.

COSMOLOGY

Lecturer: Professor **Komissarov S.S**
Room: 10.19 in Maths Satellite
email: S.S.Komissarov@leeds.ac.uk

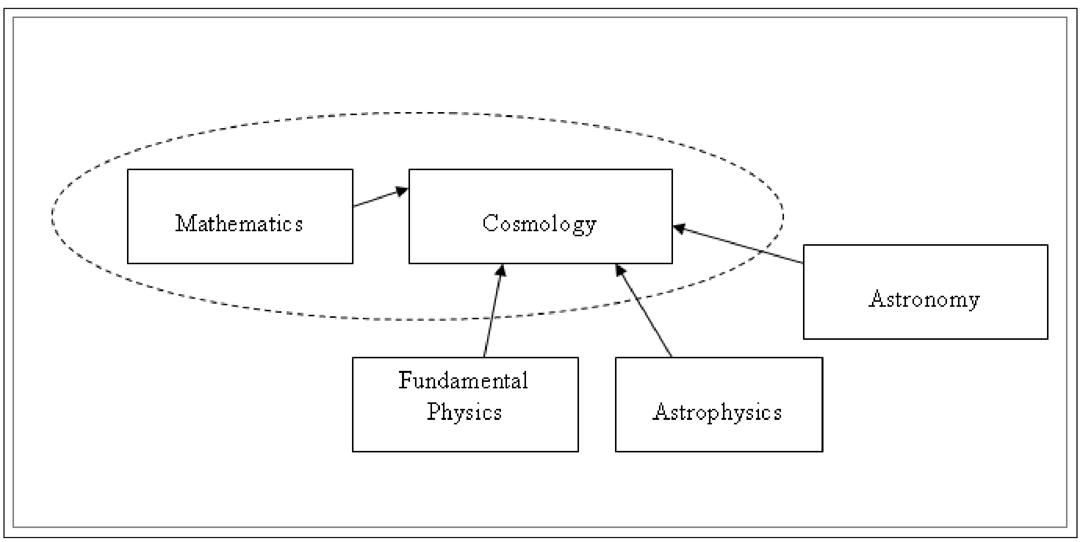
(2016)

Contents

Section 1

Introduction.

Cosmology is the science about the structure and evolution of the Universe on the large scale, its past, present and future. This is a very broad and rapidly developing science. It is based on modern fundamental Physics, Astronomy and employs a variety of mathematical methods. In our lectures, we will not be able to cover all aspects of Cosmology and will focus only on the key results which can be explained to students without strong background in Physics and Astronomy. We will use only traditional mathematics studied by all Natural Science and Engineering students, such as Calculus and Ordinary Differential Equations, so that the module can be accessible to a wide audience. From time to time, we will come across a difficult issue, which requires a more comprehensive background. In such cases, we will take the relevant results on trust.



The geometry of the Universe will be one of the main topics in our lectures. The most interesting issue here is that it is not Euclidean. The most important aspect of Cosmology is that it is relativistic – it is rooted in Einstein‘s *General relativity* with its notion of *spacetime*. As the Theory of Relativity is studied mainly by Physics students, we will outline its key ideas and describe how the basic equations of theoretical Cosmology arise from the relativistic theory of gravity. Finding solutions to these equations and their analysis is a big part of our lectures. We will see that the Universe is expanding and this expansion can take many different forms. The theoretical foundations behind the observational exploration of our Universe in attempts to understand its actual properties and contents will also be covered. We will learn how the astronomers and cosmologists deduce the presence of mysterious *dark matter* and *dark energy* and how the observations stimulate development of new Physics.

Section 2

Euclidean plane.

2.1 Cartesian coordinates $\{x, y\}$

Strict definition: *Euclidean plane is a 2-dimensional continuum of points with prescribed distances between them, which allows such coordinates $\{x, y\}$ that the distance Δl between the points (x, y) and $(x+\Delta x, y+\Delta y)$ satisfies the Pythagoras formula:*

$$\boxed{\Delta l^2 = \Delta x^2 + \Delta y^2} \tag{2.1}$$

for any x and y. These coordinates are called Cartesian after René Descartes (Cartesius). Usually, a Euclidean plane is considered as a (flat) surface of 3-dimensional (3D) Euclidean space but it does not have to, in which case it is better called a *2D Euclidean space*.

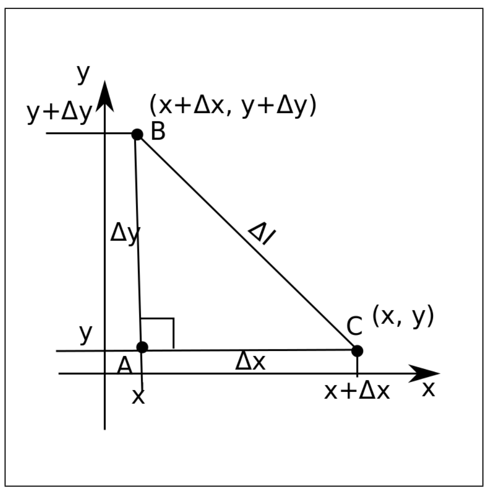


Figure 2.1: In the figure, ABC is a right angle triangle. Δy is the distance between A and B, Δx is the distance between A and C, Δl is the distance between B and C.

Obviously, the distance between (x, y) and $(x+\Delta x, y)$ is Δx and the distance between (x, y) and $(x, y+\Delta y)$ is Δy, assuming $\Delta x, \Delta y > 0$.

Lines along which $y = \text{const}$ are called x coordinate lines. Lines along which $x = \text{const}$ are called y coordinate lines. *Thus, the Cartesian coordinates refer to the distances measured along these coordinate lines.*

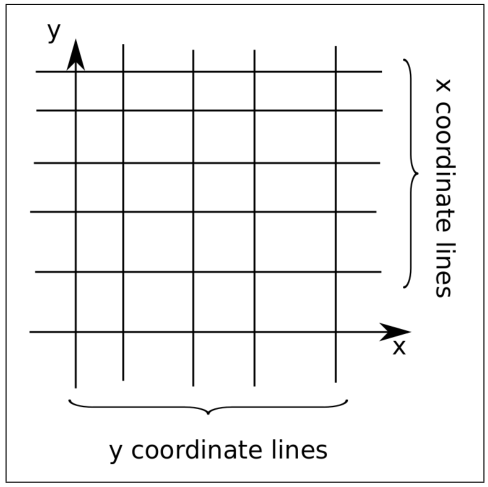


Figure 2.2: The coordinate grid of Cartesian coordinates.

For two infinitesimally close points (x,y) and $(x+dx,y+dy)$, we replace "Δ" with "d". (dx is vanishingly small compared to 1 and dx^2 is vanishingly small compared to dx. E.g. $dx+1=1$, $dx+dx^2=dx$.) The distance between them is given by

$$\boxed{dl^2 = dx^2 + dy^2} \tag{2.2}$$

This is the *metric form* of a Euclidean plane in Cartesian coordinates.

Suppose we need to find a distance Δ_{AB} between points A and B with coordinates $x_1 = x(\lambda_1)$, $y_1 = y(\lambda_1)$ and $x_2 = x(\lambda_2)$, $y_2 = y(\lambda_2)$ respectively along a curve set by the functions $x = x(\lambda)$ and $y = y(\lambda)$. Clearly this is not $\sqrt[2]{\Delta x^2 + \Delta y^2}$. Instead, we have to integrate

$$\Delta l_{AB} = \int_A^B dl\,.$$

This can be reduced to integration with respect to λ. Indeed

$$dl^2 = dx^2 + dy^2 = (\dot{x}d\lambda)^2 + (\dot{y}d\lambda)^2 = (\dot{x}^2 + \dot{y}^2)d\lambda^2.$$

Hence,

$$\Delta l_{AB} = \int_{\lambda_1}^{\lambda_2} (\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}\, d\lambda.$$

2.2 Polar coordinates $\{r, \phi\}$

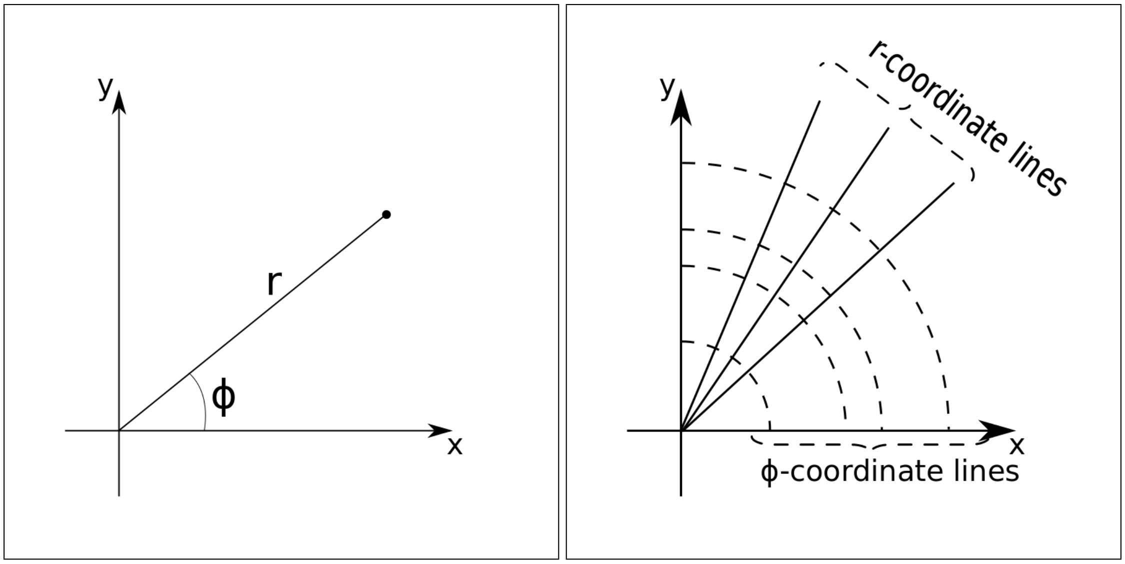


Figure 2.3: The r-coordinate lines of the polar coordinates are still straight but the ϕ-coordinate lines are *curved*.

In polar coordinates, r is the distance from the origin (the reference point of the coordinate system) to the point and ϕ is the angle between the reference direction and the direction to the point. Select the Cartesian coordinates with the same origin whose x axis is aligned with the reference direction of polar coordinates. Then

$$\begin{cases} x = r\cos\phi \\ y = r\sin\phi \end{cases} \tag{2.3}$$

Coordinate lines. The polar coordinate system is an example of *curvilinear coordinates* whose coordinate lines are not straight. Just like the Cartesian coordinates, they are orthogonal – the r- and ϕ- coordinate lines are perpendicular to each other (at the point of intersection).

What is the metric form of Euclidean plane in polar coordinates ?

Analytic derivation: This type of approach works for any coordinate transformation.

$$dx = \cos\phi\, dr - r\sin\phi\, d\phi$$
$$dy = \sin\phi\, dr + r\cos\phi\, d\phi$$

$$\begin{aligned} \Rightarrow dl^2 = dx^2 + dy^2 = \\ &= \cos^2\phi\, dr^2 + r^2\sin^2\phi\, d\phi^2 - 2r\cos\phi\sin\phi\, dr\, d\phi + \\ &+ \sin^2\phi\, dr^2 + r^2\cos^2\phi\, d\phi^2 + 2r\cos\phi\sin\phi\, dr\, d\phi = \\ &= dr^2 + r^2\, d\phi^2 \end{aligned}$$

Thus,

$$\boxed{dl^2 = dr^2 + r^2 d\phi^2,} \tag{2.4}$$

- *the metric form of a Euclidean plane in polar coordinates*.

Geometric derivation: Here we utilise the special property of the polar coordinates – their orthogonality.

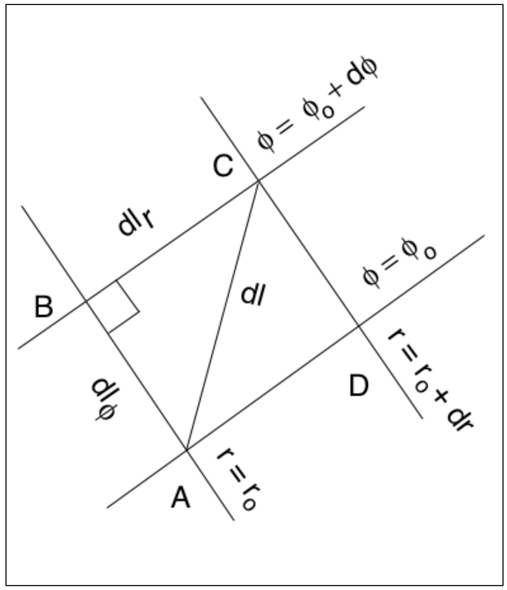


Figure 2.4: Deriving metric form using the geometric approach. The infinitesimally small parallelogram ABCD is rectangular because the polar coordinates are orthogonal.

Consider two infinitesimally close r coordinate lines $\phi = \phi_0$ and $\phi = \phi_0 + d\phi$. Since they are so close to each other they are parallel to each other. Next consider two infinitesimally close ϕ coordinate lines $r = r_0$ and $r = r_0 + dr$. They are also parallel to each other. Hence ABCD is a parallelogram. Moreover it is a rectangular as the coordinate system is orthogonal. Hence the length of its diagonal can be found from the Pythagoras theorem

$$dl^2 = dl_r^2 + dl_\phi^2 \,,$$

where $dl_r = dr$ and $dl_\phi = r d\phi$ are the lengths of its sides (see Fig.2.4). Hence

$$dl^2 = dr^2 + r^2 d\phi^2 \,.$$

2.3 General coordinates

Let us introduce new coordinates $\{x_1, x_2\}$ via

$$\begin{cases} x_1 = x, \\ x_2 = y - x. \end{cases} \tag{2.5}$$

From these,

$$\begin{aligned} dx &= dx_1 \,, \\ dy &= dx_2 + dx = dx_2 + dx_1 \,. \end{aligned}$$

Hence,

$$dl^2 = dx^2 + dy^2 = 2dx_1^2 + dx_2^2 + 2dx_1dx_2. \tag{2.6}$$

In contrast to Cartesian and polar coordinates, the metric form now includes the *mixed term* $2dx_1dx_2$. This is a direct result of the *non-orthogonality* of this coordinate system. Indeed, the x_1 *coordinate lines* are given by x_2 =const and hence satisfy

$$y = x + \text{const},$$

whereas x_2 *coordinate lines* are given by x_1 =const and hence

$$x = \text{const}.$$

Thus, they make and angle 45^o to each other.

