sqrt{\Omega_{\rm r0}a^{-4} + \Omega_{\rm m0}a^{-3} + \Omega_{K0}a^{-2} + \Omega_{\Lambda 0}} \\ &= \sqrt{\Omega_{\rm r0}(1+z)^4 + \Omega_{\rm m0}(1+z)^3 + \Omega_{K0}(1+z)^2 + \Omega_{\Lambda 0}}\ . \end{aligned} \tag{3.73}$$

We will have much use for this convenient form of the Friedmann equation. Inserting (3.73) into the relation (3.19) for the age, we find the time it takes for the universe to expand from scale factor a_1 to a_2, or from redshift z_1 to z_2,

$$\begin{aligned} t_2 - t_1 &= \int_{z_2}^{z_1} \frac{{\rm d}z}{1+z}\frac{1}{H(z)} \\ &= H_0^{-1}\int_{z_2}^{z_1} \frac{{\rm d}z}{(1+z)\sqrt{\Omega_{\rm r0}(1+z)^4 + \Omega_{\rm m0}(1+z)^3 + \Omega_{K0}(1+z)^2 + \Omega_{\Lambda 0}}} \\ &= H_0^{-1}\int_{\frac{1}{1+z_1}}^{\frac{1}{1+z_2}} \frac{{\rm d}a}{\sqrt{\Omega_{\rm r0}a^{-2} + \Omega_{\rm m0}a^{-1} + \Omega_{K0} + \Omega_{\Lambda 0}a^2}}\ , \end{aligned} \tag{3.74}$$

where the second form is more convenient due to the cancellation of some factors of $1+z$. Recall that $\Omega_{K0} = 1 - \Omega_{\rm r0} - \Omega_{\rm m0} - \Omega_{\Lambda 0} \equiv 1 - \Omega_0$. The expression (3.19) is integrable to an elementary function if two of the four terms under the root sign are absent. From this we get the age of the universe t at redshift z as

$$t(z) = H_0^{-1}\int_0^{\frac{1}{1+z}} \frac{{\rm d}a}{\sqrt{\Omega_{\rm r0}a^{-2} + \Omega_{\rm m0}a^{-1} + \Omega_{K0} + \Omega_{\Lambda 0}a^2}}\ . \tag{3.75}$$

This gives the function $t(z)$, that is, $t(a)$. Inverting this function gives us $a(t)$, the scale factor as a function of time. The scale factor $a(t)$ is not necessarily an elementary function, even if $t(a)$ is. However, even in that case we can sometimes have a parametric representation $a(\psi)$, $t(\psi)$ in terms of elementary functions.

For the present *age of the universe* we get

$$t_0 = H_0^{-1}\int_0^1 \frac{\mathrm{d}a}{\sqrt{\Omega_{\mathrm{r}0}a^{-2}+\Omega_{\mathrm{m}0}a^{-1}+\Omega_{K0}+\Omega_{\Lambda 0}a^2}}\,. \tag{3.76}$$

If $\Omega_{\Lambda 0}$ does not overwhelm the other coefficients, the value of the integral is of order unity (recall the sum rule $\Omega+\Omega_K = \Omega_\mathrm{m}+\Omega_\mathrm{r}+\Omega_\Lambda+\Omega_K = 1$). So the age of the universe is of the order of the Hubble time. In the real universe, $\Omega_{\mathrm{r}0}\approx 10^{-4}$, so dropping the radiation term causes negligible error (physically, this is because that the radiation-dominated era is short compared to the matter- and vacuum-dominated eras).

Example: Age of the open universe. Let us consider an open universe ($K<0$ or $\Omega_0<1$) without vacuum energy or radiation ($\Omega_\Lambda=\Omega_\mathrm{r}=0$). Integrating (3.76) (e.g. with the substitution $a=\frac{\Omega_{\mathrm{m}0}}{1-\Omega_{\mathrm{m}0}}\sinh^2\frac{\psi}{2}$) gives the age of the open universe as

$$\begin{aligned} t_0 &= H_0^{-1}\int_0^1 \frac{da}{\sqrt{1-\Omega_{\mathrm{m}0}+\Omega_{\mathrm{m}0}a^{-1}}} \\ &= H_0^{-1}\left[\frac{1}{1-\Omega_{\mathrm{m}0}} - \frac{\Omega_{\mathrm{m}0}}{2(1-\Omega_{\mathrm{m}0})^{3/2}}\,\mathrm{arcosh}\left(\frac{2}{\Omega_{\mathrm{m}0}}-1\right)\right]. \end{aligned} \tag{3.77}$$

A special case of the open universe is the completely empty universe, which is dominated by the spatial curvature, with $\Omega_\mathrm{m}=\Omega_\Lambda=0$ and $\Omega_K=1$. In this case we obtain from (3.73) the result $a=H_0t$, and we have $t_0=H_0^{-1}$. We get the following table for the age of the universe:

$\Omega_{\mathrm{m}0}$	$\Omega_{\Lambda 0}$	t_0H_0
0	0	1
0.1	0	0.90
0.3	0	0.81
0.5	0	0.75
1	0	2/3

The case ($\Omega_\mathrm{m}>1$, $\Omega_\Lambda=0$) is left as an exercise. The more general case ($\Omega_K\neq 1$, $\Omega_\Lambda\neq 0$) leads to elliptic functions. The results for H_0t_0 are plotted in figure 4.

Example: Expansion of the ΛCDM model. In the ΛCDM model $\Omega_K=0$, and at late times Ω_r can be neglected in the expansion law. From (3.75) we have

$$\begin{aligned} t(a) &= H_0^{-1}\int_0^a \frac{\mathrm{d}a}{\sqrt{\Omega_{\mathrm{m}0}a^{-1}+\Omega_{\Lambda 0}a^2}} \\ &= H_0^{-1}\int_0^a \frac{\mathrm{d}a\sqrt{a}}{\sqrt{\Omega_{\mathrm{m}0}+\Omega_{\Lambda 0}a^3}} \\ &= \frac{2}{3}\frac{1}{\sqrt{\Omega_{\Lambda 0}}}H_0^{-1}\int_0^y \frac{\mathrm{d}x}{\sqrt{1+x^2}} \\ &= \frac{2}{3}\frac{1}{\sqrt{\Omega_{\Lambda 0}}}H_0^{-1}\mathrm{arsinh}\left[\sqrt{\frac{\Omega_{\Lambda 0}}{\Omega_{\mathrm{m}0}}}a^{3/2}\right], \end{aligned} \tag{3.78}$$

