2 Basics of general relativity

The general theory of relativity completed by Albert Einstein in 1915 (and nearly simultaneously by David Hilbert) is the current theory of gravity. General relativity replaced the previous theory of gravity, Newtonian gravity, which can be understood as a limit of general relativity in the case of isolated systems, slow motions and weak fields. General relativity has been extensively tested during the past century, and no deviations have been found, with the possible exception of the accelerated expansion of the universe, which is however usually explained by introducing new matter rather than changing the laws of gravity [1]. We will not go through the details of general relativity, but we try to give some rough idea of what the theory is like, and introduce a few concepts and definitions that we are going to need.

The principle behind special relativity is that space and time together form four-dimensional spacetime. The essence of general relativity is that gravity is a manifestation of the *curvature* of spacetime. While in Newton's theory gravity acts directly as a force between two bodies[1], in Einstein's theory the gravitational interaction is mediated by the spacetime. In other words, gravity is an aspect of the geometry of spacetime. Matter curves the surrounding spacetime. This curvature then affects the motion of other matter (as well as the motion of the matter generating the curvature). This can be summarised as the dictum "matter tells spacetime how to curve, spacetime tells matter how to move" [2]. From the viewpoint of general relativity, gravity is not a force; if there are no forces (which could be due to particle physics interactions) acting on a body, the body is in *free fall*. A freely falling body moves along a straight line in the curved spacetime, called a *geodesic*. Forces cause the body to deviate from geodesic motion. It is important to remember that the viewpoint is that of *spacetime*, not just space. For example, the orbit of the earth around the sun is curved in space, but straight in spacetime.

If a spacetime is not curved, it is said to be *flat*, which just means that it has the geometry of Minkowski space. In the case of space (as opposed to spacetime), "flat" means that the geometry is Euclidean. (Note the possibly confusing terminology: Minkowski space*time* is called simply Minkowski space!)

To define a physical theory, we should give 1) the kinematics of the theory (closely related to the symmetry properties), 2) the degrees of freedom and 3) the laws that determine the time evolution of the degrees of freedom, consistent with the kinematics (in other words the dynamics). In Newtonian gravity, the kinematics is that of Euclidean space with the Galilean symmetry group, which is to say that the laws of physics are invariant under the transformation

$$\begin{aligned} x^i &\rightarrow x'^i = R^i{}_j x^j + A^i + v^i t \\ t &\rightarrow t' = Bt + C \ , \end{aligned} \tag{2.1}$$

where x^i are spatial coordinates, t is time, $R^i{}_j$ is a constant rotation matrix, A^i and v^i are constant vectors and B and C are constants. (Summation over repeated indices is implied; see section 2.5.) The degrees of freedom are point particles, and the dynamics is given by Newton's second law with the law of gravity, which states

[1] The way Newtonian gravity is usually formulated. It is also possible to formulate Newtonian gravity in geometric terms, so that gravity is an expression of spacetime curvature, although this is less natural than in general relativity.

that the acceleration of particle 1 due to particle 2 is

$$\ddot{\bar{x}}_1 = -G_{\rm N} m_2 \frac{\bar{x}_1 - \bar{x}_2}{|\bar{x}_1 - \bar{x}_2|^3} \ , \tag{2.2}$$

where $G_{\rm N}$ is Newton's constant and m_2 is the mass of particle 2. This law is consistent with the symmetry (2.1), but it is not uniquely specified by it.

In general relativity, Euclidean space is replaced by curved spacetime. Unlike Euclidean space or Minkowski space, a general curved spacetime has no symmetries. However, in general relativity a central role is played by *diffeomorphism invariance*, which is to say invariance under general coordinate transformations, $x^\alpha \to x'^\alpha(x^\beta)$. We use Greek indices to denote directions in spacetime, they can take any of the values 0, 1, 2, 3. Latin indices are used to denote spatial directions, i can have any of the values 1, 2, 3. In addition to the matter degrees of freedom, (which are more complicated than point particles, and have to be specified by a matter model –general relativity is not a theory about the structure of matter!), there are gravitational degrees of freedom. In Newtonian theory, gravity is just an interaction between particles, but in general relativity, it is an aspect of the geometry of spacetime and its degrees of freedom are described by the *metric*. The equation of motion is the Einstein equation. We will below first go through some kinematics of curved spacetime, and then briefly discuss the Einstein equation and its relation to Newtonian gravity.