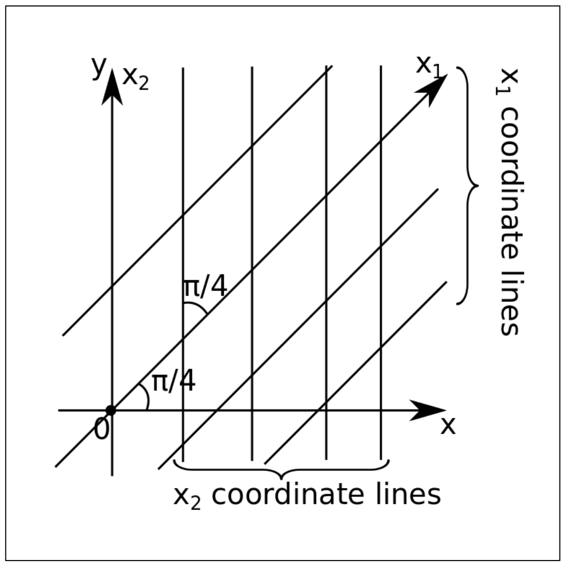


Figure 2.5: The coordinate lines are not orthogonal.

In the general case of arbitrary (curvilinear) coordinates $\{x_1, x_2\}$, the metric form of a Euclidean plane can be written as

$$\boxed{dl^2 = g_{11}dx_1^2 + g_{12}dx_1dx_2 + g_{21}dx_2dx_1 + g_{22}dx_2^2 = \sum_{i,j=1}^{2} g_{ij}dx_idx_j}, \tag{2.7}$$

where

$$g_{ij} = g_{ji} \tag{2.8}$$

are functions of x_1 and x_2. Since by definition $dl^2 > 0$, the metric form must be a *positive-definite quadratic form*.

Along the x_i coordinate line $dl^2 = g_{ii}dx_i^2$ and hence

$$dl_i = \sqrt{g_{ii}}dx_i, \tag{2.9}$$

where we assume that $dx_i > 0$. Thus, if $g_{ii} = 1$ then like in the Cartesian coordinates $dl_i = dx_i$. Such coordinates are called *normalised*. In the case of *orthogonal* coordinates $g_{12} = g_{21} = 0$.

Eq.(2.1) for the distance between the points A and B along a curve (or the length of this curve between A and B) generalises to

$$\Delta l_{AB} = \int_{\lambda_1}^{\lambda_2} \left(\sum_{i,j=1}^{2} g_{ij}\dot{x}_i\dot{x}_j \right)^{\frac{1}{2}} d\lambda. \tag{2.10}$$

In this equation, the functions $x_1(\lambda)$ and $x_2(\lambda)$ define the curve make $g_{ij}(x_1, x_2)$ a function of λ.

Metric form is what introduces distances between the points of a continuum (space, surface or manifold) in theoretical geometry. Given the metric form in some coordinates one can find how it looks in any other coordinates via corresponding coordinate transformations.

2.4 Area element in orthogonal coordinates.

The general expression for the area element (and volume in 3D Euclidean space) is somewhat complicated. In the case of orthogonal coordinates, it takes on a much simpler form.

Consider the patch of the plane define via $[x_1, x_1 + dx_1] \times [x_2, x_2 + dx_2]$. This is the parallelogram ABCD shown in the figure below.

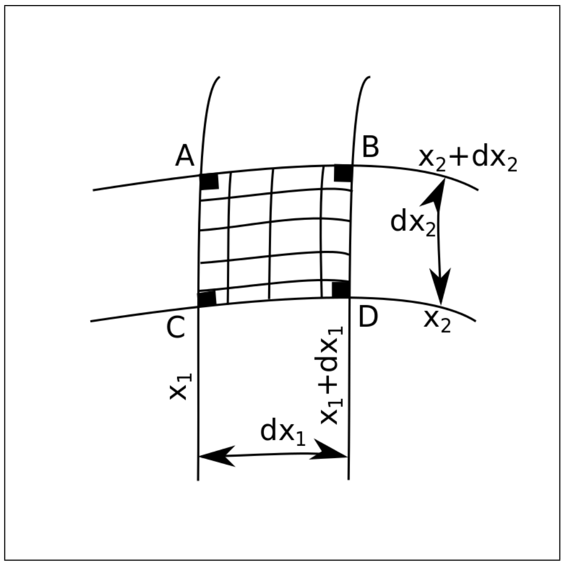


Figure 2.6: The infinitesimal patch $[x_1, x_1 + dx_1] \times [x_2, x_2 + dx_2]$ is a parallelogram. In the case of orthogonal coordinates, it is a rectangular.

The lengths of its sides are:

$$dl_1 = |AB| = |CD| = \sqrt{g_{11}} dx_1,$$
$$dl_2 = |AC| = |BD| = \sqrt{g_{22}} dx_2,$$

In *orthogonal coordinates*, this is a rectangular and its area is simply $dA = dl_1 dl_2$. Hence,

$$\boxed{dA = \sqrt{g_{11} g_{22}} dx_1 dx_2.} \tag{2.11}$$

For example, in polar coordinates $g_{rr} = 1$ and $g_{\phi\phi} = r^2$, hence $dA = r dr d\phi$.

Obviously, finding the area of a finite shape involves integration

$$A = \int dA = \int \int \sqrt{g_{11} g_{22}} dx_1 dx_2 .$$

Section 3

Three-dimensional Euclidean space.

3.1 Cartesian coordinates $\{x, y, z\}$

Definition: *Three-dimensional (3D) Euclidean space is a 3-dimensional continuum of points with prescribed distances between them, which allows such coordinates, $\{x, y, z\}$, that the distance between points (x, y, z) and $(x + \Delta x, y + \Delta y, z + \Delta z)$ satisfies the generalised Pythagoras formula*

$$\boxed{\Delta l^2 = \Delta x^2 + \Delta y^2 + \Delta z^2} \tag{3.1}$$

for any x, y, z. These coordinates are called Cartesian.

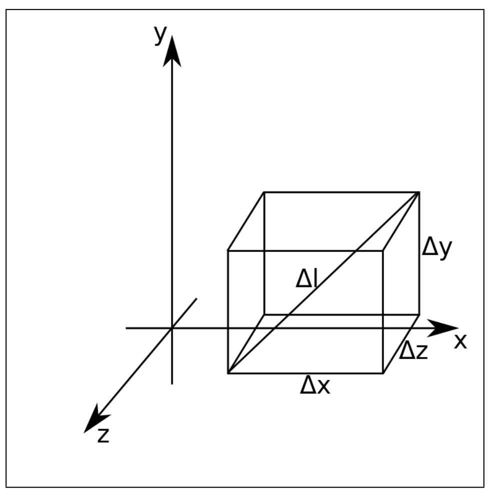


Figure 3.1: The rectangular parallelepiped with edges of lengths $\Delta x, \Delta y, \Delta z$, and the diagonal of length Δl.

For any two infinitely close points we have

$$\boxed{dl^2 = dx^2 + dy^2 + dz^2.} \tag{3.2}$$

This is the metric form of 3D Euclidean space in Cartesian coordinates.

3.2 Spherical coordinates $\{r, \theta, \phi\}$

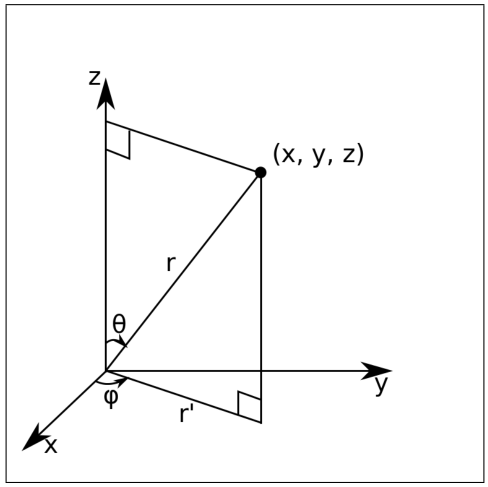


Figure 3.2: Spherical coordinates and aligned Cartesian coordinates.

Spherical coordinates employ one reference point (the origin) and two orthogonal reference directions, the polar direction and the azimuthal direction. The r coordinate is the distance from the origin (the reference point of the coordinate system) to the point. The θ coordinate (the polar angle) is the angle between the polar direction and the direction to the point; $\theta \in [0, \pi]$. The ϕ coordinate (the azimuthal angle) is the angle between the azimuthal reference direction and the direction to the orthogonal projection of the point onto the equatorial plane; $\phi \in [0, 2\pi]$. The definition implies that this coordinate system is orthogonal (see Figure 3.3).

Select the Cartesian coordinates with the same origin whose x axis is aligned with the azimuthal direction and z axis with the polar direction of spherical coordinates. Then

$$\begin{cases} z = r\cos\theta, \\ x = r'\cos\phi, \\ y = r'\sin\phi, \end{cases} \tag{3.3}$$

where $r' = r\sin\theta$ (see Figure 3.2). Hence,

$$\begin{cases} z = r\cos\theta, \\ x = r\sin\theta\cos\phi, \\ y = r\sin\theta\sin\phi. \end{cases}$$

It is easy to show that in spherical coordinates the metric form of Euclidean space is

$$\boxed{dl^2 = dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right).} \tag{3.4}$$

$\{x, y, z\}$ coordinates measure distances along their coordinate lines. The radial coordinate r does too (this is a normalised coordinate), but θ and ϕ do not (they are not normalised coordinates):

$$dl_r = dr, \quad dl_\theta = r\,d\theta, \quad dl_\phi = r\sin\theta\, d\phi \tag{3.5}$$

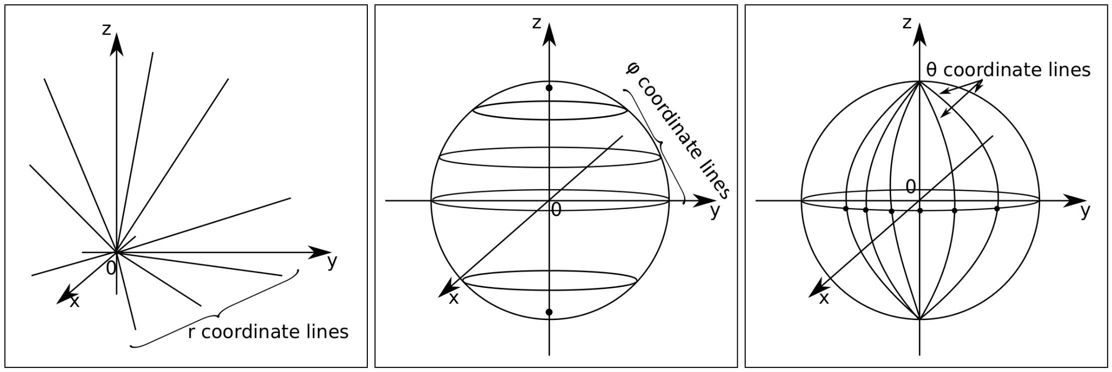


Figure 3.3: Coordinate lines of spherical coordinates.

3.3 Cylindrical coordinates $\{\varpi, \phi, \zeta\}$

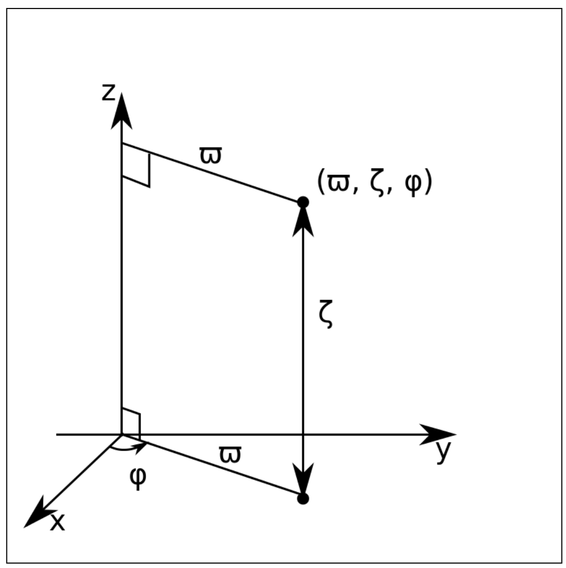


Figure 3.4: Cylindrical coordinates and aligned Cartesian coordinates

Cylindrical coordinates also employ one reference point (the origin), and two reference directions (the polar and azimuthal) but in a different way. The ϖ coordinate (the cylindrical radius) is the distance from the polar axis, the ζ coordinate (the elevation) is the distance from the equatorial plane and the ϕ coordinate (the azimuthal angle) is the angle between the azimuthal direction and the direction to the point's projection on the equatorial plane. The definition implies that this coordinate system is orthogonal.

Select the Cartesian coordinates with the same origin whose x axis is aligned with the azimuthal direction and z axis with the polar direction of cylindrical coordinates (see Figure 3.4). Hence

$$\begin{cases} x = \varpi \cos\phi, \\ y = \varpi \sin\phi, \\ z = \zeta\,. \end{cases}$$

It is easy to show that in cylindrical coordinates the metric form of Euclidean space is

$$\boxed{dl^2 = d\varpi^2 + d\zeta^2 + \varpi^2 d\phi^2\,.} \tag{3.6}$$

The distances along the coordinate lines

$$dl_\zeta = d\zeta, \quad dl_\varpi = d\varpi, \quad dl_\phi = \varpi\, d\phi. \tag{3.7}$$

3.4 General coordinates x_k

One can see that the metric form of a 3D Euclidean space is still a positive-definite quadratic form but the higher dimension leads to higher number of terms. In the general case of arbitrary (curvilinear) coordinates

$$\boxed{dl^2 = \sum_{i,j=1}^{3} g_{ij} dx_i dx_j}, \tag{3.8}$$

where

$$g_{ij} = g_{ji} \tag{3.9}$$

are functions of x_i. Compared to the 2D expression (Eq.2.7), the only difference is the upper limit in the summation, which is now 3.