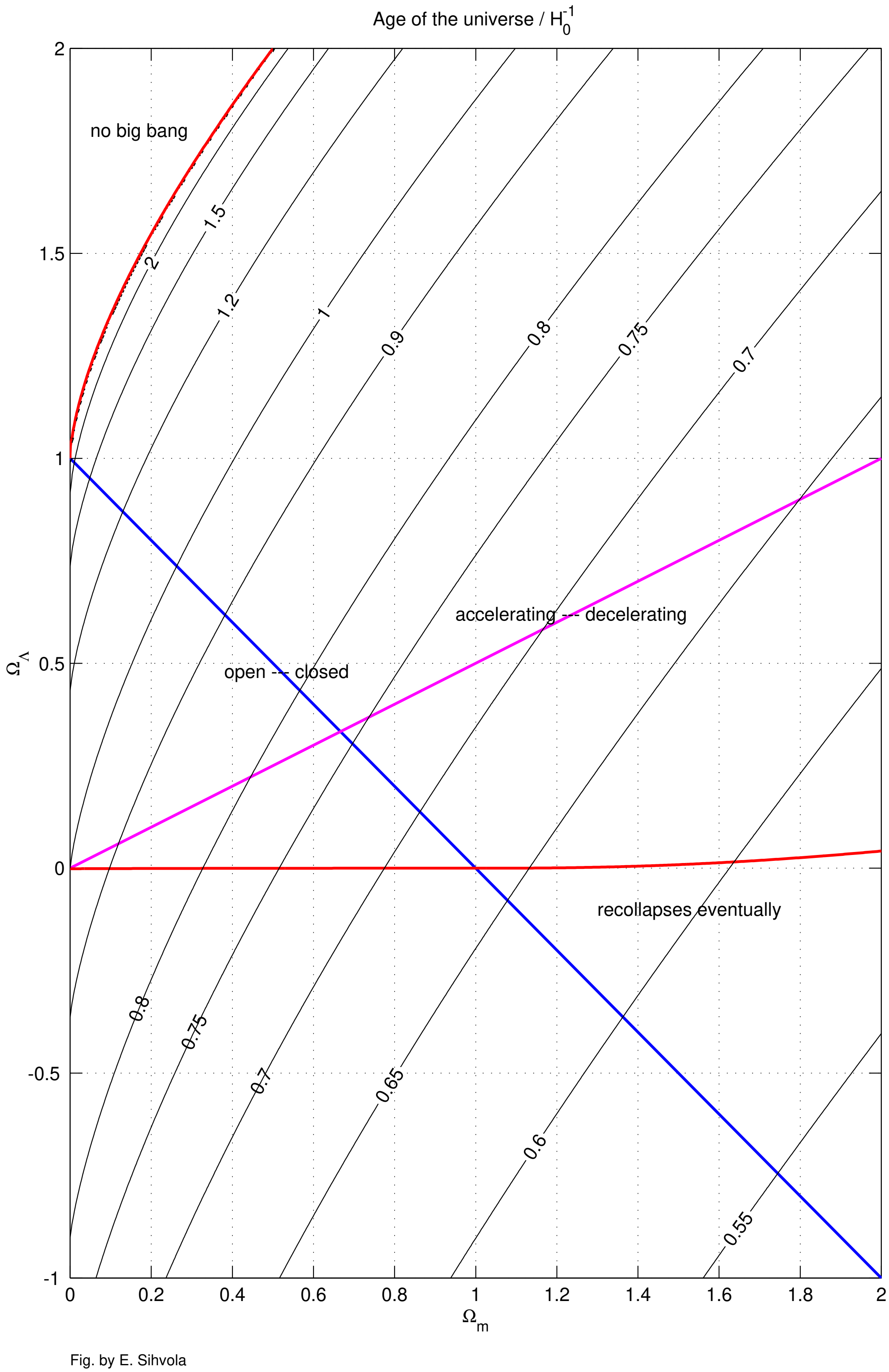

Figure 5: The age of the universe as a function of Ω_{m0} and $\Omega_{\Lambda 0}$.

where we have used the change of variables $y = \sqrt{\Omega_{\Lambda 0}/\Omega_{m0}} a^{3/2}$. Inverting, we get the scale factor as a function of time,

$$a(t) = \left(\frac{\Omega_{m0}}{\Omega_{\Lambda 0}}\right)^{1/3} \sinh^{2/3}\left(\frac{3}{2}\sqrt{\Omega_{\Lambda 0}} H_0 t\right) . \tag{3.79}$$

At early times, $t \ll H_0^{-1}$, the universe expands as in the matter-dominated case, $a \propto t^{2/3}$. At late times, $t \gg H_0^{-1}$, the expansion is exponential, $a \propto e^{\sqrt{\Omega_{\Lambda 0}} H_0 t}$. The value of the cosmological constant is related to the density parameter $\Omega_{\Lambda 0}$ as $\Omega_{\Lambda 0} = \frac{\Lambda}{3H_0^2}$, so the expansion at late times can be written as $a \propto e^{\sqrt{\frac{\Lambda}{3}} t}$. This expansion law (together with radiation at early times, and the treatment of deviations from homogeneity and isotropy, which we will discuss in the second part of the course) fits wide variety of cosmological data very well. However, when fitting it to the CMB observations and other measurements related to early times in the universe, gives a smaller value of H_0 than determining H_0 from nearby observations, like the local Hubble law, as discussed in chapter 1. We will discuss how H_0 is determined from the CMB in the second part of the course.

Exercise: In the ΛCDM model, calculate when (in terms of redshift and time) we have $\Omega_\Lambda = \Omega_m$. Calculate at which redshift and time the universe transitions from deceleration to acceleration, $\ddot{a} = 0$. What is the age of the universe today, in units of H_0^{-1}?

Measurements of the age of the universe have provided important evidence against the Einstein-de Sitter model, and early indications for vacuum energy in the 1990s. For example, a model-independent estimate of the age of the universe based on ages of globular clusters, which are compact groups of stars in our galaxy, give a 95% probability lower limit on the age of the universe of 11 Gyr, and a best-fit age of about 13.4 Gyr [5]. The Hubble time is $H_0^{-1} \approx h^{-1} 9.8$ Gyr, so we get $H_0 t_0 \gtrsim 1.14h$ as the lower limit, and $H_0 t_0 \approx 1.37h$ as the preferred value. So the age of the universe implies that models with only matter and curvature need a small Hubble parameter. For a spatially flat matter model, we would need $h = 0.48$, and an open model with $\Omega_{m0} \approx 0.3$ (as indicated by observations) would need $h \approx 0.6$. The measured value $h = 0.7$ gives $H_0 t_0 = 0.96$, using $t_0 = 13.4$ Gyr. (Using the CMB value $h = 0.67$ gives $H_0 t_0 = 0.92$, while the value from local distance measurements $h = 0.73$ gives $H_0 t_0 = 1.00$.) So just from measurements of H_0 and t_0 we see that models with only spatial curvature and matter have trouble fitting the observations. However, the strongest evidence against a model with no vacuum energy (or other form of dark energy) comes from distance measurements.