2.1 Curved 2D and 3D space

(If you are familiar with the concept of curved space and how its geometry is given by the metric, you can skip the following and go straight to section 2.3.)

To help to visualise a four-dimensional curved spacetime, it may be useful to consider curved two-dimensional spaces embedded in a flat three-dimensional space.[2] So let us consider first a 2D space. Imagine there are 2D beings living in this 2D space. They have no access to a third dimension. How can they determine whether the space they live in is curved? By examining whether the laws of Euclidean geometry hold. If the space is flat, then the sum of the angles of any triangle is 180°, and the circumference of any circle with radius χ is $2\pi\chi$. If by measurement they find that this does not hold for some triangles or circles, then they can conclude that the space is curved.

A simple example of a curved 2D space is the sphere. The sum of angles of any triangle on a sphere is greater than 180°, and the circumference of any circle drawn on the surface of a sphere is less than $2\pi\chi$. (Straight lines on the sphere are sections of *great circles*, which divide the sphere into two equal hemispheres.)

In contrast, the surface of a cylinder has Euclidean geometry, i.e. there is no way that 2D beings living on it could conclude that it differs from a flat surface, and thus by our definition it is a flat 2D space. (By travelling around the cylinder

[2]This embedding is only a visualisation aid. A curved 2D space is defined completely in terms of its two independent coordinates, without any reference to a higher dimension. The geometry is given by the metric (part of the definition of the 2D space), which is a function of these coordinates. Some such curved 2D spaces have the same geometry as a 2D surface in flat 3D space. We then say that the 2D space can be embedded in flat 3D space. But there are curved 2D spaces which have no such corresponding surface, i.e. not all curved 2D spaces can be embedded in flat 3D space.

Figure 1: Cylinder and sphere.

they could conclude that their space has a non-trivial *topology*, but the geometry is anyway flat.)

In a similar manner we could try to determine whether the 3D space around us is curved, by measuring whether the sum of angles of a triangle is 180° or whether a sphere with radius r has surface area $4\pi r^2$. The space around Earth is indeed curved due to Earth's gravity, but the curvature is so small that more sophisticated measurements than the ones described above are needed to detect it.

2.2 The metric of 2D and 3D space

The tool to describe the geometry of space is the *metric*. The metric is given in terms of a set of coordinates. The coordinate system can be an arbitrary curved coordinate system. The coordinates are numbers which identify locations, but do not, by themselves, say anything about physical distances. The distance information is in the metric.

To introduce the concept of a metric, let us first consider Euclidean two-dimensional space with Cartesian coordinates x,y. Take a parametrised curve $x(\eta)$, $y(\eta)$ that begins at η_1 and ends at η_2. The length of the curve is

$$s = \int ds = \int \sqrt{dx^2 + dy^2} = \int_{\eta_1}^{\eta_2} \sqrt{x'^2 + y'^2} d\eta \,, \tag{2.3}$$

where $x' \equiv dx/d\eta$, $y' \equiv dy/d\eta$. Here $ds = \sqrt{dx^2 + dy^2}$ is the *line element*. The square of the line element, the *metric*, is

$$ds^2 = dx^2 + dy^2 \,. \tag{2.4}$$

The line element has the dimension of distance. As a working definition for the *metric*, we can say that *the metric is an expression which gives the square of the line element in terms of the coordinate differentials.*

We could use another coordinate system on the same 2-dimensional Euclidean space, e.g. polar coordinates. Then the metric is

$$ds^2 = dr^2 + r^2 d\varphi^2 \,, \tag{2.5}$$

Figure 2: A parametrised curve in Euclidean 2D space with Cartesian coordinates.

giving the length of a curve as

$$s = \int ds = \int \sqrt{dr^2 + r^2 d\varphi^2} = \int_{\eta_1}^{\eta_2} \sqrt{r'^2 + r^2 \varphi'^2} d\eta \,. \tag{2.6}$$

In a similar manner, in 3-dimensional Euclidean space, the metric is

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{2.7}$$

in Cartesian coordinates, and

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2 \tag{2.8}$$

in spherical coordinates (where the r coordinate has the dimension of distance, but the angular coordinates θ and φ are dimensionless).