3.5 Volume element of 3D Euclidean space in orthogonal coordinates

Consider the infinitesimal parallelepiped $[x_1, x_1+dx_1] \times [x_2, x_2+dx_2] \times [x_3, x_3+dx_3]$, where $\{x_i\}$ are some coordinates. If these are orthogonal coordinates then this parallelepiped is rectangular and its volume

$$dV = dl_1 dl_2 dl_3 \,,$$

where dl_i are the lengths of its sides. Replacing dl_i with $\sqrt{g_{ii}}dx_i$, we obtain

$$\boxed{dV = \sqrt{g_{11}g_{22}g_{33}}dx_1 dx_2 dx_3 \,.} \tag{3.10}$$

In the spherical coordinates, this reads

$$dV = r^2 \sin\theta \, dr \, d\theta \, d\phi \tag{3.11}$$

and in the cylindrical coordinates

$$dV = \varpi \, d\zeta \, d\varpi \, d\phi . \tag{3.12}$$

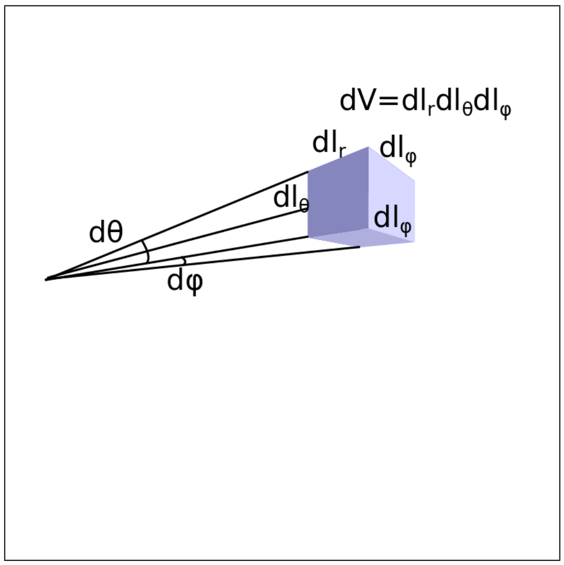


Figure 3.5: Volume element of spherical coordinates.

Section 4

Spherical geometry

Suppose that $\{x, y, z\}$ are Cartesian coordinates in 3D Euclidean space. Hence,

$$dl^2 = dx^2 + dy^2 + dz^2.$$

The surface defined by the equation

$$z = z_o \tag{4.1}$$

is a plane. The distance between any two infinitesimally close points of the plane is

$$dl^2 = dx^2 + dy^2. \tag{4.2}$$

The original coordinates x and y of the 3D space serve fine as coordinates on the plane plane and hence (4.2) is the metric form of the plane. From this it is manifest that the plane is a 2D Euclidean space and $\{x, y\}$ are its Cartesian coordinates.

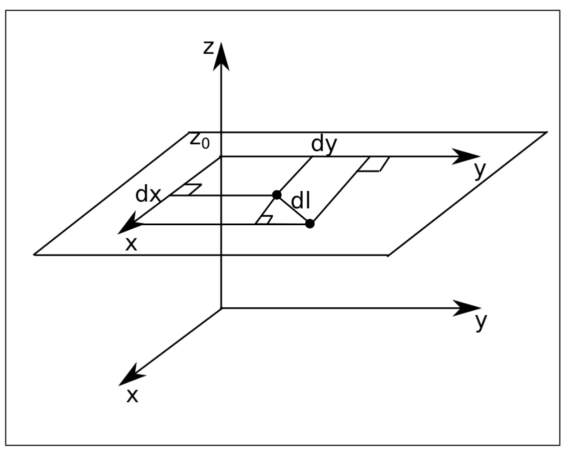


Figure 4.1: A plane in 3D Euclidean space

However, not all surfaces are planes and their geometry can be very different from Euclidean. The simplest example of such a surface is a sphere.

4.1 Generalised polar coordinates $\{\rho, \phi\}$

Any two points of the sphere are also points of the Euclidean space to which this sphere belongs. Hence, we can use the metric properties of the Euclidean space to determine the metric properties of the sphere. In spherical coordinates, the metric form of Euclidean space is

$$dl^2 = dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right).$$

Let us choose the origin of these coordinates to be at the centre of the sphere (point O) and denote as O' the point where the z axis intersects the sphere (the “North pole”; see figure 4.1). O' will be the origin of our generalised polar coordinates on the surface of the sphere.

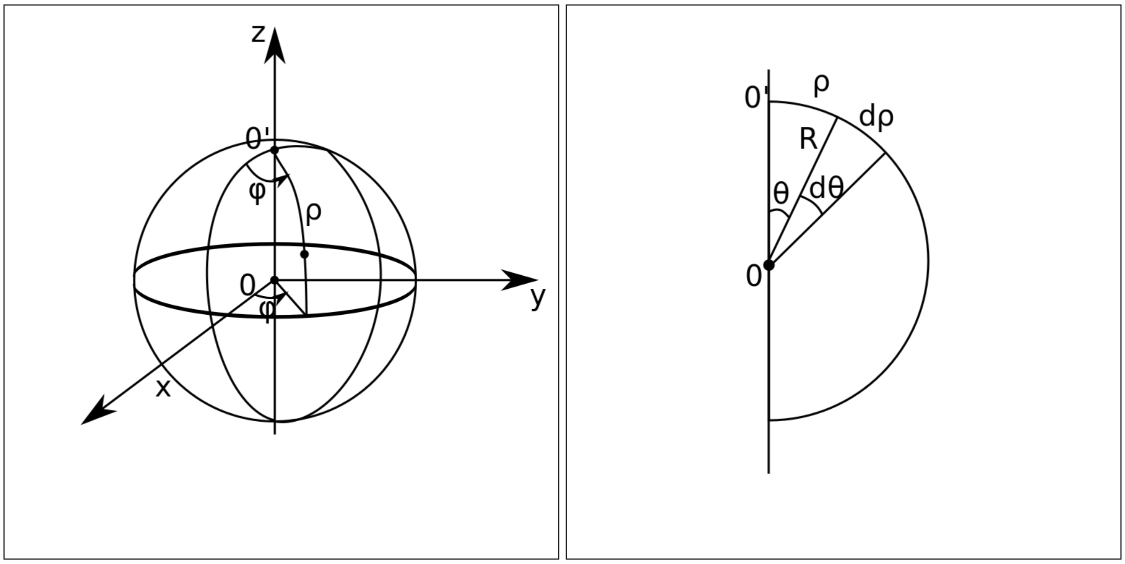


Figure 4.2: The generalised polar coordinate of Euclidean sphere.

Since on the sphere $r = R$ is constant, the distance between any two infinitesimally close points of the sphere is

$$dl^2 = R^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right). \tag{4.3}$$

Now let us replace θ with the new coordinate

$$\rho = R\theta\,. \tag{4.4}$$

Obviously, $\rho \in [0, \pi R]$. Since

$$d\rho = Rd\theta \tag{4.5}$$

we can write Eq.4.3 as

$$\boxed{dl^2 = d\rho^2 + R^2\sin^2\left(\frac{\rho}{R}\right)d\phi^2\,.} \tag{4.6}$$

This is the metric form of the sphere in coordinates $\{\rho, \phi\}$. Both these coordinates have a simple interpretation in terms of the measurements made on the surface of the sphere, which is illustrated in Fig.4.1. First, we consider the intersection of the XOZ plane with the sphere. This is circle of radius R, which will be called the zero meridian. Next consider the plane determined by the z axis and the point of the sphere whose coordinates we wish to determine. The intersection of this plane and the sphere is another circle of radius R - the point’s meridian. The ϕ coordinate is obviously the angle between these meridians as measured at the North pole. The ρ coordinate is the distance from the North pole along the point’s meridian. From (4.6), it follows that ρ and ϕ are orthogonal coordinates.

Choose a *small patch* on the surface with $\rho \ll R$. Then

$$\sin\left(\frac{\rho}{R}\right) \simeq \frac{\rho}{R}$$

and (4.6) becomes

$$\boxed{dl^2 = d\rho^2 + \rho^2 d\phi^2.} \tag{4.7}$$

This is the same as the metric form of a Euclidean plane in polar coordinates. Hence on small scales ($\ll R$) the geometry of a sphere is the same as Euclidean. This is why our ancestors believed that the Earth was *flat*. However on the large scale ($\sim R$), the difference between (4.7) and (4.6) is significant. This property is reflected in describing the spherical geometry as *locally Euclidean*. The same is true for any smooth surface of Euclidean space.

Suppose that functions $\rho(\lambda)$ and $\phi(\lambda)$ define a curve on the sphere. Then the length of this curve between the points A and B is

$$\Delta l_{AB} = \int_A^B dl = \int_{\lambda_A}^{\lambda_B} (\dot{\rho}^2 + R^2 \sin^2\left(\frac{\rho}{R}\right)\dot{\phi}^2)^{1/2} d\lambda\,.$$

4.2 Geodesics

Many basic geometric constructions of Euclidean plane involve straight lines. We need a generalisation of straight lines which could be applicable to curved surfaces. These are known as geodesics.

Definition: *A line of a surface is called geodesic of this surface if it satisfies the following condition. Consider any two points of this line. Then the distance between them as measured along this line is smaller compared to the distance measured along any other line of the surface.*

Geodesics satisfy a second order ODE called the *geodesic equation*. This is an example of the famous Euler-Lagrange equation. Given a metric form of a surface in some coordinates, one has to solve this equation in order to find geodesics of this surface. The geodesic equation:

$$\frac{d}{d\lambda}\frac{\partial L}{\partial \dot{x}^k} - \frac{\partial L}{\partial x^k} = 0 \quad (k = 1,2) \tag{4.8}$$

where

$$L = \sum_{i,j=1}^{2} g_{ij}\dot{x}_i\dot{x}_j. \tag{4.9}$$

Here x_i are coordinates on the surface, $g_{ij}(x_1,x_2)$ come from the metric form of the surface in these coordinates, functions $x_i(\lambda)$ describe a curve on the surface in parametric form, and $\dot{x}_k = dx_k/d\lambda$. For example, for a Euclidean plane in Cartesian coordinates $L = \dot{x}^2 + \dot{y}^2$. Hence

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = 0, \qquad \frac{\partial L}{\partial \dot{x}} = 2\dot{x}, \qquad \frac{\partial L}{\partial \dot{y}} = 2\dot{y}$$

and the geodesic equation reduces to

$$\ddot{x} = 0 \quad \text{and} \quad \ddot{y} = 0.$$

Integrating these equations, we find that $y = ax + b$, where a,b are constants. This is the familiar equation of straight line. In a similar way, one can show that all geodesics of a sphere are its *great circles*.

Definition: *A great circle of a sphere is the curve obtained via intersection of this sphere and a plane passing through its origin. A small circle of the sphere is its intersection with a plane that does not pass through the origin.*

Great circles are geodesics, whereas small ones are not.

As far as the coordinate lines of the generalised polar coordinates are concerned, (i) all the ρ coordinate lines (meridians) are great circles and hence geodesics, (ii) with one exception, the ϕ coordinate lines (parallels) are small circles and hence not geodesics. The only exception is the equator, $\rho = (\pi/2)R$, which is also a great circle (see the left panel of Fig.4.4).

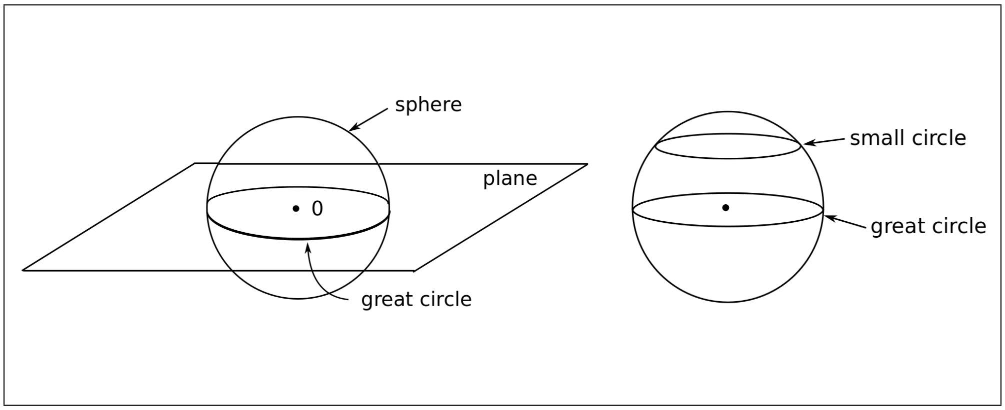


Figure 4.3: Great and small circles of spherical geometry.

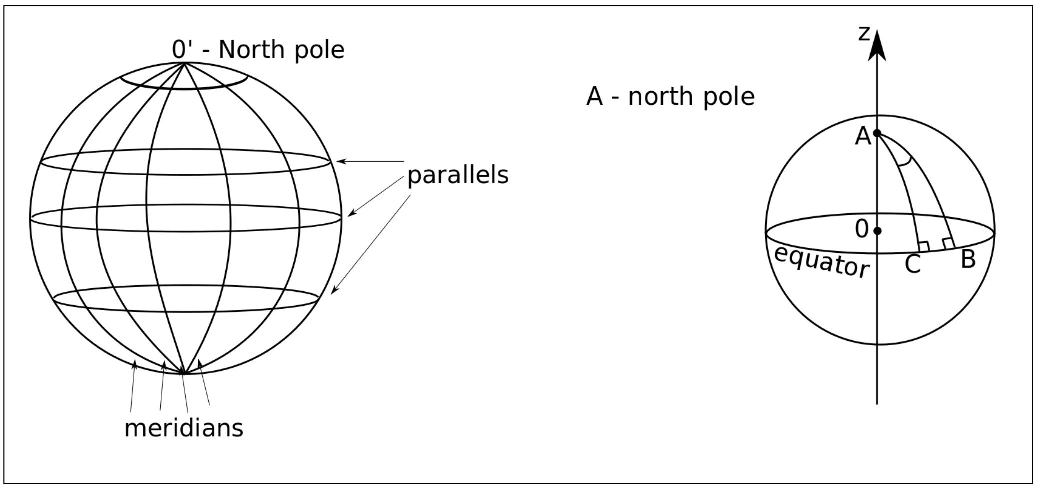


Figure 4.4: *Left:* The coordinate lines of the generalised polar coordinates. *Right:* A spherical triangle made of arcs of two meridians and the equator.