3.2.7 Distances in the universe

Earlier, we derived the angular diameter distance, luminosity distance and proper distance for a general FLRW spacetime. We can now plug the parametrised expansion rate $H(z)$ into these expressions, like we did for the age of the universe. The comoving proper distance to a comoving object seen at redshift z is, from (3.42) and

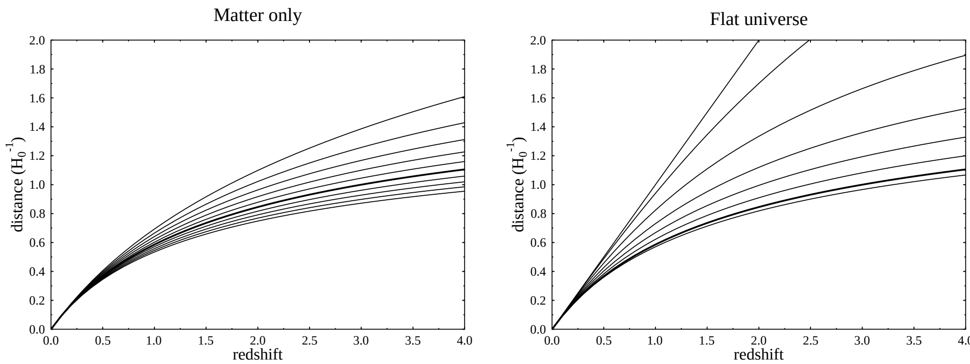

Figure 6: The proper distance (3.80) for a) the matter-only universe $\Omega_\Lambda = 0$, $\Omega_{\rm m0} = 0$, 0.2,. . . ,1.8 (from top to bottom) b) the spatially flat universe $\Omega = 1$ ($\Omega_\Lambda = 1 - \Omega_{\rm m}$), $\Omega_{\rm m0} = 0$, 0.05, 0.2, 0.4, 0.6, 0.8, 1.0, 1.05 (from top to bottom). The thick line in both cases is the Einstein–de Sitter model with $\Omega_m = 1$, $\Omega_\Lambda = 0$.

(3.73),

$$\begin{aligned} D_P^c(z) &= \int_0^z \frac{\mathrm{d}z'}{H(z')} \\ &= H_0^{-1} \int_0^z \frac{\mathrm{d}z'}{\sqrt{\Omega_{\rm r0}(1+z')^4 + \Omega_{\rm m0}(1+z')^3 + \Omega_{K0}(1+z')^2 + \Omega_{\Lambda0}}} \\ &= H_0^{-1} \int_{\frac{1}{1+z}}^1 \frac{\mathrm{d}a}{\sqrt{\Omega_{\rm r0} + \Omega_{\rm m0}a + \Omega_{K0}a^2 + \Omega_{\Lambda0}a^4}} \\ &\approx H_0^{-1} \int_{\frac{1}{1+z}}^1 \frac{\mathrm{d}a}{\sqrt{\Omega_0(a-a^2) - \Omega_{\Lambda0}(a-a^4) + a^2}} \ , \end{aligned} \tag{3.80}$$

where on the last line we have dropped the $\Omega_{\rm r0}$ term which has negligible effect, and used $\Omega_{\rm m0} = \Omega_0 - \Omega_{\Lambda0}$. The proper distance depends on three independent cosmological parameters, for which we have taken H_0, Ω_0 and $\Omega_{\Lambda0}$, and the distance at a given redshift is proportional to the Hubble distance, H_0^{-1}. If we give the distance in units of H_0^{-1}, then it depends only on Ω_0 and $\Omega_{\Lambda0}$.

If we increase Ω_0 while keeping $\Omega_{\Lambda0}$ constant (meaning that we increase $\Omega_{\rm m0}$), the distance corresponding to a given redshift decreases. This is because the universe has expanded faster in the past, so there is less time between a given value of the scale factor $a = 1/(1+z)$ and the present. The distance to the galaxy with redshift z is shorter because photons have had less time to travel. Whereas if we increase $\Omega_{\Lambda0}$ with a fixed Ω_0 (meaning that we decrease $\Omega_{\rm m0}$), we have the opposite situation and the distance increases. In figure 6 we show the proper distance for some parameter values.

In the case $\Omega_\Lambda = 0$, we have

$$D_P^c(z) = H_0^{-1} \frac{2}{\Omega_{\rm m0}^2(1+z)} \left(\Omega_{\rm m0}z - (2-\Omega_{\rm m0})(\sqrt{1+\Omega_{\rm m0}z} - 1) \right) \ . \tag{3.81}$$

A subcase of this is the Einstein–de Sitter universe, which has $\Omega = \Omega_{\rm m} = 1$, $\Omega_K =$

$\Omega_\Lambda = 0$,

$$D_P^c(z) = 2H_0^{-1}\left(1 - \frac{1}{\sqrt{1+z}}\right) . \tag{3.82}$$

The comoving horizon distance today is

$$D_{\rm hor}^c = H_0^{-1}\int_0^1 \frac{{\rm d}a}{\sqrt{\Omega_{\Lambda 0}a^4 + (1-\Omega_0)a^2 + \Omega_{\rm m0}a + \Omega_{\rm r0}}} . \tag{3.83}$$

In figure 7 the comoving horizon distance is plotted for various choices of parameters.

The angular diameter distance is, from (3.29) and (3.80),

$$\begin{aligned} D_A(z) &= (1+z)^{-1}\frac{1}{\sqrt{-K}}\sinh\left(\sqrt{-K}\int_0^z \frac{{\rm d}z'}{H(z')}\right) \\ &= H_0^{-1}(1+z)^{-1}\frac{1}{\sqrt{\Omega_{K0}}}\sinh\left(\sqrt{\Omega_{K0}}\int_{\frac{1}{1+z}}^1 \frac{da}{\sqrt{\Omega_{\rm r0} + \Omega_{\rm m0}a + \Omega_{K0}a^2 + \Omega_{\Lambda 0}a^4}}\right) \\ &\approx H_0^{-1}(1+z)^{-1}\frac{1}{\sqrt{1-\Omega_0}}\sinh\left(\int_{\frac{1}{1+z}}^1 {\rm d}a \frac{\sqrt{1-\Omega_0}}{\sqrt{\Omega_0(a-a^2) - \Omega_{\Lambda 0}(a-a^4) + a^2}}\right) , \end{aligned} \tag{3.84}$$

where we have used the definition $\Omega_{K0} = -K/H_0^2 = 1 - \Omega_0$ and have on the last line again dropped $\Omega_{\rm r0}$. The angular diameter distance is plotted in figure 8 for some values of the parameters; figure 9 shows the same plot for the comoving angular diameter distance. As always, the luminosity distance is $D_L = (1+z)^2 D_A$.