Now we can go to our first example of a curved (2-dimensional) space, the sphere. Let the radius of the sphere be a. For the two coordinates on this 2D space we can take the angles θ and φ. We get the metric from the Euclidean 3D metric in spherical coordinates by setting $r = r_0$,

$$ds^2 = r_0^2 \left(d\theta^2 + \sin^2\theta d\varphi^2\right) \ . \tag{2.9}$$

The length of a curve $\theta(\eta), \varphi(\eta)$ on this sphere is given by

$$s = \int ds = \int_{\eta_1}^{\eta_2} r_0 \sqrt{\theta'^2 + \sin^2\theta \varphi'^2} d\eta \ . \tag{2.10}$$

For later application in cosmology, it is instructive to consider the coordinate transformation $r = \sin\theta$ (this new coordinate r has nothing to do with the earlier r of 3D space, it is a coordinate on the sphere growing in the same direction as θ, starting at $r = 0$ from the North Pole ($\theta = 0$)). Since now $dr = \cos\theta d\theta = \sqrt{1 - r^2} d\theta$, the metric becomes

$$ds^2 = \frac{dr^2}{1 - r^2} + r^2 d\varphi^2 \ . \tag{2.11}$$

Figure 3: A parametrised curve on a 2D sphere with spherical coordinates.

Figure 4: The part of the sphere covered by the coordinates in (2.11).

For $r \ll 1$ (in the vicinity of the North Pole), this metric is approximately the same as in (2.5), so on the "Arctic plane" the metric looks flat and the coordinates look like polar coordinates. As r grows, the deviation from flat geometry becomes more apparent. Note that we run into a problem when $r = 1$. This corresponds to $\theta = 90°$, i.e. the "equator". After this $r = \sin\theta$ begins to decrease again, repeating the same values. Also, at $r = 1$, the $1/(1 - r^2)$ factor in the metric becomes infinite. We say there is a *coordinate singularity* at the equator. There is nothing wrong with the space itself, but our chosen coordinate system applies only for a part of this space, the region "north" of the equator.

2.3 4D flat spacetime

The coordinates of the four-dimensional spacetime are (x^0, x^1, x^2, x^3), where $x^0 = t$ is a time coordinate. Some examples are "Cartesian" (t, x, y, z) and spherical (t, r, θ, φ) coordinates.

The metric of the Minkowski space of *special relativity* is

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \ , \tag{2.12}$$

in Cartesian coordinates. In spherical coordinates it is

$$ds^2 = -dt^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\varphi^2 \ . \tag{2.13}$$

The fact that time appears in the metric with a different sign reflects the special geometric features of Minkowski space. (We assume that the reader is familiar with

Figure 5: The light cone.

special relativity, and won't go into details.) There are three kinds of distance intervals,

- timelike, $ds^2 < 0$
- lightlike, $ds^2 = 0$
- spacelike, $ds^2 > 0$.

The lightlike directions form the observer's future and past *light cones*. Light moves along the light cone, so everything we see with light lies on our past light cone, and we can receive signals slower than light from everywhere inside it. To see us as we are now (using light), the observer has to lie on our future light cone, and we can send timelike signals to everywhere inside it. As we move in time along our world line, we drag our light cones with us so that they sweep over the spacetime. The motion of a massive body is always timelike, and the motion of massless particles is always lightlike.