Using geodesics one can generalise many geometrical constructs of Euclidean planes. Let us say we wont to introduce triangles on the surface of a sphere. Clearly, such a triangle should have three vertices and three sides. Vertices are easy – they are just points. The sides are lines (curves) and one can connect two points with many different lines. By analogy with the Euclidean triangles these sides should be the shortest lines connecting the vertices. Hence geodesics.

In the right panel of Fig.4.4, AC, AB and BC are arcs of great circles of a sphere. Hence, ABC is a spherical triangle.

It is easy to see that

$$A\hat{C}B = 90^\circ \quad A\hat{B}C = 90^\circ \quad C\hat{A}B > 0^\circ$$

Thus, the sum of angles of the triangle ABC exceeds 180°.

This shows that the geometry of spheres (spherical geometry) is non-Euclidean. Another property of the spherical geometry which sets it apart from the geometry of Euclidean plane, with its parallel lines, is that

Any two geodesics of a sphere intersect each other. This includes the geodesics AC and AB of Fig.4.4, which can be considered as parallel since they both intersect the geodesic BC at 90^o.

4.3 Circumference of a spherical circle

Not only triangles, but many other constructs of Euclidean geometry have counterparts in spherical geometry. Consider for example a circle. In a plane, this is the set of all points at the same distance from the centre point. In the same way, one can define a circle on the surface of a sphere of radius R. But now the distance should be measured along the geodesics of the sphere.

Let the circle centre to coincide with the origin O' of generalised polar coordinates. The the circle of radius ρ is the set of all points with the radial coordinate ρ. (Incidentally, it is also a circle in the 3D Euclidean space to which the sphere belongs but in this space it has a different radius and a different centre.).

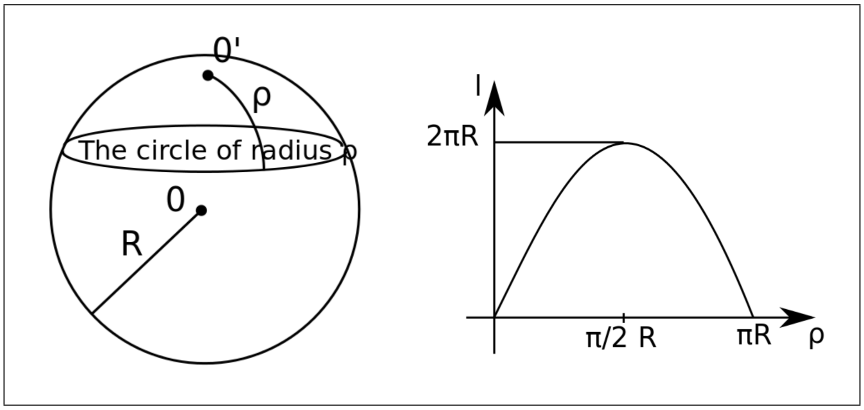


Figure 4.5: *Left:* Spherical circle of the radius ρ and centre O'. *Right:* Circumference of the circle as a function of its radius.

In the generalised polar coordinates centred on O', $d\rho = 0$ along the circle. Hence from (4.6) we have

$$dl = R\sin\left(\frac{\rho}{R}\right)d\phi\,.$$

Integrating this over the full turn, $(0, 2\pi)$, one finds the length of the circumference to be

$$\boxed{l = 2\pi R\sin\left(\frac{\rho}{R}\right)\,.} \tag{4.10}$$

From this one finds

$$l = \begin{cases} 2\pi R & \text{for } \rho = \frac{\pi}{2}R, \\ 0 & \text{for } \rho = \pi R. \end{cases}$$

This are clearly non-Euclidean results. However, for $\rho \ll R$ we have $\sin\left(\frac{\rho}{R}\right) \simeq \frac{\rho}{R}$ and

$$l \simeq 2\pi\rho\,. \tag{4.11}$$

Thus, the Euclidean result is recovered for very small circles. This example shows that hypothetical 2D creatures living in a 2D non-Euclidean world should be able to establish the true geometry of their world by exploring larger and larger regions.

4.4 The ”co-moving” radial coordinate χ

Replace the radial coordinate ρ with

$$\chi = \sin\left(\frac{\rho}{R}\right) = \sin\theta\,. \tag{4.12}$$

Then

$$
\begin{aligned}
d\chi &= \frac{1}{R}\cos\left(\frac{\rho}{R}\right)d\rho \\
&= \frac{1}{R}\left(1-\sin^2\frac{\rho}{R}\right)^{\frac{1}{2}}d\rho \\
&= \frac{1}{R}\left(1-\chi^2\right)^{\frac{1}{2}}d\rho
\end{aligned}
$$

or

$$
d\rho = R\frac{d\chi}{(1-\chi^2)^{\frac{1}{2}}} \tag{4.13}
$$

Then we can rewrite the metric form (4.6) as

$$
\boxed{dl^2 = R^2\left(\frac{d\chi^2}{(1-\chi^2)} + \chi^2 d\phi^2\right).} \tag{4.14}
$$

In this equation, R is a scaling factor. Its variation amounts to a uniform expansion/contraction of the sphere, where the distances between all its points, as defined by their coordinates χ and ϕ, vary by the same factor.

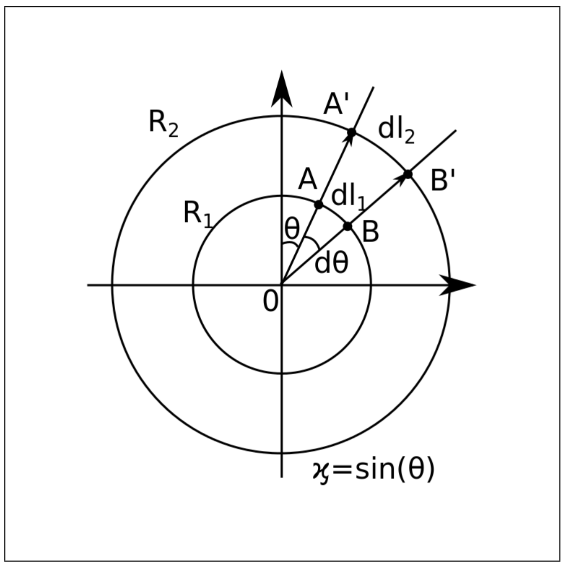


Figure 4.6: Geometric interpretation of the co-moving coordinate χ of a sphere.

Consider a spherical balloon which expands uniformly when inflated (with no rotation). Then the angular coordinates ϕ and θ, and hence ϕ and χ, for any point of the balloon remain unchanged. The coordinate grid of ϕ and χ coordinates expands with the balloon as if it was painted on it. Hence the name *co-moving*. For the distances between points with fixed χ and ϕ we have

$$
\frac{dl_2}{dl_1} = \frac{R_2}{R_1} \quad \text{or} \quad dl \propto R\,. \tag{4.15}
$$

- A similar co-moving coordinate can be introduced for a uniformly *stretched plane sheet*.

Just replace the polar radius ρ with χ via

$$\rho = R\chi. \tag{4.16}$$

Then the metric form of a plane in polar coordinates

$$dl^2 = d\rho^2 + \rho^2 d\phi^2$$

becomes

$$dl^2 = R^2 \left(d\chi^2 + \chi^2 d\phi^2\right). \tag{4.17}$$

The role of R as a *scaling factor* in this equation is manifest.

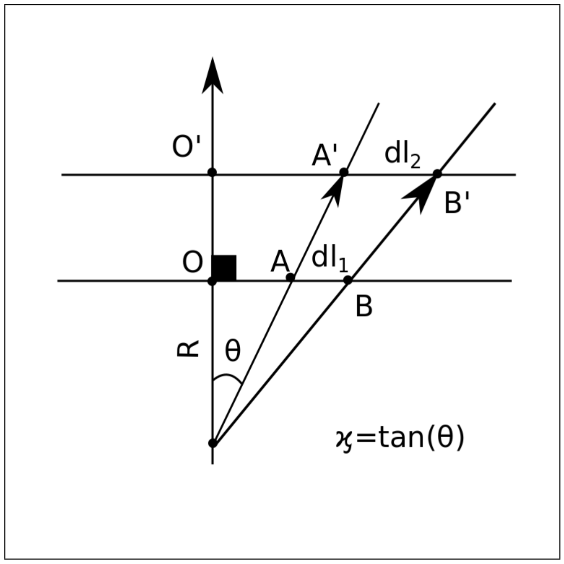


Figure 4.7: Geometric interpretation of the co-moving coordinate χ in the case of a uniformly stretched plane.

- The metrics (4.14) and (4.17) can be combined into

$$\boxed{dl^2 = R^2 \left(\frac{d\chi^2}{1-k\chi^2} + \chi^2 d\phi^2\right)}, \tag{4.18}$$

where $k = 0$ for a plane and $k = 1$ for a sphere.

4.5 General coordinates and Curvature

Since, the sphere is a two-dimensional surface, it should be no surprise that in general coordinates its metric form looks the same way as that of a plane

One can introduce many different kinds of coordinates for a sphere. They do not have to be neither orthogonal nor normalised. So in general, its metric form will be just a positive definite quadratic form

$$\boxed{dl^2 = \sum_{i,j=1}^{2} g_{ij} dx_i dx_j}, \tag{4.19}$$

with all $g_{ij} \neq 0$ depending on coordinates in some way. In fact, the same is true for a Euclidean plane which raises the question: “Given some metric form, is it possible to tell if it describes a Euclidean plane or some curved surface instead?" The answer is "Yes" but it is not easy to prove. The prove involves the notion of curvature (curvature tensor), which can be calculated from the metric. If and only if it vanishes everywhere then the surface is a Euclidean plane.

Section 5

4D Euclidean space.

Although we are unable to visualise 4D geometrical constructions (images) in our head we can still

1. use analytic approach to develop its theory, and
2. utilise various projections of 4D constructions into a 3D Euclidean space (hyper-planes of 4D Euclidean space).

5.1 Cartesian coordinates

Definition: *Four-dimensional (4D) Euclidean space* is a 4-dimensional continuum of points with prescribed distances between them, which allows such coordinates $\{x_1, x_2, x_3, x_4\}$ that for any two points (x_1, x_2, x_3, x_4) and
$(x_1 + \Delta x_1, x_2 + \Delta x_2, x_3 + \Delta x_3, x_4 + \Delta x_4)$, the distance between them is given by

$$\boxed{\Delta l^2 = (\Delta x_1)^2 + (\Delta x_2)^2 + (\Delta x_3)^2 + (\Delta x_4)^2\,.} \tag{5.1}$$

These coordinates are called *Cartesian*.

Consider the 2-dimensional continuum of points, $x_3 = \text{const}, x_4 = \text{const}$. Then from (5.1) we have that for this continuum

$$\Delta l^2 = (\Delta x_1)^2 + (\Delta x_2)^2\,. \tag{5.2}$$

This is the metric of a *Euclidean plane*.
Consider the 3-dimensional continuum, $x_4 = \text{const}$. For this continuum we have

$$\Delta l^2 = (\Delta x_1)^2 + (\Delta x_2)^2 + (\Delta x_3)^2\,. \tag{5.3}$$

This is not a plane, but its dimension is lower than the dimension of the space (4D). Its geometry is that of a 3D Euclidean space and it is called a *hyper-plane*.

5.2 Generalised spherical coordinates $\{\eta, \psi, \theta, \phi\}$

η - the radial coordinate (distance from the origin),
θ, ψ - *the two polar angles*, measured from two orthogonal reference directions, $\theta, \psi \in [0, \pi]$,
ϕ - *the azimuthal angle*, measured from a third reference direction in the *plane normal to the first two reference directions*, $\phi \in [0, 2\pi]$.

Introduce the Cartesian system of coordinates such that

1. it has the same origin
2. the axis x_4 is the reference direction for ψ,

3. the axis x_3 is the reference direction for θ,
4. the axis x_1 is the reference direction for ϕ.

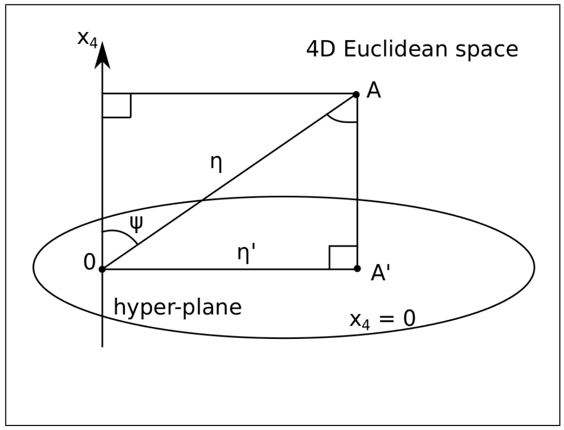


Figure 5.1: A' is the orthogonal projection of point A onto the coordinate hyper-plane $x_4 = 0$.

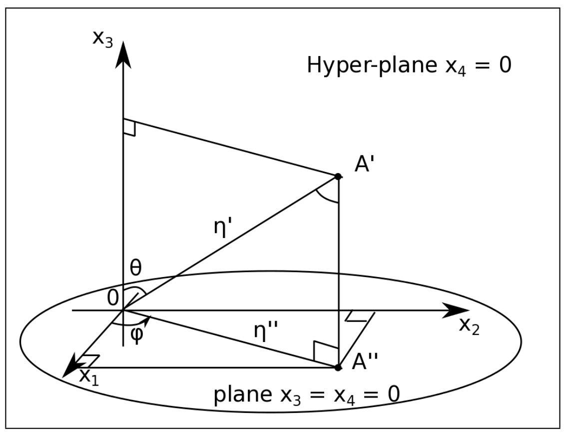


Figure 5.2: A'' is the orthogonal projection of point A' onto the coordinate plane $x_4 = 0, x_3 = 0$.

From these we conclude

$$\eta' = \eta \sin\psi, \quad \eta'' = \eta' \sin\theta = \eta \sin\psi \sin\theta, \tag{5.4}$$

$$\Rightarrow \begin{cases} x_4 = \eta \cos\psi, \\ x_3 = \eta' \cos\theta = \eta \sin\psi \cos\theta, \\ x_2 = \eta'' \sin\phi = \eta \sin\psi \sin\theta \sin\phi, \\ x_1 = \eta'' \cos\phi = \eta \sin\psi \sin\theta \cos\phi. \end{cases} \tag{5.5}$$

By construction, the distances along η-, ψ-, θ- and ϕ-coordinate lines are $dl_\eta = d\eta$, $dl_\psi = \eta d\psi$, $dl_\theta = \eta' d\theta = \eta \sin\psi d\theta$ and $dl_\phi = \eta'' d\phi = \eta \sin\psi \sin\theta \, d\phi$ respectively. (See figure 5.3 if not convinced.)