In a spatially flat universe the angular diameter distance is equal to the proper distance,

$$\begin{aligned} D_A^c(z) &= D_P^c(z) \\ &= \int_0^z \frac{{\rm d}z'}{H(z')} \\ &= H_0^{-1}\int_{\frac{1}{1+z}}^1 \frac{{\rm d}a}{\sqrt{\Omega_{\rm r0} + \Omega_{\rm m0}a + \Omega_{K0}a^2 + \Omega_{\Lambda 0}a^4}} . \end{aligned} \tag{3.85}$$

From anisotropies of the CMB we can infer rather model-independently that $D_A(1090) = 13$ Mpc [6]. In the second part of course we will discuss in detail where this length scale comes from. But given this number, it is simple to use it as a cosmological constraint. The parameters of any model have to be such that this distance is reproduced – fitting a specific model directly to the data instead of using model-independent numbers will give more precise constraints for that model.

Exercise. Show that in order to fit this distance in the Einstein–de Sitter model, the Hubble constant has to be smaller than the observed value $h = 0.7$.

3.2.8 Illustrating the distances

Any flat map of the surface of the Earth is necessarily distorted, and the same holds for plots of curved spacetime. Even in the simplest spatially flat case, the expansion rate affects the mapping. Thus any spacetime diagram is a distortion of

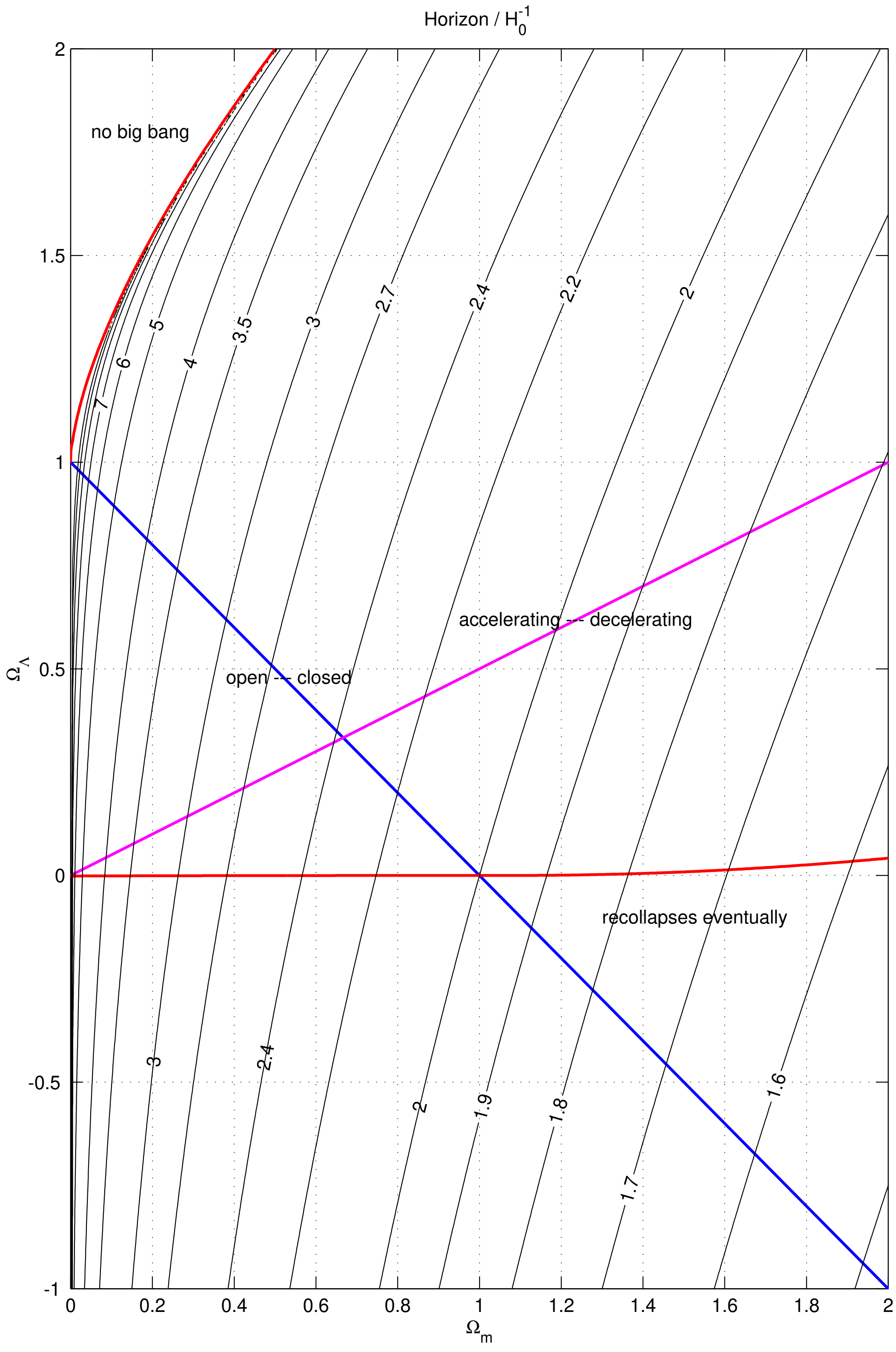

Fig. by E. Sihvola

Figure 7: The comoving horizon as a function of Ω_{m0} and $\Omega_{\Lambda 0}$.