2.4 Curved spacetime

These features of the Minkowski space are inherited by the spacetime of general relativity. However, spacetime is now *curved*, whereas Minkowski spacetime is flat. (Recall that when we say space is flat, we mean it has Euclidean geometry; when we say spacetime is flat, we mean it has Minkowski geometry.) The (proper) length of a spacelike curve is $\Delta s \equiv \int ds$. Light moves along lightlike world lines, $ds^2 = 0$, massive objects along timelike world lines $ds^2 < 0$. The time measured by a clock carried by the object, the *proper time*, is $\Delta\tau = \int d\tau$, where $d\tau \equiv \sqrt{-ds^2}$, so $d\tau^2 = -ds^2 > 0$. The proper time τ is a natural parameter for the world line, $x^\mu(\tau)$. The *four-velocity* of an object is defined as

$$u^\mu = \frac{dx^\mu}{d\tau}. \tag{2.14}$$

Figure 6: Two coordinate systems with different time slicings.

The zeroth component of the 4-velocity, $u^0 = dx^0/d\tau = dt/d\tau$ relates the proper time τ to the *coordinate time* t, and the other components of the 4-velocity, $u^i = dx^i/d\tau$, to the *coordinate velocity* $v^i \equiv dx^i/dt = u^i/u^0$. To convert this coordinate velocity into a "physical" velocity (with respect to the coordinate system), we need to use the metric, see (2.20).

In an *orthogonal* coordinate system the coordinate lines are everywhere orthogonal to each other. The metric is then diagonal, meaning that it contains no cross-terms like $dxdy$. We will only use orthogonal coordinate systems in this course.

The three-dimensional subspace, or *hypersurface* $t = const.$ of spacetime is called the space (or the *universe*) at time t, or a *time slice* of the spacetime. It is possible to slice the same spacetime in many different ways i.e. to make different choices of the time coordinate t.

2.5 Vectors, tensors, and the volume element

The *metric* $g_{\mu\nu}$ of spacetime is related to the distance interval by

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \equiv \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} g_{\mu\nu}dx^\mu dx^\nu. \tag{2.15}$$

We introduce the Einstein *summation rule*: we always sum over repeated indices, even if we don't bother to write down the summation sign $\sum$. This also applies to Latin indices, $g_{ij}dx^i dx^j \equiv \sum_{i=1}^{3}\sum_{j=1}^{3} g_{ij}dx^i dx^j$. The objects $g_{\mu\nu}$ are the components of the *metric tensor*. They are usually taken to be dimensionless, but sometimes (particularly in the case of angular coordinates) it is more useful to keep the coordinates dimensionless and put the dimension in the metric. The components of the metric tensor form a symmetric 4×4 matrix.

In the case of Minkowski space, the metric tensor in Cartesian coordinates is called $\eta_{\mu\nu} \equiv \mathrm{diag}(-1, 1, 1, 1)$. In matrix notation we have for Minkowski space

$$g_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.16}$$

in Cartesian coordinates, and

$$g_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{bmatrix} \tag{2.17}$$

in spherical coordinates.

As another example, the metric tensor for a sphere (discussed above as an example of a curved 2D space) has the components

$$[g_{ij}] = \begin{bmatrix} r_0^2 & 0 \\ 0 & r_0^2 \sin^2\theta \end{bmatrix}. \tag{2.18}$$

The vectors that occur naturally in relativity are *four-vectors*, with four components, as with the four-velocity discussed above. We will use the short term "vector" to refer both to three-vectors and four-vectors, as it should be obvious from the context which one we mean. As in three-dimensional flat geometry, the values of the components depend on the basis used. For example, if we move along the coordinate x^1 so that it changes by dx^1, the distance travelled is $ds = \sqrt{g_{11} dx^1 dx^1} = \sqrt{g_{11}} dx^1$. Similarly, the components of a vector do not give the physical magnitude of the quantity. In the case when the metric is diagonal, we just multiply by the relevant metric component to get the physical magnitude,

$$w^{\hat{\alpha}} \equiv \sqrt{|g_{\alpha\alpha}|} w^\alpha \ , \tag{2.19}$$

where w^α is the component of a vector in the basis where the metric is $g_{\alpha\beta}$, and $w^{\hat{\alpha}}$ is the correctly normalised physical magnitude of the vector. (In the above, there is no summation over α.)