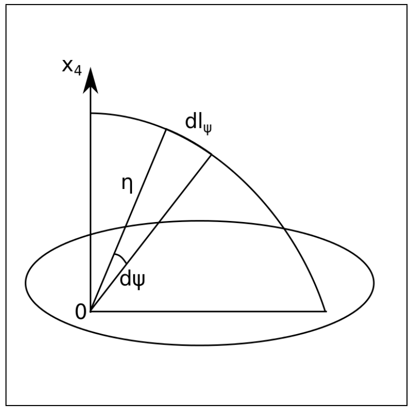


Figure 5.3: The ψ-coordinate line and its infinitesimal element.

By construction, the generalised spherical coordinates are *orthogonal* and hence $[\eta, \eta + d\eta] \times [\psi, \psi + d\psi] \times [\theta, \theta + d\theta] \times [\phi, \phi + d\phi]$ is a rectangular parallelepiped with edges of lengths dl_η, dl_ψ, dl_θ and dl_ϕ respectively. The length of its diagonal is given by the generalised Pythagoras equation

$$\begin{aligned} dl^2 &= dl_\eta^2 + dl_\psi^2 + dl_\theta^2 + dl_\phi^2 \\ &= d\eta^2 + \eta^2 d\psi^2 + \eta^2 \sin^2\psi \, d\theta^2 + \eta^2 \sin^2\psi \sin^2\theta \, d\phi^2. \end{aligned}$$

Thus,

$$\boxed{dl^2 = d\eta^2 + \eta^2 \left[d\psi^2 + \sin^2\psi \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right].} \tag{5.6}$$

Section 6

Riemann's hypersphere.

Definition:
A hypersphere is a 3D continuum of points in a 4D Euclidean space defined by the equation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = R^2$$

in Cartesian coordinates, or

$$\eta = R$$

in the generalised spherical coordinates. $R > 0$ is the radius of the hypersphere.

This is not a hyper-plane and its geometry is not Euclidean. Let us determine its metric form. The metric form of 4D Euclidean space is

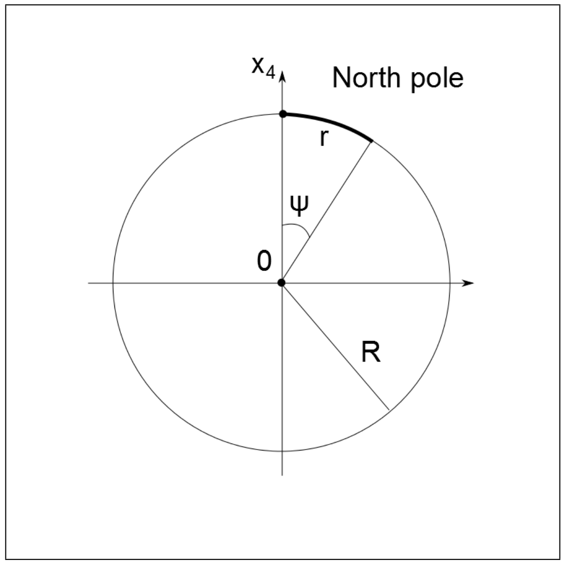
$$dl^2 = d\eta^2 + \eta^2 \left[d\psi^2 + \sin^2\psi \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right]$$

On the surface of the hypersphere $\eta = R$ and $d\eta = 0$. Hence,

$$\boxed{dl^2 = R^2 d\psi^2 + R^2 \sin^2\psi \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right).} \tag{6.1}$$

6.1 Generalised spherical coordinates $\{r, \theta, \phi\}$

Clearly, $R\psi = r$ is the distance from the "North pole" along the ψ coordinate line. Let us replace ψ with r.



Now (6.1) can be written as

$$\boxed{dl^2 = dr^2 + R^2 \sin^2\left(\frac{r}{R}\right)\left(d\theta^2 + \sin^2\theta\, d\phi^2\right).} \tag{6.2}$$

From its definition it follows that $r \in [0, \pi R]$. For $r \ll R$ we have $\sin\left(\frac{r}{R}\right) \simeq \frac{r}{R}$ and

$$\boxed{dl^2 = dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right).} \tag{6.3}$$

This is the metric form of 3D Euclidean space in spherical coordinates. Compare this with the metric form of a sphere of radius R in the generalised polar coordinates $\{\rho, \phi\}$,

$$dl^2 = d\rho^2 + R^2 \sin^2\left(\frac{\rho}{R}\right) d\phi^2,$$

which for $\rho \ll R$ becomes the metric form of 2D Euclidean space

$$dl^2 = d\rho^2 + \rho^2 d\phi^2.$$

Obviously, the system of coordinates $\{r, \theta, \phi\}$ of a hypersphere is a generalisation of the spherical coordinates of Euclidean space in the same sense as the generalised polar coordinates of a sphere is a generalisation of the polar coordinates of Euclidean plane. These coordinates have exactly the same meaning as in a 3D Euclidean space: r - radial distance, θ - polar angle, ϕ - azimuthal angle. The system is *orthogonal* (no mixed terms like $dr\,d\theta$).

The non-Euclidean geometry of a hypersphere was first discovered and studied by the German mathematician *Bernhard Riemann* in 1854. At some point *Albert Einstein* believed that our Universe had the geometry of a hypersphere. Imagine some intelligent beings living in a Universe with the geometry of Riemann's hypersphere. Having access only to a very small part of the Universe ($r \ll R$), they would conclude that their Universe has Euclidean geometry. Only when they manage to explore their Universe on large-scales ($r \sim R$) the deviations from the Euclidean geometry become apparent.

6.2 Geodesics of hypersphere

By solving the geodesic equations for hypersphere in the generalised spherical coordinates, it is easy to show that (i) all the radial coordinate lines (θ, ϕ = const) of this system are geodesics; (ii) out of all θ coordinate lines (r, ϕ = const), only those with $r = (\pi/2)R$ are geodesics; (iii) out of all ϕ coordinate lines (r, θ =

const), only the one with $\theta = \pi/2$, $r = (\pi/2)R$ is a geodesic. The r coordinate lines generalise the meridians of standard spherical coordinates. The ϕ and θ geodesics form a sphere of radius $r_s = (\pi/2)R$. This sphere generalises the equator of standard spherical coordinates. It splits the hypersphere into two halves of equal volume.

6.3 Circles of hypersphere.

Use the metric form (6.2)

$$dl^2 = dr^2 + R^2 \sin^2\left(\frac{r}{R}\right)\left(d\theta^2 + \sin^2\theta\, d\phi^2\right).$$

Consider the circle $\phi = \text{const}, r = \text{const}$. Along this circle

$$dl^2 = R^2 \sin^2\left(\frac{r}{R}\right) d\theta^2 \quad \Rightarrow \quad dl = R \sin\left(\frac{r}{R}\right) d\theta\,. \tag{6.4}$$

Hence its circumference

$$l = 2R\sin\left(\frac{r}{R}\right)\int_0^\pi d\theta = 2\pi R \sin\left(\frac{r}{R}\right) \tag{6.5}$$

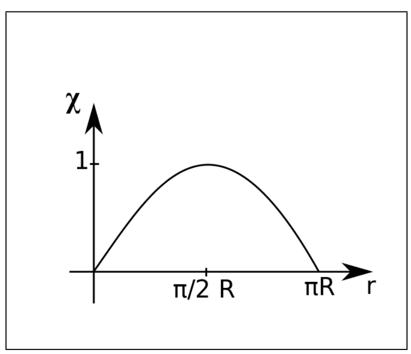
One can see that

1. for $r \ll R$, $l \simeq 2\pi r$, the Euclidean results,
2. for $r = \frac{\pi}{2}R$, $l = l_{\text{max}} = 2\pi R$,
3. for $r = \pi R$, $l = 0$ again.

Obviously, $r_{\text{max}} = \pi R$ is the distance to the most remote, from the centre of the circle, point of the hypersphere (the one which is exactly on the opposite side of the hypersphere). For $r = \pi R$ the circle contracts into this point.

6.4 The co-moving radial coordinate χ.

Instead of the radial coordinate r, introduce new coordinate $\chi = \sin(r/R) = \sin\psi$. Obviously $\chi \in [0,1]$, and the mapping $r \to \chi$ is *not one-to-one*.



$$d\chi = \frac{1}{R}\cos\left(\frac{r}{R}\right) dr$$

$$\Rightarrow dr = \frac{R d\chi}{\sqrt{1-\sin^2\left(\frac{r}{R}\right)}} = \frac{R}{\sqrt{1-\chi^2}} d\chi.$$

Thus the metric form of Riemann's hypersphere

$$dl^2 = dr^2 + R^2 \sin^2\left(\frac{r}{R}\right)\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)$$

becomes

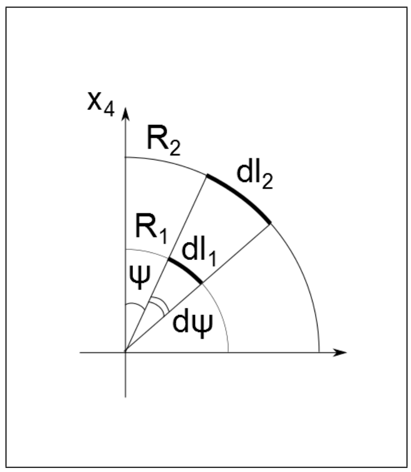
$$\boxed{dl^2 = R^2\left[\frac{d\chi^2}{1-\chi^2} + \chi^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right].} \tag{6.6}$$

Both in (6.1) and in (6.6) we have

$$\boxed{dl \propto R.} \tag{6.7}$$

This has the simple interpretation: when R increases, the distance between points with fixed coordinates ψ, θ, ϕ (or χ, θ, ϕ) increases with the same rate as R.

$$\frac{dl_2}{dl_1} = \frac{R_2}{R_1}$$



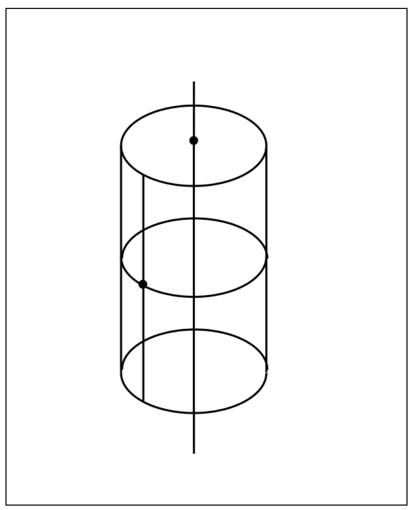
Section 7

Homogeneous and isotropic geometries.

Euclidean plane is *homogeneous (uniform)* – As far as its geometry is concerned, one cannot differentiate between its points. For example, no matter where one makes a right angle triangles it will satisfy the Pythagoras theorem.

Euclidean plane is *isotropic* – at any point, one cannot differentiate between directions. For example, no matter what the orientation of a right-angle triangle, it still satisfies the Pythagoras theorem.

A sphere is the only one another example of a surface with homogeneous and isotropic geometry. For example, a cylinder is homogeneous but not isotropic - the direction along its axis is obviously different from one perpendicular to it.



Similarly, a hyper-plane and a hypersphere are the only two examples of homogeneous and isotropic hyper-surfaces of a 4D Euclidean space. Using the *co-moving coordinate* χ, their metrics can be written as

$$\boxed{dl^2 = R^2 \left[\frac{d\chi^2}{1-k\chi^2} + \chi^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right]}, \tag{7.1}$$

where $k = 0$ corresponds to a *hyper-plane* and $k = 1$ corresponds to a *hypersphere*. (Compare this with equation (4.18).)

It turns out that $k = -1$ gives another type of geometry which is also homogeneous and isotropic. However, *it does not correspond to any hyper-surface of Euclidean space*[1]. It is known as the *Lobachevskian geometry (1830)*, or the geometry of hyperbolic space. Similar to the Euclidean space and the hypersphere,

[1] One can show that hyperbolic space is a hyper-surface of Minkowskian space.

one can introduce generalised spherical coordinates $\{r, \theta, \phi\}$ for the hyperbolic space. In these coordinates, its metric form is

$$dl^2 = dr^2 + R^2 \sinh^2\left(\frac{r}{R}\right)(d\theta^2 + \sin^2\theta d\phi^2)\,. \tag{7.2}$$

Notice that when $\chi \ll 1$ the metric form (7.1) reduces to that of Euclidean space,

$$dl^2 = R^2\left[d\chi^2 + \chi^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right]\,,$$

for all allowed values of k. Thus, the local geometry both of hypersphere and hyperbolic space is Euclidean. The fact that in our every-day life experience Euclidean geometry appears to provide an accurate description of our physical space may simply reflect the fact that we are dealing with too tiny a part of this space.

This issue is important due to the *Cosmological principle*, which postulates that the Universe is homogeneous and isotropic, and so its geometry. This means that its metric is limited (7.1) with $k = -1, 0, +1$. Which of the three describes the real world is to be established via experiments or observations. Notice that in all three cases we are dealing with worlds without bounds. It is also an infinite world for $k = 0$ or $k = -1$ - their volumes are infinite. For $k = +1$ (the hypersphere) the world is finite. It has a finite volume and its radial coordinate $r \in [0, \pi R]$.

Section 8

Space, time, and motion in Newtonian physics.

Here we summarise the key ideas defining Newtonian Physics.

- Absolute space:

  In addition to the apparent ability of physical objects to move *relative* to each other, one can also consider the motion in space itself, which exists independent of anything else. One is able to tell if an object is moving or at rest in the *absolute sense*, that is *relative to the space itself*. The geometry of absolute space is Euclidean.

- Absolute time:

  There is one and only one meaningful order of events in the Universe, the absolute order. One can always tell if an event A precedes, simultaneous with, or follows an event B in the absolute sense.

- Absolute motion of free bodies:

  A physical bodies, which are not subject to forces, moves in the absolute space along straight line with constant speed.