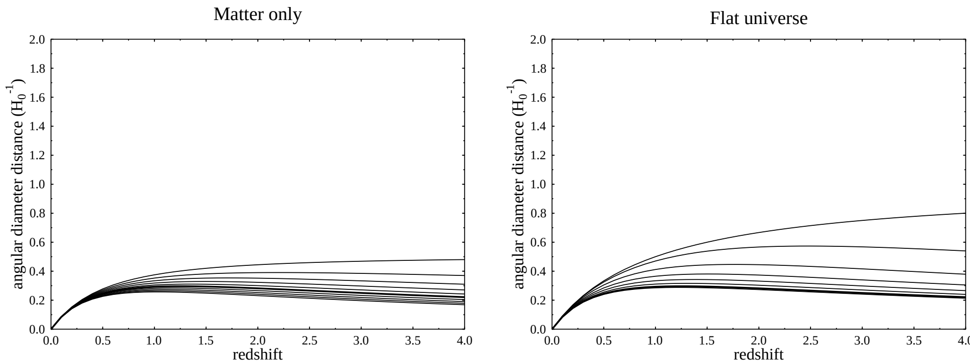

Figure 8: The angular diameter distance, for a) the matter-only universe $\Omega_\Lambda = 0$, $\Omega_{m0} = 0$, 0.2,…,1.8 (from top to bottom); b) the spatially flat universe $\Omega = 1$ ($\Omega_\Lambda = 1 - \Omega_m$), $\Omega_{m0} = 0$, 0.05, 0.2, 0.4, 0.6, 0.8, 1.0, 1.05 (from top to bottom). The thick line in both cases is the Einstein–de Sitter model with $\Omega_m = 1$, $\Omega_\Lambda = 0$. Note how the angular diameter distance decreases for large redshifts, meaning that the object that is farther away may appear larger on the sky. In the spatially flat case, this is an expansion effect. In the matter-only case, the effect is enhanced by spatial curvature in the closed ($\Omega_m > 1$) models.

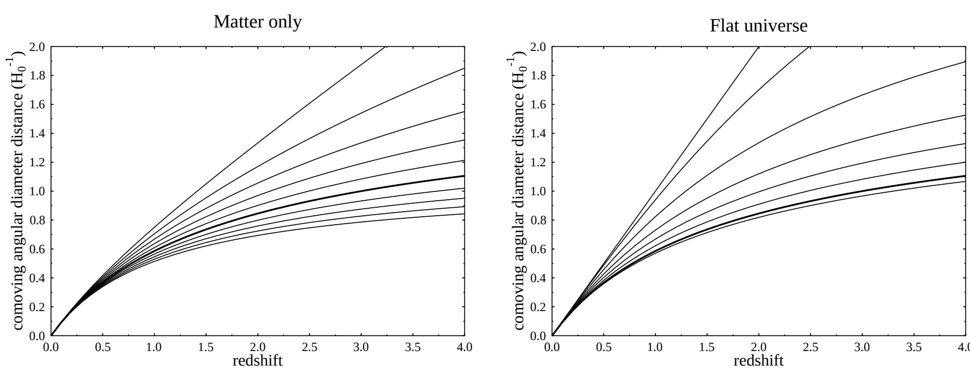

Figure 9: Same as figure 8, but for the *comoving* angular diameter distance. For the closed models (for $\Omega_m > 1$ in the case of $\Omega_\Lambda = 0$) even the comoving angular diameter distance may begin to decrease if we look at large enough redshifts. This happens when we are looking beyond $\chi = \pi/2$, where the universe "begins to close up'" as we pass the equator of the hypersphere. The figure does not go to high enough z to show this for the parameters used. Note how for the flat universe the comoving angular diameter distance is equal to the comoving distance (see figure 6).

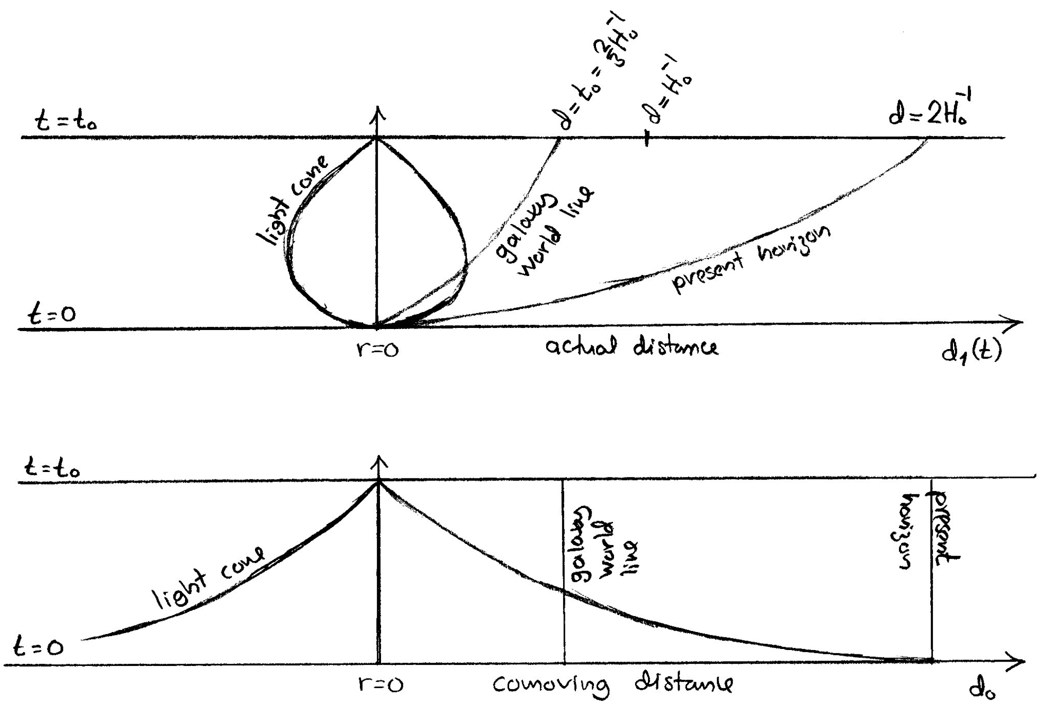

Figure 10: Spacetime diagrams for a flat universe giving a) the actual distance b) the comoving distance from origin as a function of cosmic time.

the true situation. In figures 10 and 11 there are three different ways of drawing the same spacetime diagram for the simplest cosmological model, the Einstein–de Sitter model which has $\Omega_m = 1$. In the first one the vertical distance is proportional to the cosmic time t, the horizontal distance to the actual distance at that time, d_1. The second one is in the comoving coordinates (t, r), so that the horizontal distance is proportional to the comoving proper distance D_P^c. (Recall that for in the case $K = 0$ we have $D_P^c = r$, see (3.40).) The third one uses the conformal coordinates (η, r). The last one has the advantage that light cones are always at a 45° angle. It is a spacetime analogue of the Mercator projection.