For example, the physical velocity of an object is[3]

$$v^{\hat{i}} = \sqrt{g_{ii}} dx^i / \sqrt{|g_{00}|} dx^0 \,, \tag{2.20}$$

and the spatial components are always smaller than one.

The volume of a region of space (given by some range in the spatial coordinates x^1, x^2, x^3) is given by

$$V = \int_V dV = \int_V \sqrt{\det[g_{ij}]} dx^1 dx^2 dx^3 \tag{2.21}$$

where $dV \equiv \sqrt{\det[g_{ij}]} dx^1 dx^2 dx^3$ is the *volume element*. Here $\det[g_{ij}]$ is the determinant of the 3×3 submatrix of the metric tensor components corresponding to the spatial coordinates. For an orthogonal coordinate system, the volume element is

$$dV = \sqrt{g_{11}} dx^1 \sqrt{g_{22}} dx^2 \sqrt{g_{33}} dx^3. \tag{2.22}$$

Similarly, the surface area of a two-dimensional spatial region is S

$$S = \int_S dS = \int_S \sqrt{\det[g_{ij}]} dx^1 dx^2 \tag{2.23}$$

[3]When $g_{00} = -1$, this simplifies to $\sqrt{g_{ii}} dx^i / dt$.

where $dS \equiv \sqrt{\det[g_{ij}]}dx^1dx^2$ is the *area element*. Here $\det[g_{ij}]$ is the determinant of the 2×2 submatrix of the metric tensor components corresponding to the subvolume with constant x^0 and x^3. For an orthogonal coordinate system, we again have

$$dS = \sqrt{g_{11}}dx^1\sqrt{g_{22}}dx^2 \ . \tag{2.24}$$

The metric tensor is used for taking scalar products of four-vectors,

$$\mathbf{w}\cdot\mathbf{u} \equiv g_{\alpha\beta}u^\alpha w^\beta \, . \tag{2.25}$$

The (squared) *norm* of a four-vector $\mathbf{w}$ is

$$\mathbf{w}\cdot\mathbf{w} \equiv g_{\alpha\beta}w^\alpha w^\beta . \tag{2.26}$$

Exercise: Show that the norm of the four-velocity is always -1.

2.6 Contravariant and covariant components

(This subsection is not needed for the course, but it may help to clarify things for those who wonder why we sometimes write indices up and sometimes down, and what the difference is.)

Sometimes the index is written as a subscript, sometimes as a superscript. We will not be doing index gymnastics in the course, but for completeness' sake, let us say a few words about this. The component w^α of a four-vector is called a *contravariant* component. The corresponding *covariant* component is defined as

$$w_\alpha \equiv g_{\alpha\beta}w^\beta . \tag{2.27}$$

The norm is now simply

$$\mathbf{w}\cdot\mathbf{w} = w_\alpha w^\alpha . \tag{2.28}$$

In particular, for the 4-velocity we always have

$$u_\mu u^\mu = g_{\mu\nu}u^\mu u^\nu = \frac{ds^2}{d\tau^2} = -1. \tag{2.29}$$

In Minkowski space written in Cartesian coordinates, the only difference is in the sign of the 0-component, but in curved spacetime (or in curved coordinates), the covariant and contravariant vectors can be quite different.

We defined the metric tensor through its covariant components (Eq. 2.15). We now define the corresponding covariant components $g^{\alpha\beta}$ as the inverse matrix of the matrix $[g_{\alpha\beta}]$,

$$g_{\alpha\beta}g^{\beta\gamma} = \delta_\alpha{}^\gamma \ . \tag{2.30}$$

Now

$$g^{\alpha\beta}w_\beta = g^{\alpha\beta}g_{\beta\gamma}w^\gamma = \delta^\alpha{}_\gamma w^\gamma = w^\alpha . \tag{2.31}$$

The metric tensor can be used to lower and raise indices. For tensors we have

$$\begin{aligned} A_\alpha{}^\beta &= g_{\alpha\gamma}A^{\gamma\beta} \\ A_{\alpha\beta} &= g_{\alpha\gamma}g_{\beta\delta}A^{\gamma\delta} \\ A^{\alpha\beta} &= g^{\alpha\gamma}g^{\beta\delta}A_{\gamma\delta}. \end{aligned} \tag{2.32}$$

Note that in general $A_\alpha{}^\beta \neq A^\beta{}_\alpha$ unless the tensor is symmetric.