- Inertial frames:

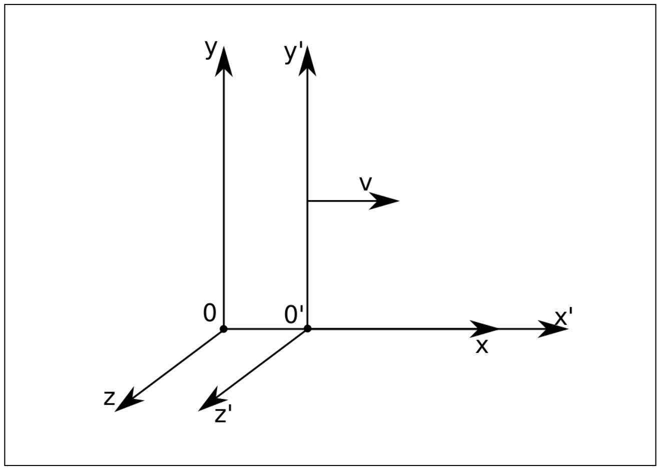
  An inertial frame is a non-rotating grid of spatial coordinates with fixed distances between its points which moves as a free body in the absolute space. This is a theoretical concept used to introduce mathematical laws of motion.

- Galilean relativity:

  All laws of mechanics are the same in all inertial frames. Hence mechanical phenomena can not be used to detect ”absolute motion”. This observation somewhat undermines the notion of absolute space.

- Galilean transformation.

  Consider two inertial frames moving relative to each other along the x-axis. At time $t = 0$ their origins coincide.



$$\begin{cases} z = z', \\ y = y', \\ x = x' + vt \end{cases} \tag{8.1}$$

  Here $\{x', y', z'\}$ are the Cartesian coordinates of a give point in space as measured in the "moving" frame, which we denote as O'. $\{x, y, z\}$ are the Cartesian coordinates of the same point as measured in the frame at "rest", which we denote as O. Equation (8.1) can be supplemented with

$$t = t'. \tag{8.2}$$

  According to the Galilean relativity, all mathematical equations describing the laws of mechanics must be invariant under the Galilean transformation (that is to have exactly the same form).

- Velocity addition law.

  From the Galilean transformation, it follows the Newtonian velocity addition law

$$w = w' + v, \tag{8.3}$$

  where $w = dx/dt$ and $w = dx'/dt'$ are the velocities of a body as measured in both the frames.

Section 9

Space, time and motion in Special Relativity.

Here is a brief summary of basic Special Relativity.

- Einstein's Special Relativity principle:

  ”All laws of physics are the same in all inertial frames. ” This implies that it is no special inertial frame, which can be associated with the absolute space. Hence the absolute space does not exist. The notion of inertial frames is refined. They are defined without any reference to the absolute space. Namely, *an inertial frame is a frame where free bodies move with constant velocities.* It is assumed that for any inertial frame one can introduce a Cartesian grid. Hence the geometry of the physical space associated with this frame is assumed to be Euclidean.

- The speed of light principle:

  ”The speed of light (waves or particles) is the same in all inertial frames. ” It is denoted as c. This can be understood not as an independent principle but as a result following from the first principle, which requires the Maxwell equations of electrodynamics to be the same in all inertial frames.

- Lorentz transformation:

  In order to satisfy the speed of light principle, the Galilean transformation is replaced with the Lorentz transformation:

  $$\boxed{\begin{cases} t = \gamma\left(t' + \frac{v}{c^2}x'\right), \\ x = \gamma(x' + vt'), \\ y = y', \\ z = z', \end{cases}} \tag{9.1}$$

  where

  $$\boxed{\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}}} \tag{9.2}$$

  is called the *Lorentz factor*. *Note that* $\gamma \geq 1$.

- Velocity composition law.

  Consider a particle moving with speed $w' = dx'/dt'$ along the x axis as measured in the frame O', and $w = dx/dt$ as measured in the frame O. From (9.1) we have

$$dt = \gamma\left(dt' + \frac{v}{c^2}dx'\right)$$
$$dx = \gamma\left(dx' + vdt'\right),$$

and

$$w = \frac{dx}{dt} = \frac{dx' + vdt'}{dt' + \frac{v}{c^2}dx'} = \frac{\frac{dx'}{dt'} + v}{1 + \frac{v}{c^2}\frac{dx'}{dt'}} = \frac{w' + v}{1 + \frac{vw'}{c^2}},$$

Hence

$$\boxed{w = \frac{w' + v}{1 + \frac{vw'}{c^2}}} \tag{9.3}$$

Suppose that $w' = c$. Then

$$w = \frac{c+v}{1 + \frac{vc}{c^2}} = \frac{c+v}{\frac{1}{c}(c+v)} = c.$$

Thus, *the Lorentz transformation is consistent with the speed of light principle*.

In order to satisfy the Special Relativity principle the mathematical equations describing laws of physics must be invariant under the Lorentz transformations.

- Time.

  In (9.1)

$$t = \gamma\left(t' + \frac{v}{c^2}x'\right). \tag{9.4}$$

Hence $t \neq t'$. This means that **each inertial frame has its own time**. How is this time measured? Using a *time grid*. This is a *collection of standard clocks* located at the grid points of the spatial grid of a given inertial frame. *These clocks are synchronised* in this frame. How do they get synchronised? Using *light pulses* as synchronisation signals. When an event takes place, its time is recorded by the clock of the frame which has the same instantaneous location as this event.

1. Relativity of simultaneity and temporal order.

   Consider two events. Suppose that in the frame O' they are separated by the time interval $\Delta t'$ and their x' coordinates differ by $\Delta x'$. Then in the frame O they are separated by the time interval

$$\Delta t = \gamma\left(\Delta t' + \frac{v}{c^2}\Delta x'\right). \tag{9.5}$$

   Suppose that $\Delta t' = 0$ (these events are simultaneous in the frame O'). Then

$$\Delta t = \frac{\gamma v}{c^2}\Delta x', \quad \text{which is} \quad \begin{cases} > 0 \text{ if } \Delta x' > 0, \\ = 0 \text{ if } \Delta x' = 0, \\ < 0 \text{ if } \Delta x' < 0, \end{cases}$$

   (here we assume $v > 0$). Hence, the order of events is different in different inertial frames.

2. Time dilation effect.

   Consider a *standard* clock at rest in the frame O'. Denote as τ the time of this clock (its *proper time*). Then for this clock, $\Delta x' = 0$ and $\Delta t' = \Delta\tau$ and (9.5) gives us

$$\boxed{\Delta t = \gamma\Delta\tau.} \tag{9.6}$$

Since $\gamma > 1$, the clock appears to run slower when observed in the frame O. Similarly, any standard clock at rest in the frame O will appear to run slower when observed in the frame O'. Thus, the rate of a moving clock slows down compared to the time of inertial frame where this clock is observed.

- Space.

1. Relativity of spatial order. Consider two events. Suppose that in the frame O' they are separated by the time interval $\Delta t'$ and their x' coordinates differ by $\Delta x'$. Then in the frame O their spatial separation along the x axis is

$$\Delta x = \gamma\left(\Delta x' + v\Delta t'\right) . \tag{9.7}$$

Suppose that $\Delta x' = 0$. Then

$$\Delta x = \gamma v \Delta t', \quad \text{which is} \quad \begin{cases} > 0 \text{ if } \Delta t' > 0, \\ = 0 \text{ if } \Delta t' = 0, \\ < 0 \text{ if } \Delta t' < 0, \end{cases}$$

(here we assume $v > 0$). Hence, the spatial order of events may differ in different inertial frames. Obviously, a similar results exists in Newtonian physics but the existence of absolute space make one of the frame privileged and hence implies unique spatial order of events.

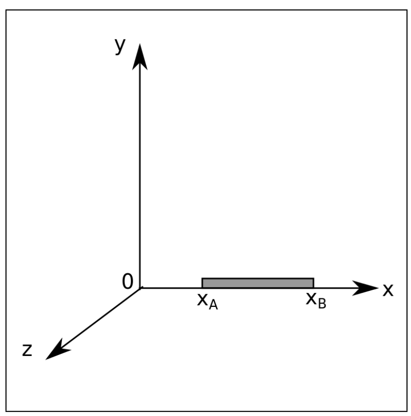
2. Length contraction effect.
From (9.1) we have

$$x = \gamma\left(x' + vt'\right) . \tag{9.8}$$

Consider a bar at *rest* in the frame O. Denote as l_0 its length as measured in this frame (its *proper length*). Suppose that it is aligned with the x axis. Then

$$l_0 = x_B - x_A \equiv \Delta x,$$

where x_A and x_B are the coordinates of its ends. It does not matter when x_A and x_B are measured.



In the frame O', this bar is *moving* and obviously it matters when the coordinates of its ends are measured. To make any sense, they have to be measured simulteneously in this frame. Thus, in the frame O' the length of this bar is

$$l = x'_B - x'_A = \Delta x',$$

where the coordinates x'_A, x'_B *are measured simultaneously* ($\Delta t' = 0$). Hence, from (9.8) we have

$$\Delta x = \gamma\left(\Delta x' + v\Delta t'\right) = \gamma \Delta x' .$$

Thus,

$$\boxed{l = \frac{l_0}{\gamma}.} \tag{9.9}$$

Thus, in the frame O', where the bar is moving, its length is smaller. This is called the *length contraction effect*.

Consider two identical bars, A and B, moving relative to each other. The above analysis implies that, in the frame where A is at rest, A is longer than B. On the contrary, in the frame where B is at rest, A is shorter than B.

Now consider two inertial frames O and O' moving relative to each other in the x derection. Each frame has it own Cartesian spatial grid (one can imagine it as built out of solid rods). The length contraction effect implies that according to measurements made in the frame O the grid of the frame O' is not Cartesian but squashed along the x axis and the other way around. This further crystallises the understanding that in Special Relativity *each inertial frame has its own space*.

Section 10

Spacetime.

10.1 The spacetime interval

The Theory of Relativity has this name because it insists that the absolute space and time do not exist and to describe any physical process one first has to specify which inertial frame is used. The description can differ significantly in another frame. There is no absolute spatial or temporal order to events - these are relative concepts. Yet, the Theory of Relativity replaces the absolute space and time with spacetime. Most importantly, it does not matter which frame is used to make the space and time measurements; the spacetime description will be the same. This makes spacetime an absolute concept of a new kind.

Consider an inertial frame O with its Cartesian coordinates $\{x, y, z\}$ and time t. Let $\Delta t, \Delta x, \Delta y, \Delta z$ be the differences between the coordinates of two events as measure in this frame. Combine them in this way

$$\boxed{\Delta s^2 = -c^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 .} \tag{10.1}$$

(Δs (and Δs^2) is called the *spacetime interval*.) Now let us take the measurements carried out in a different frame, frame O'. and combine $\Delta t', \Delta x', \Delta y', \Delta z'$ in the same way. It turns out that in both frames the numerical result is the same

$$-c^2 \left(\Delta t'\right)^2 + \left(\Delta x'\right)^2 + \left(\Delta y'\right)^2 + \left(\Delta z'\right)^2 = -c^2 \left(\Delta t\right)^2 + \left(\Delta x\right)^2 + \left(\Delta y\right)^2 + \left(\Delta z\right)^2 .$$

This is easily shown using the Lorentz transformation (9.1).

(10.1) looks as a metric form of a 4-dimensional space in orthogonal coordinates with Δs playing the role of distance. This space is called the *Minkowskian spacetime*. The points of this space are called *events*. The t coordinate is the only one here which is not normalised but this is easily corrected. Just introduce the new coordinate ct. To stress that this is just one of four coordinates of a 4D space, the uniform coordinate notation $x_0 = ct, x_1 = x, x_2 = y, x_3 = z$ is used. In these coordinates, the metric form of spacetime reads

$$\boxed{ds^2 = -\left(dx_0\right)^2 + \left(dx_1\right)^2 + \left(dx_2\right)^2 + \left(dx_3\right)^2 .} \tag{10.2}$$

This looks like the metric form of 4D Euclidean space, apart from the sign ”-” in front of dx_0^2. Because of this sign, the geometry of spacetime is hugely different, reflecting the indisputable difference between physical space and time. The Minkowskian spacetime is often described as a *pseudo-Euclidean* space and the $\{x_\alpha\}$ coordinates of Eq.10.2 are called *pseudo-Cartesian*.

10.2 Three types of spacetime intervals

Equation (10.2) can be written as

$$ds^2 = -c^2 dt^2 + dl^2, \tag{10.3}$$

where dt and dl are the time interval and the distance between two events as measured in some inertial frame, respectively. This form is useful for classifying the types of separation between events in spacetime.

There are three three different types of spacetime intervals in Minkowskian spacetime, space-like, time-like and null, determined by the sign of ds^2.

1. Space-like:
   If $ds^2 > 0$ then there exists a frame where $dt = 0$ and
   $$ds^2 = dl^2. \tag{10.4}$$

2. Time-like:
   If $ds^2 < 0$ then there exists a frame where $dl = 0$ and
   $$ds^2 = -c^2 dt^2. \tag{10.5}$$

3. Null:
   If $ds^2 = 0$ (but $dt \neq 0$) then
   $$\left|\frac{dl}{dt}\right| = c. \tag{10.6}$$
   Here we are dealing with two events which can be connected via a *light signal*; e.g. two events in the life of a photon (a particle of light).

10.3 What is space in the spacetime formulation of Special Relativity?

Figure 10.1 gives a the graphic representation of spacetime, where every point corresponds to some physical event. The t and x axes of frame O are shown as two mutually orthogonal lines (the y and z directions are not shown there). Using the Lorentz transformation it is easy to find the equations for the corresponding axes of the frame O'. The results are illustrated in the same figure.

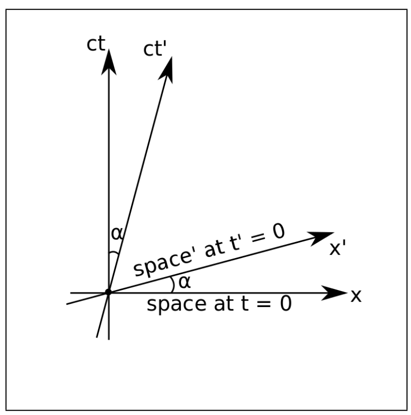


Figure 10.1: Splitting of spacetime into space and time of two different inertial frames.

The angle between the lines is
$$\alpha = \tan\left(\frac{v}{c}\right). \tag{10.7}$$

The space of frame O at time $t = 0$ is the hyper-plane $t = 0$ of the Minkowskian spacetime. In the figure, it is represented by the x axis. Any line parallel to the axis is the space of frame O as well, but at a different time.