3.2.9 Luminosity distance and observations

Using the luminosity distance to constrain the cosmological model, we would ideally have a set of *standard candles*, objects that have the same absolute luminosity L. From their observed redshifts z and fluxes $F(z)$ we would then get an observed luminosity-distance-redshift relationship $D_L(z) = \sqrt{L/4\pi F}$, to be compared to the theoretical predictions to find the values of the cosmological parameters which give the best fit between theory and observations.

In astronomy, luminosity is often expressed in terms of *magnitude*. This system hails back to the ancient Greeks, who classified stars visible to the naked eye into six classes according to their brightness. Magnitude in modern astronomy is defined so that it roughly matches this classification, but it is not restricted to positive integers. The magnitude scale is logarithmic in such a way that a difference of 5

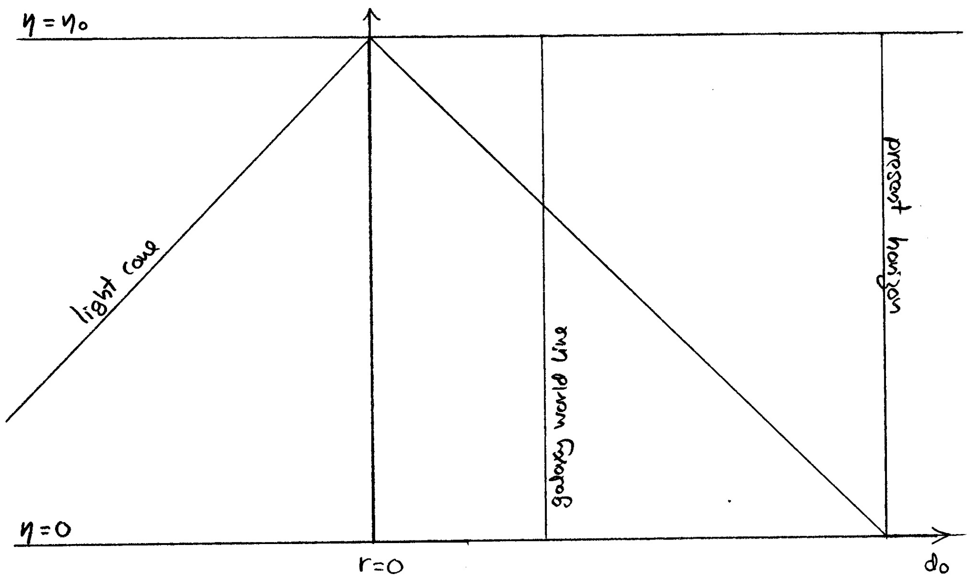

Figure 11: Spacetime diagram for a flat universe in conformal coordinates.

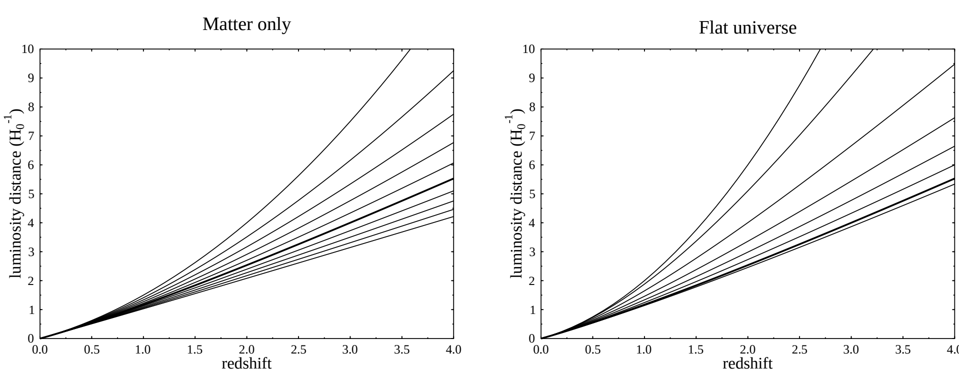

Figure 12: Same as figure 8, but for the *luminosity* distance. Note how the vertical scale now extends to 10 Hubble distances instead of just 2, to have room for the much more rapidly increasing luminosity distance.

magnitudes corresponds to a factor of 100 in luminosity[6]. The *absolute magnitude* M and *apparent magnitude* m of an object are defined as

$$M \equiv -2.5 \log_{10} \frac{L}{L_0} , \qquad m \equiv -2.5 \log_{10} \frac{F}{F_0} , \tag{3.86}$$

so from (3.32) we get the distance modulus $m - M$ in terms of the luminosity distance:

$$\begin{aligned} m - M &= -2.5 \log_{10} \left(\frac{F}{L} \frac{L_0}{F_0} \right) \\ &= 5 \log_{10} D_L + 2.5 \log_{10} \left(4\pi \frac{F_0}{L_0} \right) \\ &= -5 + 5 \log_{10} \frac{D_L}{\text{pc}} , \end{aligned} \tag{3.87}$$

where L_0 and F_0 are reference luminosity and flux. There are actually different magnitude scales corresponding to different regions of the electromagnetic spectrum, with different reference luminosities. The *bolometric* magnitude and luminosity refer to the power or flux integrated over all frequencies, whereas the *visual* magnitude and luminosity refer only to the visible light. In the bolometric magnitude scale $L_0 = 3.0 \times 10^{28}$ W. The reference flux F_0 for the apparent scale has been chosen so in relation to the absolute scale that a star whose distance is $D = 10$ pc has $m = M$. For a set of standard candles, all having the same absolute magnitude M, their apparent magnitudes m should be related to their redshift z as

$$\begin{aligned} m(z) &= M - 5 + 5 \log_{10} \frac{D_L}{\text{pc}} \\ &= M - 5 - 5 \log_{10} H_0 + 5 \log_{10} \left[(1+z) \frac{1}{\sqrt{1-\Omega_0}} \times \right. \\ & \left. \times \sinh \left(\int_{\frac{1}{1+z}}^{1} \mathrm{d}a \frac{\sqrt{1-\Omega_0}}{\sqrt{\Omega_0 (a - a^2) - \Omega_{\Lambda 0} (a - a^4) + a^2}} \right) \right] . \end{aligned} \tag{3.88}$$

The Hubble constant H_0 contributes only a constant of proportionality to this *magnitude-redshift relationship*. If we just know that all the objects have the same M, but do not know the value of M, we cannot use the observed $m(z)$ to determine H_0, since both M and H_0 contribute to this constant term. On the other hand, the *shape* of the $m(z)$ curve depends only on the parameters Ω_0 and $\Omega_{\Lambda 0}$.