The symbols $\delta_{\alpha\beta}$ and $\eta_{\alpha\beta}$ are not tensors, and the location of their indices carries no meaning.

2.7 The Einstein equation

Given that the degrees of freedom of the spacetime are given by the metric and we want to have equations of motion which are second order, they can only involve the metric and its first and second derivatives,

$$g_{\mu\nu}, \qquad \partial g_{\mu\nu}/\partial x^\sigma, \qquad \partial^2 g_{\mu\nu}/(\partial x^\sigma \partial x^\tau) \ , \tag{2.33}$$

as well as the matter degrees of freedom. The requirement of invariance under general coordinate transformations restricts the equation of motion (in four dimensions) to have the form

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_{\rm N} T_{\mu\nu} \ , \tag{2.34}$$

where $G_{\mu\nu}$ is a unique tensor constructed from the metric and its first and second derivatives and Λ is a constant called the *cosmological constant* (the reason for the name will become apparent in the next chapter) and $T_{\mu\nu}$ is the *energy-momentum tensor*, also known as the *stress-energy tensor*. This equation specifies how the geometry of spacetime and its matter content interact, in other words it is the *law of gravity* according to general relativity. We will not discuss the Einstein tensor or this equation in much detail in this course. In the first part of the course we only need it in the case of the homogeneous and isotropic approximation, and in the second part we will look at small perturbations around this. However, we have explained a little bit about general relativity to give some idea of the mathematical structure which underlies the Friedmann-Robertson-Walker models.

The energy-momentum tensor describes all properties of matter which affect the spacetime, namely energy density, momentum density, pressure, and stress. For frictionless continuous matter, a *perfect fluid*, it has the form

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu + p g_{\mu\nu} \ , \tag{2.35}$$

where ρ is the energy density and p is the pressure measured by an observer moving with four-velocity u^μ (such an observer is in the *rest frame* of the fluid). In cosmology we can usually assume that the energy tensor has the perfect fluid form. If we consider a fictitious observer who is comoving not with the fluid, but the coordinates (i.e. her four-velocity is $w^\alpha = \delta^{\alpha 0}$), T_{00} gives the energy density she measures, T_{i0} gives the momentum density, which is equal to the energy flux, T_{0i} and T_{ij} gives the flux of momentum i-component in j-direction.

In Newton's theory the source of gravity is mass, in the case of continuous matter, the mass density ρ_m. According to Newton, the gravitational field $\vec{g}_N$ is given by the equation

$$\nabla^2 \Phi = -\nabla \cdot \vec{g}_N = 4\pi G_{\rm N} \rho_m \ , \tag{2.36}$$

where Φ is the gravitational potential. (We earlier discussed Newton's law in the form of the force law for point particles; this potential formulation for a continuous medium is equivalent, for finite systems.) Comparing (2.36) to (2.34), the mass density ρ_m has been replaced by $T_{\mu\nu}$, and $\nabla^2\Phi$ has been replaced by the Einstein tensor $G_{\mu\nu}$, which is a short way of writing a complicated expression built from $g_{\mu\nu}$ and its first and second derivatives of. Thus the gravitational potential is replaced by the 10-component tensor $g_{\mu\nu}$.

Figure 7: Defining the angular diameter distance.