The space of frame O' at time $t' = 0$ is the hyper-plane $t' = 0$ of the Minkowskian spacetime. In the figure, it is represented by the x' axis. Any line parallel to this axis is the space of frame O' as well, but at a different time.

10.4 World-lines.

Consider a particle moving according to the equations

$$x_i = f_i(t) \qquad\qquad (i = 1,2,3)\,.$$

In the spacetime, these equations define a line which is called *the world-line* of this particle. The angle between a line and the time axis can never exceed $\pi/4$, otherwise the particle speed would exceed the speed of light. In figure 10.1, the t axis is the world line of the origin of the O frame and the t' axis is that of the origin of the O' frame.

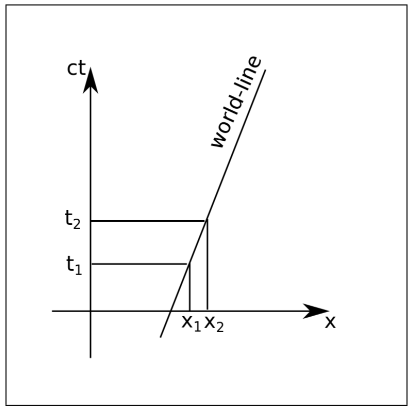


Figure 10.2: The world-line of a particle.

Consider a standard clock and its world-line. One can always find a frame where this clock is at rest (instantaneously). In this frame, for two infinitesimally close events in the “life” of this clock we have $dl = 0$ and $dt = d\tau$, where τ is the clock's proper time. Hence

$$\begin{aligned} ds^2 &= -c^2dt^2 + dl^2 \\ &= -c^2d\tau^2. \end{aligned} \tag{10.8}$$

Since $ds^2 < 0$ everywhere along the world-line of this clock, this line can be classified as *time-like*. The same is true for world-lines of all massive particles (particles of non-vanishing mass) which move with speeds below that of light. Massless particles, like photons, move at the speed of light. Along their world lines $dl = cdt$ and hence $ds^2 = 0$. Thus, their world-lines are *null*.

Consider some massive particle and imagine a standard clock moving with the particle. Its time is called the *proper time of the particle* and it can be considered as a parameter τ of the world-line, $t = t(\tau)$, $x_i = x_i(\tau)$. To any $\tau = \tau_0$ there corresponds a point of the world-line – the particle's location in spacetime at its proper time τ_0. As τ increases, the particle moves along the world-line from the direction of its past into the direction of its future.

Section 11

Mass-energy relation.

The laws of Newtonian mechanics are incompatible with the principles of Special Relativity. Hence, all laws of mechanics had to be rewritten in such a way that (1) they remain invariant under the Lorentz transformations; (2) they reduce to the laws of Newtonian mechanics in the so-called Newtonian limit, essentially the limit of low (compared to the speed of light) speeds. The revision has lead to few spectacular results. One of them is the famous

$$\boxed{E = mc^2\,.} \tag{11.1}$$

Here E is the *total energy* and m is the *inertial mass* of a physical body, both measured in the same inertial frame. Since E includes the kinetic energy, it depends on the choice of the frame and so does m. The energy measured in the frame where the body is at rest is minimum and so is its mass. This minimum mass is called the *rest mass*. it is denoted as m_0. The corresponding energy

$$E_0 = m_0 c^2\,, \tag{11.2}$$

is called the *rest mass-energy*. In a frame where the body is in motion, $E > E_0$ and hence $m > m_0$. In fact, the Theory of Relativity yields

$$m^2 = m_0^2 + \frac{p^2}{c^2}, \tag{11.3}$$

where $p = mv$ is the body's *momentum*. Hence

$$m^2 = m_0^2 + m^2\frac{v^2}{c^2}$$

or,

$$\boxed{m = m_0\gamma,} \tag{11.4}$$

where $\gamma = 1/\sqrt{1-v^2/c^2}$ is the Lorentz factor.

Any mole of gas, consists of many particles in motion. As it gets heated, the kinetic energy of this motion increases, leading to an increase of the mole's total energy and hence its inertial mass. However, this effect is strong only when the speed of thermal motion approaches the speed of light. Denote as n the number density of gas particles, ρ the mass density, e_t the thermal energy density and P the gas pressure. The total mass-energy density (ρc^2 or just ρ) of this gas is

$$\rho = mn\,.$$

One can distinguish two limiting regimes, where the relation between the total mass-energy and the thermal energy (and pressure) of gas is particularly simple.

1. If the speed of thermal motion of the gas particles $v_t \ll c$ then $\gamma_t \simeq 1$ and $m \simeq m_0$. Hence the total mass-energy density of this gas is dominated by the rest mass-energy of its particles and

$$\rho \simeq m_0 n,$$

   where m_0 is the mean rest mass of the particles. In this case,

$$\boxed{e_t, P \ll \rho c^2,} \tag{11.5}$$

   where e_t is the gas thermal energy density and P is its pressure.

2. If $v_t \simeq c$ and hence $\gamma_t \gg 1$ then $m \gg m_0$. In this case, the mass-energy density is dominated by the thermal energy density e_t:

$$\rho \simeq \frac{e_t}{c^2}.$$

   In fact, one can show that

$$\boxed{\rho = \frac{3P}{c^2},} \tag{11.6}$$

   where P *is the gas pressure*.

In Cosmology, any gas which satisfies (11.5) is traditionally called *matter* (or cold matter), and any gas which satisfies (11.6) is called *radiation*. The latter includes *photon gas*.

Section 12

Spacetime of General Relativity.

12.1 Brief summary

- Both in *Newtonian mechanics* and in *Special Relativity*, forces result in accelerated motion. Newtonian gravity is an example of such force.

- General Relativity treats gravity in a completely different way to any other type of force. In this theory, *gravity is identified with deformation of spacetime itself*, so its geometry is no longer Minkowskian. This deformed (warped or curved) spacetime *does not allow pseudo-Cartesian coordinates* and hence *the notion of a global inertial frame becomes obsolete*. The accelerated motion under the action of gravity is replaced by free motion in warped spacetime. The world-lines of free particles are geodesics of this spacetime.

- Locally, for a small "patch" of spacetime, its geometry is still Minkowskian, but globally it is not. This is similar to the local geometry of a curved smooth hyper-surface of 4D Euclidean space being Euclidean (See §6). Hence the notion of a *local inertial frame*: Within a small region of space and time on can set a nearly Cartesian frame with standard clocks and, according to the measurements of physical processes within this small region, they obey the laws of Special Relativity. On larger scales, the effect of spacetime curvature become significant.

- *The key equation of General relativity* is

  $$\boxed{\mathscr{R}_{\nu\mu} - \frac{1}{2}\mathscr{R}g_{\nu\mu} = \frac{8\pi G}{c^4}T_{\nu\mu}.} \tag{12.1}$$

  Here: G is the gravitational constant,
  $\mathscr{R}_{\nu\mu}$ is the Ricci tensor,
  $\mathscr{R}$ is the curvature scalar,
  $T_{\nu\mu}$ is the stress-energy-momentum tensor,
  $g_{\nu\mu}$ is the metric tensor (coefficients of the metric form).

  $\mathscr{R}_{\nu\mu}$ and $\mathscr{R}$ describe the curvature of spacetime. Their components can be expressed in terms of the components of the metric form and their first and second order derivatives.
  $T_{\nu\mu}$ describes the state of motion of matter (and radiation).

  This remarkable equation does not only describe *how matter deforms spacetime* but also *how matter moves in such a deformed spacetime*.

  In particular, it yields the second order PDEs which determine $g_{\nu\mu}$ as functions of spacetime coordinates.

  In flat spacetime $\mathscr{R}_{\nu\mu} = 0$ and $\mathscr{R} = 0$. In empty spacetime (free of matter and radiation), $T_{\nu\mu} = 0$. Thus, it is matter (and radiation) that warps spacetime.

Although Eq.12.1 looks rather neat, it is in fact very complicated and we are not in a position to analyse it in details in our lectures. The key equations of Cosmology, the Friedmann equations, follow directly from Eq.12.1. Unfortunately, the derivations are too involved and will not be shown here.

12.2 Splitting of spacetime into space and time

The most general metric form of spacetime can be written as

$$\boxed{ds^2 = -\alpha c^2 dt^2 + \sum_{i=1}^{3} \beta_i dt\, dx^i + \sum_{i,j=1}^{3} \gamma_{ij} dx^i\, dx^j.} \tag{12.2}$$

where spacetime is split into time and space (This can be done in many different ways.). Here t is a *time-like coordinate* of spacetime. It plays the role of global time.

In this 3+1 splitting of spacetime, the hyper-surface $t = \text{const}$ is the space at time t. Its metric form is

$$\boxed{dl^2 = \sum_{i,j=1}^{3} \gamma_{ij} dx^i\, dx^j,} \tag{12.3}$$

where $\{x^i\}$ are the coordinates of this space. The quadratic form of (12.3) is positive-definite. The geometry of this space generally changes with t, so γ_{ij} are not only functions of $\{x^i\}$ but also of t.

The *shift-vector* β_i describes the motion of the spatial grid through this space - it is the velocity vector of the grid.

$\alpha > 0$ is called the *lapse function*. It describes the rate of the global time compared to the time as measured by standard clocks co-moving with the spatial grid. Indeed, along the world line of any standard clock

$$ds^2 = -c^2 d\tau^2$$

where τ is the proper time of this clock (as indicated by its dial; see equation (10.8) of Section 10). Along the world line of any body comoving with spatial grid, $dx^i = 0$ and (12.2) gives us

$$ds^2 = -\alpha c^2 dt^2.$$

Hence

$$\alpha c^2 dt^2 = c^2 d\tau^2$$

or

$$\boxed{\frac{d\tau}{dt} = \alpha^{1/2}.} \tag{12.4}$$

Both α and β_i are functions of both t and x_i.

Section 13

Robertson-Walker metric and Friedmann equations

13.1 The Robertson-Walker metric of homogeneous and isotropic Universe.

Astronomical observations tell us that although there are lots of different structures in the Universe, stars, galaxies, clusters of galaxies etc., on sufficiently large scales it looks remarkably the same in all directions. Hence, either 1) the Universe has a well-defined centre and by accident we are located very close to it or 2) the Universe is *homogeneous* and hence appears isotropic to all observers in all locations. The latter assumption is adopted in modern Cosmology and it is often referred to as the *Cosmological Principle*.

From the geometric prospective, the Cosmological Principle implies that the spacetime can be split into time and space in such a way that

1. *The geometry of space is homogeneous and isotropic*. Robertson and Walker were first to prove that the metric form of any such space can be written as

$$dl^2 = R^2 \left[\frac{d\chi^2}{1-k\chi^2} + \chi^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right], \tag{13.1}$$

   where k is a constant parameter. $k > 0$, $k = 0$ and $k < 0$ correspond to three different geometries: Riemann's hypersphere, Euclidean space and Lobachevsky's hyperbolic space. It is easy to see that via rescaling of the coordinate χ, any metric with positive k can be reduced to the one with $k = 1$ and any metric with negative k can be reduced to the one with $k = -1$.

2. *The spatial grid is not moving through this space*,

$$\beta_i = 0. \tag{13.2}$$

3. *Everywhere, the global time t runs at the same rate relative to a standard clock at rest in this space.* Hence α does not depend on the spatial coordinates. Formally, it may depend on t but this dependence can be eliminated using suitable time variable $\tilde{t}$ such that

$$\sqrt{\alpha(t)}dt = d\tilde{t}.$$

The renormalised time $\tilde{t}$ runs with same rate as a standard clock at rest in the space. Combining (12.2) with (13.1)-(12.4) we obtain

$$\boxed{ds^2 = -c^2dt^2 + R^2 \left[\frac{d\chi^2}{1-k\chi^2} + \chi^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right].} \tag{13.3}$$

This is known as the *Robertson-Walker metric* (though introduced before them first by Friedmann and later by Lemaitre). Here R can be a function of time, $R(t)$, which gives us a dynamic Universe with "evolving" geometry. In fact, $R(t)$ is the only unknown function in this metric. Substitution of this metric into the Einstein's equation (12.1) yields ODE governing the evolution of R.

13.2 The Cosmological constant.

Since gravity is by far the strongest "force" on cosmological scale, any theory of gravity dictates particular properties to the Universe as a whole. General Relativity does it differently to Newtonian gravity and its emergence prompted the development of new cosmological models. In his initial study of this issue, Einstein was looking for solutions describing static (R =const) Universe filled with cold matter ($P = 0$) of constant density ρ. This was motivated by contemporary astronomical data. In fact, he anticipated the space to be warped by this matter and to have the geometry of Riemann's hypersphere. To his great surprise he found that the static solution was possible only for $\rho = 0$, which also implied flat space with $k = 0$. This made no sense to Einstein and forced him to modify the key equation (12.1) by including an additional term, namely $(\Lambda/c^2)g_{\nu\mu}$, so that

$$\boxed{\mathscr{R}_{\nu\mu} - \frac{1}{2}\mathscr{R}g_{\nu\mu} = \frac{8\pi G}{c^4}T_{\nu\mu} - \frac{\Lambda}{c^2}g_{\nu\mu},} \tag{13.4}$$

where Λ is a universal constant called the *Cosmological constant*. This modification has two important implications: 1) In the Newtonian limit, the law of gravity has a new component to the gravity force. While the standard Newtonian component is an attraction force the new one is a repulsion force. This did not look disastrous as the anti-gravity was be very small on non-cosmological scales and did not lead to conflict with the observations of planetary motion. 2) The spacetime can now be curved by itself, without any matter. Indeed, in flat spacetime $\mathscr{R} = 0$, $\mathscr{R}_{\mu\nu} = 0$, which is incompatible with non-vanishing Λ (in empty Universe $T_{\mu\nu} = 0$.).

13.3 Friedmann's equations

In contrast to Einstein, Russian mathematician *Alexander Friedmann* was interested in *time-dependent* cosmological solutions of Einstein's equations. Hence he assumed that $R = R(t)$, $\rho = \rho(t)$ and $P = P(t)$. Substituting the RW-metric into the Einstein's equations (13.4), he derived (back in 1922) the following two independent equations governing the evolution of R, ρ and P:

$$\left(\frac{\dot{R}}{R}\right)^2 + \frac{kc^2}{R^2} = \frac{8\pi}{3}G\rho + \frac{\Lambda}{3}, \tag{13.5}$$

(Here $\dot{A} = dA/dt$.) This equation is often called the Friedmann equation.