Unfortunately, there are no known good standard candles. However, the absolute peak luminosity of type Ia supernovae (SNe Ia) is correlated with the shape of the observed luminosity as a function of time. Therefore, calibrating off nearby supernovae whose distance can be determined independently, it is possible to find the absolute luminosity of individual SNe Ia, making them reliable distance indicators[7].

Observations of SNe Ia published in 1998 provided the strong evidence that the expansion of the universe had accelerated. (There were a number of hints already

[6] So a difference of 1 magnitude corresponds to a factor of $100^{1/5} \approx 2.512$ in luminosity.

[7] The variation in luminosity is about a factor of ten, so type Ia supernovae are far from standard candles, though they are often incorrectly referred to as such. Another, less incorrect, expression is "standardisable candle'".

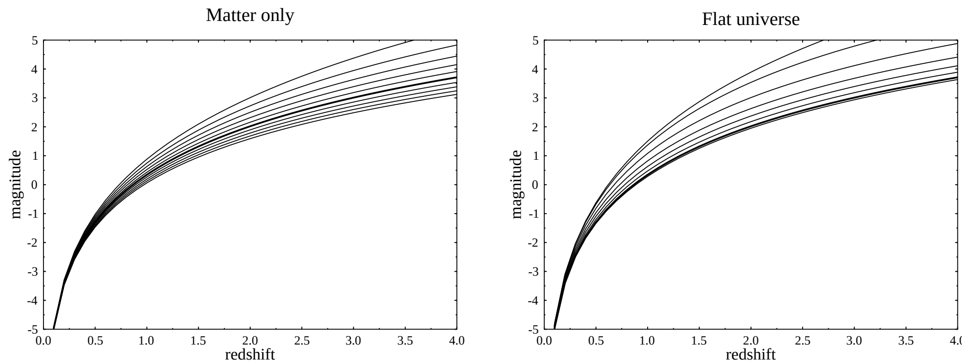

Figure 13: Same as figure 8, bur for the *magnitude-redshift* relationship. The constant $M - 5 - 5\log_{10} H_0$ in (3.88), which is different for different classes of standard candles, has been arbitrarily set to 0.

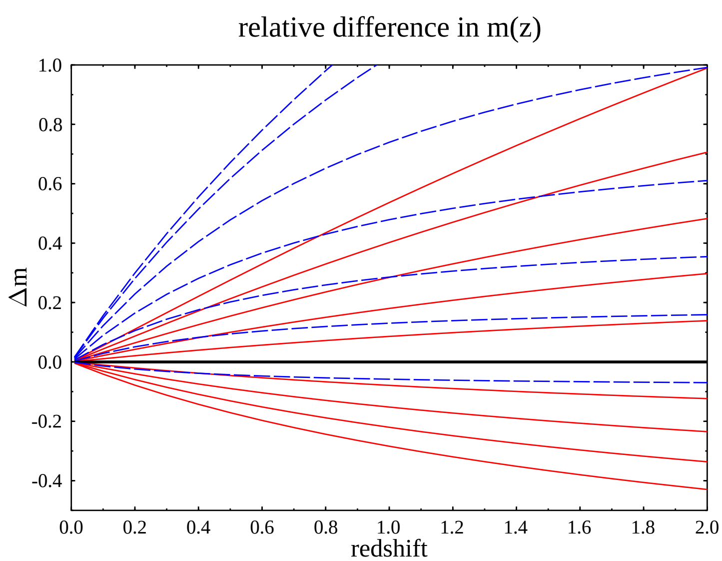

Figure 14: The difference between the magnitude-redshift relationship of the different models in figure 13 from the reference model $\Omega_{\rm m} = 1$, $\Omega_\Lambda = 0$ (which appears as the horizontal thick line). The red (solid) lines are for the matter-only ($\Omega_\Lambda = 0$) models and the blue (dashed) lines are for the flat ($\Omega_0 = 1$) models.

earlier –such as the age of the universe coupled with the value of H_0– that something was wrong with the SCDM model.) Two groups, the Supernova Cosmology Project and the High-Z Supernova Search Team made independent observations and analyses of SNe Ia up to redshifts $z \sim 1$ to determine the values of the cosmological parameters Ω_0 and $\Omega_{\Lambda 0}$ [7, 8]. Their observations were inconsistent with a decelerating matter- or curvature-dominated universe, i.e. with $\Omega_\Lambda = 0$. Instead, they found that the expansion of the universe accelerates.

Later, more accurate observations by these and other groups have confirmed this result. The SNIa data is one of the main arguments for the existence of dark energy in the universe, supported by many other observations. Figure 15 shows determination of $\Omega_{\mathrm{m}0}$ and $\Omega_{\Lambda 0}$ from more recent data, in 2022. The coloured blob is Pantheon+ SNIa data, while the lines marked SDSS and Planck use data from large-scale structure and the CMB, respectively, requiring more assumptions than just the homogeneous and isotropic FLRW metric.

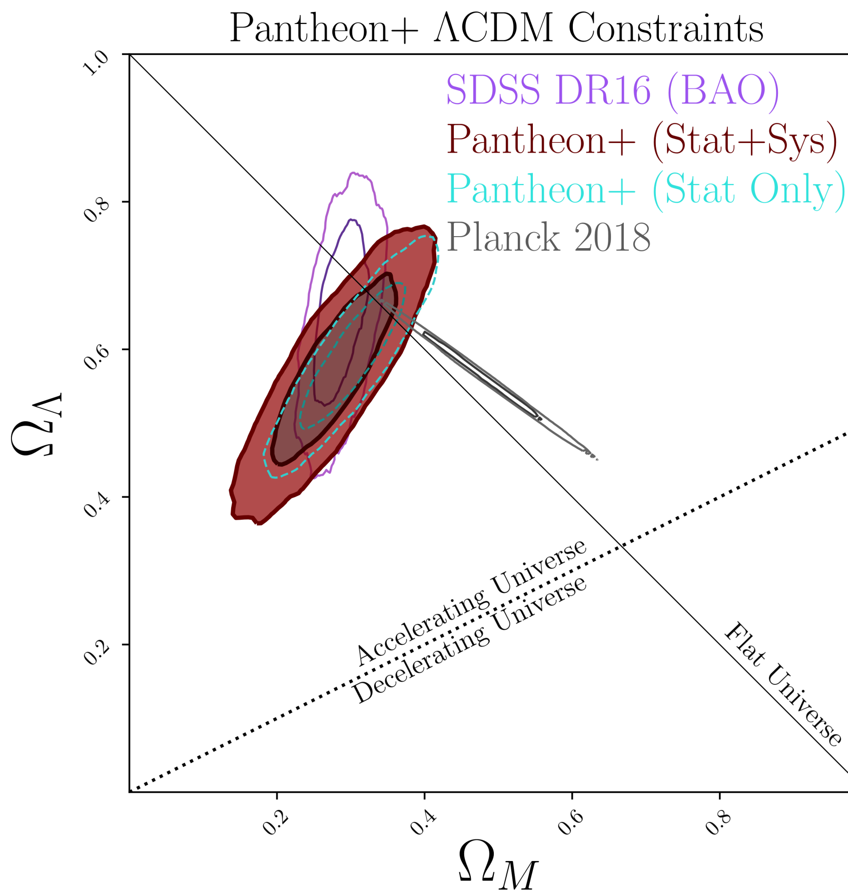

Figure 15: The densities $\Omega_{\mathrm{m}0}$ and $\Omega_{\Lambda 0}$ determined from Pantheon+ SNIa data. This figure is from [9].