In the case of a weak gravitational field, the metric is close to the Minkowski metric, and it can be written as

$$\mathrm{d}s^2 = -(1+2\Phi)\mathrm{d}t^2 + (1-2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \ , \tag{2.37}$$

where $|\Phi| \ll 1$. The Einstein equation then reduces to

$$\nabla^2\Phi = 4\pi G_{\mathrm{N}}(\rho + 3p) - \Lambda \ . \tag{2.38}$$

Comparing this to (2.36), we see that the mass density ρ_m has been replaced by $\rho+3p$, leaving aside the cosmological constant for a moment. For relativistic matter, where mass is not the dominant contribution to the energy density and p can be of the same order of magnitude as ρ, this is an important modification to the law of gravity. For nonrelativistic matter, where the particle velocities are $v \ll 1$, we have $p \ll \rho \simeq \rho_m$, and we get the Newtonian equation. The cosmological constant corresponds to a Newtonian potential that is quadratic in the coordinates, and thus a linear repulsive force (for $\Lambda > 0$). We will see the repulsive effect of the cosmological constant even more clearly when we discuss cosmology in the next chapter.

2.8 Distance, luminosity, and magnitude

In general relativity, it is possible to define spacelike distances just as in special relativity and Newtonian physics. One simply draws a spacelike line and integrates $\sqrt{|ds^2|}$ along the line. However, in a space that evolves in time, it is impossible to measure such distances, because they are defined only at one particular moment in time. The observer necessarily moves forward in time, and can never travel in a spacelike direction. (In other words, the space changes as the observer goes about measuring it.) Even if we lived in a static universe where such measurements would be possible in principle, they could not be done in practice for cosmology, since we cannot move for cosmologically significant distances. Therefore, in cosmology the observationally relevant distances are those defined with respect to light. They are not distances in space but distances along lightlike directions in spacetime. The two main distances used in cosmology are the *angular diameter distance* and the *luminosity distance*.

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

$$\mathrm{d}\theta = \frac{\mathrm{d}S}{d} \ . \tag{2.39}$$

In general relativity, we therefore *define* the angular diameter distance of an object with proper size R and angular size θ as

$$d_A \equiv \frac{\mathrm{d}S}{\mathrm{d}\theta} \ . \tag{2.40}$$

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 size of the object we are observing. In cosmology, this is can be done reliably only in a few cases, the most notable of which is the pattern of the anisotropy of the CMB, which we will discuss in the second part of the course.

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)

$$F = \frac{L}{4\pi d^2} \ . \tag{2.41}$$

In general relativity, the luminosity distance d_L is defined as

$$d_L \equiv \sqrt{\frac{L}{4\pi F}} \ . \tag{2.42}$$

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 behaves qualitatively as expected from Euclidean intuition, unlike d_A.)

In any spacetime, the two distances are related by $d_L = (1+z)^2 d_A$, so there in practice there is only one independent observational cosmological distance[4].

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 magnitudes corresponds to a factor of 100 in luminosity[5]. The *absolute magnitude* M and *apparent magnitude* m of an object are defined as

$$\begin{aligned} M &\equiv -2.5 \log_{10} \frac{L}{L_0} \\ m &\equiv -2.5 \log_{10} \frac{F}{F_0} \, , \end{aligned} \tag{2.43}$$

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 is chosen so in relation

[4]The *parallax distance*, related to the change of angular position of objects on the sky with the movement of the observer, does not reduce to the angular diameter distance, but at present it has not been measured on cosmological scales. That is expected to change with ESA's Gaia satellite, launched in December 2013 and currently taking data.

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

to the absolute scale that a star whose distance is $d = 10$ pc has $m = M$. From this, (2.42) and (2.43) follows that the difference between apparent and absolute magnitudes is related to the luminosity distance as

$$m - M = -5 + 5 \log_{10}(D_L/\text{pc}) \ . \tag{2.44}$$

References

[1] C.M. Will, *The confrontation between general relativity and experiment*, *Living Rev. Rel.* **9** (2006) 3, http://www.livingreviews.org/lrr-2006-3 [arXiv:gr-qc/0510072]

[2] C.W. Misner, K.S. Thorne, J.A. Wheeler, Gravitation (Freeman 1973)