We have in the preceding assumed that the mysterious dark energy component of the universe is vacuum energy, for which $p_{\mathrm{de}} = -\rho_{\mathrm{de}}$. Instead allowing the equation of state parameter $w_{\mathrm{de}} \equiv p_{\mathrm{de}}/\rho_{\mathrm{de}}$ for dark energy to be an arbitrary constant[8], we see that w_{de} is restricted to be close to -1; see figure 16.

It is worth emphasising that what the supernova observations (and observations of the angular diameter distance to the CMB) show is that the distances are longer than in the Einstein–de Sitter model. If the distance observations are interpreted

[8] If w is close to -1, it will be close to a constant. If it is far from -1, there is no theoretical reason for it to be a constant; it it just assumed here for simplicity.

assuming that the FLRW approximation is valid (i.e. that the FLRW relation (3.29) between the distance and the expansion rate holds), it follows that the expansion rate has accelerated. Assuming that the Friedmann equations hold (i.e. general relativity is valid), (3.47) shows that the total pressure then has to negative. Observations of t_0 and H_0 (and other observables, such as the growth rate of cosmic structures) are consistent with this interpretation. There are now also direct observations of $H(z)$ independent of the distances, which support accelerated expansion [10]. Note that the only cosmological effect of vacuum energy is to increase the expansion rate and correspondingly increase the distances. Its success in fitting various cosmological observations is thus strong evidence for faster expansion, but it may be that the explanation for the faster expansion is not vacuum energy but something else, be it a more complicated form of dark energy, modified gravity or a breakdown of the FLRW approximation due to cosmological structure formation. In this course, we will not discuss these possibilities, and will stick with vacuum energy.

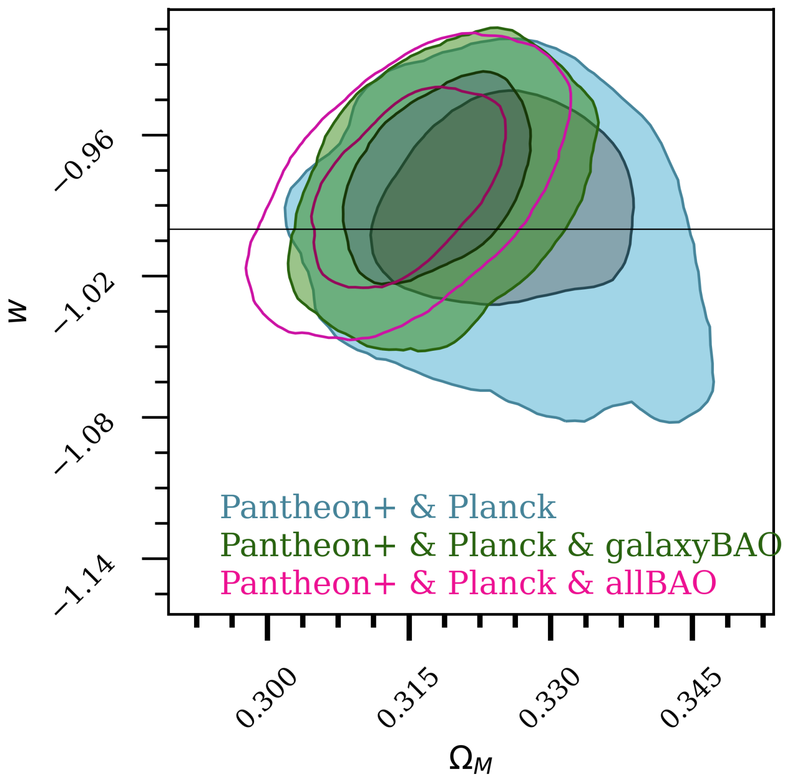

Figure 16: The matter density $\Omega_{\rm m0}$ and the dark energy equation of state w determined from SNIa, CMB and large-scale structure data, assuming spatial flatness. This figure is from [9].

References

[1] S. Räsänen, *On the relation between the isotropy of the CMB and the geometry of the universe*, Phys. Rev. D **79** (2009) 123522, [0903.3013].

[2] R. Maartens, *Is the Universe homogeneous?*, *Phil. Trans. Roy. Soc. Lond. A* **369** (2011) 5115–5137, [1104.1300].

[3] S. Nadathur, *Seeing patterns in noise: Gigaparsec-scale 'structures' that do not violate homogeneity*, *Mon. Not. Roy. Astron. Soc.* **434** (2013) 398–406, [1306.1700].

[4] Planck collaboration, N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, [1807.06209].

[5] L. M. Krauss and B. Chaboyer, *Age Estimates of Globular Clusters in the Milky Way: Constraints on Cosmology*, *Science* **299** (2003) 65–70.

[6] M. Vonlanthen, S. Räsänen and R. Durrer, *Model-independent cosmological constraints from the CMB*, *JCAP* **08** (2010) 023, [1003.0810].

[7] Supernova Search Team collaboration, A. G. Riess et al., *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, *Astron. J.* **116** (1998) 1009–1038, [astro-ph/9805201].

[8] Supernova Cosmology Project collaboration, S. Perlmutter et al., *Measurements of Ω and Λ from 42 High Redshift Supernovae*, *Astrophys. J.* **517** (1999) 565–586, [astro-ph/9812133].

[9] D. Brout et al., *The Pantheon+ Analysis: Cosmological Constraints*, *Astrophys. J.* **938** (2022) 110, [2202.04077].

[10] N. Borghi, M. Moresco and A. Cimatti, *Toward a Better Understanding of Cosmic Chronometers: A New Measurement of H(z) at $z \sim 0.7$*, *Astrophys. J. Lett.* **928** (2022) L4, [2110.04304].