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right) + \frac{\Lambda}{3}, \tag{13.6}$$

This equation is called the *acceleration equation*. To have a closed system, these equations have to be complemented with an equation of state (EOS), which relates P and ρ (so we have three equations for three unknowns). You may have noticed that none of these equations looks as an evolution equation for ρ – they do not involve time derivatives of ρ. This can be helped. Differentiate equation (13.5) once to obtain

$$2\left(\frac{\dot{R}}{R}\right)\left[\frac{\ddot{R}}{R} - \left(\frac{\dot{R}}{R}\right)^2 - \frac{kc^2}{R^2}\right] = \frac{8\pi}{3}G\dot{\rho}.$$

Substitute into the expression in square brackets $\ddot{R}/R$ from (13.6) and $\dot{R}/R$ from (13.5) to obtain

$$\dot{\rho} = -3\left(\frac{\dot{R}}{R}\right)\left(\rho + \frac{P}{c^2}\right), \tag{13.7}$$

Naturally, this equation is called the *fluid equation*. As we have noted in Sec.12, Einstein equations not only describe how matter (and radiation) warps spacetime but also how it evolves in such spacetime. Here we have a nice illustration of this property.

Section 14

Einstein's static solution.

We start our study of relativistic cosmological models with historically the first one, the model of static Universe, developed by Einstein. In our review, we proceed in a somewhat different to Einstein way and start with the set of Friedmann equations.

Let us seek steady-state solutions of the Friedmann equations for Universe filled with cold matter ($P = 0$). Put $\dot{R} = \ddot{R} = 0$ into the Friedmann equations to find

$$\frac{kc^2}{R^2} = \frac{8\pi}{3}G\rho + \frac{\Lambda}{3}\,, \tag{14.1}$$

and

$$-\frac{4\pi G}{3}\rho + \frac{\Lambda}{3} = 0\,. \tag{14.2}$$

From these, it is easy to obtain the Einstein's results

$$\frac{kc^2}{R^2} = \Lambda \tag{14.3}$$

and

$$4\pi G\rho = \Lambda\,. \tag{14.4}$$

Notice that $\Lambda = 0$ implies $\rho = 0$ (empty Universe) and $k = 0$ (flat Universe). Thus, static solutions with non-vanishing matter density are only possible if $\Lambda \neq 0$ (cf. Section 13.2). Moreover, since $\rho > 0$ only $\Lambda > 0$ are allowed, implying $k > 0$ and hence Universe with the geometry of Riemann's hypersphere.

The fact that a set of ODEs, describing a physical system, allows static solutions, does not mean that the system may settle to such a static state. For this the solution have to be stable. It turns out that the Einstein's static solution is not. Let us study the stability of this solution, following the original analysis by Friedmann. To this aim we need to use the time-dependent equations. For cold matter, equation (13.7) reads

$$\dot{\rho} = -3\frac{\dot{R}}{R}\rho\,. \tag{14.5}$$

This separable equation is easily integrated

$$
\begin{aligned}
\frac{1}{\rho}\frac{d\rho}{dt} &= -3\frac{dR}{dt}\frac{1}{R}, \\
\Rightarrow \frac{1}{\rho}d\rho &= -3\frac{1}{R}dR \\
\Rightarrow \int\frac{1}{\rho}d\rho &= -3\int\frac{1}{R}dR \\
\Rightarrow \ln\rho &= -3\ln R + \text{const} \\
\Rightarrow \ln\left(\rho R^3\right) &= \text{const} \\
\Rightarrow \rho R^3 &= \text{const}\,.
\end{aligned}
$$

Denoting the constant of integration as M we write

$$\boxed{\rho R^3 = M > 0\,.} \tag{14.6}$$

Substituting ρ from (14.6) into (13.6) we obtain

$$\frac{\ddot{R}}{R} = -\frac{4\pi}{3}\frac{GM}{R^3} + \frac{\Lambda}{3}, \tag{14.7}$$

For the static solution $R(t) = R_0$, this gives

$$\Lambda = 4\pi\frac{GM}{R_0^3} > 0\,. \tag{14.8}$$

Now, slightly perturb the static solution $R(t) = R_0\,(1+\varepsilon(t))$, where $\varepsilon \ll 1$. Substitute this into (14.7) to obtain

$$\frac{d^2}{dt^2}\left(R_0\,(1+\varepsilon)\right) = -\frac{4\pi GM}{3R_0^2}\frac{1}{(1+\varepsilon)^2} + \frac{\Lambda}{3}R_0\,(1+\varepsilon)\,. \tag{14.9}$$

Since for $\varepsilon \ll 1$ the Maclaurin expansion gives $(1+\varepsilon)^{-2} = 1 - 2\varepsilon + O(\varepsilon^2)$, equation (14.9) can be written as

$$R_0\ddot{\varepsilon} = -\frac{4\pi}{3}\frac{GM}{R_0^2}(1-2\varepsilon) + \frac{\Lambda}{3}R_0\,(1+\varepsilon) + O\left(\varepsilon^2\right). \tag{14.10}$$

Here $O(\varepsilon^2)$ denotes a *small* term of the order ε^2, which we can ignore when $\varepsilon \ll 1$. Hence

$$
\begin{aligned}
\ddot{\varepsilon} &= -\frac{4\pi}{3}\frac{GM}{R_0^3} + \frac{8\pi GM}{3R_0^3}\varepsilon + \frac{\Lambda}{3} + \frac{\Lambda}{3}\varepsilon \\
&= \left(\frac{\Lambda}{3} - \frac{4\pi GM}{3R_0^3}\right) + \left(\frac{8\pi GM}{3R_0^2} + \frac{\Lambda}{3}\right)\varepsilon.
\end{aligned}
$$

Using equation (14.8) we can write this as

$$\ddot{\varepsilon} = \Lambda\varepsilon. \tag{14.11}$$

Since $\Lambda > 0$ the general solution of (14.11) is

$$\varepsilon = Ae^{\sqrt{\Lambda}t} + Be^{-\sqrt{\Lambda}t}. \tag{14.12}$$

Only for very special (singular) initial conditions, one may have $A = 0$. For an arbitrary initial perturbation, one expects both $B \neq 0$ and $A \neq 0$ and because of the first term in eq.(14.12), the perturbation will not remain small but will grow, and hence the static solution is unstable.

Section 15

Friedmann's models of dynamic Universe.

Once Friedmann discovered that the static solution was unstable, he dismissed the modification of the GR equations, made by Einstein only to allow such static solutions, and instead focused of the original equations without the cosmological term. Following Einstein he assumed that $P = 0$ and proceeded with studying the time-dependent models. It turned out that the evolution of the Universe depended on the assumed spatial geometry and was qualitatively different for models with $k = 1$, $k = -1$ and $k = 0$.

Here we derive Friedmann's solutions and study their properties. For $\Lambda = P = 0$, the Friedmann equations read

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{R^2}\,, \tag{15.1}$$

$$\dot{\rho} = -3\left(\frac{\dot{R}}{R}\right)\rho\,, \tag{15.2}$$

$$\left(\frac{\ddot{R}}{R}\right) = -\frac{4\pi G}{3}\rho\,. \tag{15.3}$$

The last equation tells us that $\ddot{R} < 0$ and thus the *expansion of the Universe described by these equation will always be slowing down.* From this we may even speculate that at some point the Universe may start to contract. However, the answer depends of the value of k.

Repeating the calculations of Sec.13 we obtain

$$\rho R^3 = M\,(= \text{const.})\,, \tag{15.4}$$

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3}\frac{M}{R^3} - \frac{kc^2}{R^2}\,. \tag{15.5}$$

Equation (15.5) can be written as

$$\boxed{\dot{R}^2 = c^2\left(\frac{\alpha^2}{R} - k\right),} \tag{15.6}$$

where

$$\alpha^2 = \frac{8\pi G}{3}\frac{M}{c^2}\,. \tag{15.7}$$

15.1 The flat Universe ($k = 0$).

(The geometry of space is Euclidean, hence the name *flat Universe*.) For $k = 0$, eq.(15.6) reads

$$\dot{R}^2 = \frac{c^2\alpha^2}{R}.$$

From this equation it is clear that $\dot{R}$ cannot not vanish at any finite R and hence there can be no turning point in the evolution. *A now expanding Universe will keep expanding forever*. In this case,

$$\dot{R} = \frac{a}{\sqrt{R}},$$

where the constant a is positive ($a = +\sqrt{c^2\alpha^2}$). Hence

$$\begin{aligned}
R^{\frac{1}{2}}\dot{R} &= a \\
\Rightarrow R^{\frac{1}{2}}\,dR &= a\,dt \\
\Rightarrow \int R^{\frac{1}{2}}\,dR &= \int a\,dt \\
\Rightarrow \frac{2}{3}R^{\frac{3}{2}} &= at + b,
\end{aligned}$$

where $b =$ const. It is easy to see, that $R = 0$ when $t = -b/a$. Thus the solution implies the "moment of creation" for the Universe. Resetting the "time zero" (introducing new time variable equal to $t + b/a$) so that $R(0) = 0$, we get

$$R \propto t^{\frac{2}{3}}. \tag{15.8}$$

This result can also be written as

$$R = R_0\left(\frac{t}{t_0}\right)^{\frac{2}{3}}. \tag{15.9}$$

where $R_0 = R(t_0)$. Since

$$\dot{R} = \frac{2}{3}At^{-\frac{1}{3}} \to 0 \text{ as } t \to \infty, \tag{15.10}$$

the expansion rate always reduces with time and tends to zero, but never reaches it.

15.2 The open Universe ($k = -1$).

For $k = -1$ the geometry of space is Lobachevskian. Hence the name *open*, meaning infinite. Equation (15.6) now reads

$$\dot{R}^2 = c^2\left(\frac{\alpha^2}{R} + 1\right). \tag{15.11}$$

From this we have $\dot{R} \neq 0$ for any R and hence there can be no turning point as well – *the open Universe also expands forever*.

Equation (15.11) is still separable, but its integration is a bit more involved:

$$\frac{dR}{dt} = c\left(\frac{\alpha^2}{R}+1\right)^{\frac{1}{2}},$$
$$\Rightarrow \frac{dR}{\left(\frac{\alpha^2}{R}+1\right)^{\frac{1}{2}}} = c\,dt,$$
$$\Rightarrow \int \frac{R^{\frac{1}{2}}\,dR}{(\alpha^2+R)^{\frac{1}{2}}} = ct\,.$$

Introduce new variable x via

$$R = \alpha^2 \sinh^2 x, \quad x \in (0,+\infty)\,. \tag{15.12}$$

Hence,

$$dR = 2\alpha^2 \sinh x \cosh x\, dx$$

and

$$\int \frac{R^{\frac{1}{2}}\,dR}{(\alpha^2+R)^{\frac{1}{2}}} = \int \frac{2\alpha^2 \sinh x \cosh x\, dx\, \alpha \sinh x}{\alpha\left(1+\sinh^2 x\right)^{\frac{1}{2}}} = \int \frac{2\alpha^2 \sinh^2 x \cosh x}{\cosh x}\, dx = 2\alpha^2 \int \sinh^2 x\, dx,$$

(where we used $\cosh^2 x - \sinh^2 x = 1$).

$$\sinh x = \frac{e^x - e^{-x}}{2},$$
$$\Rightarrow \sinh^2 x = \left(\frac{e^x - e^{-x}}{2}\right)^2$$
$$= \frac{e^{2x} + e^{-2x} - 2}{4}$$
$$= \frac{1}{2}\left(\cosh 2x - 1\right)$$

Hence,

$$\int \sinh^2 x\, dx = \frac{1}{2}\int \left(\cosh 2x - 1\right)\, dx$$
$$= \frac{1}{2}\left(\frac{1}{2}\sinh 2x - x\right) + A\,,$$

where $A = \text{const}$. Hence,

$$\alpha^2\left(\frac{1}{2}\sinh 2x - x\right) = ct + A.$$

Resetting "time zero" so that $R(0) = 0$, we obtain $A = 0$ and hence

$$\alpha^2\left(\frac{1}{2}\sinh 2x - x\right) = ct\,. \tag{15.13}$$

Together, equations (15.12) and (15.13) define the function $R(t)$ in an implicit way.

It is easy to show that in the case of open Universe $\dot{R}$ *does not vanish as* $t \to +\infty$ *but approaches the speed of light.*

15.3 The closed Universe ($k = +1$).

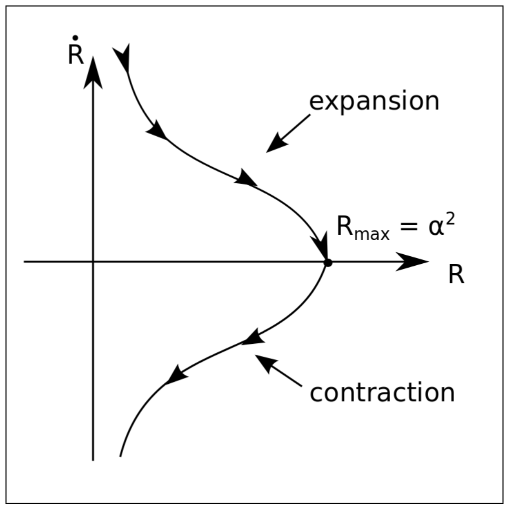
Now the geometry of space is that of Riemann's hypersphere. Hence the name *closed* (wrapped onto itself). For $k = +1$, equation (15.6) reads

$$\dot{R}^2 = c^2\left(\frac{\alpha^2}{R} - 1\right). \tag{15.14}$$

Now $\dot{R} = 0$ when $R = R_{max} = \alpha^2$. Hence there can be a turning point. From equation (15.3) we have

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\rho < 0.$$

Hence the turning point is a maximum and so the initial expansion of the Universe will eventually turn into a contraction.



Equation (15.14) yields

$$\dot{R} = \pm c\left(\frac{\alpha^2}{R} - 1\right)^{\frac{1}{2}}, \tag{15.15}$$

where "+" corresponds to the expansion phase and "-" corresponds to the contraction phase. Since $\dot{R}$ does not vanish when $R = 0$ we immediately conclude that *the closed Universe eventually collapses*.

Let us show that the contraction phase is a mirror image of the expansion phase, that is

$$R(t_m - \tau) = R(t_m + \tau), \tag{15.16}$$

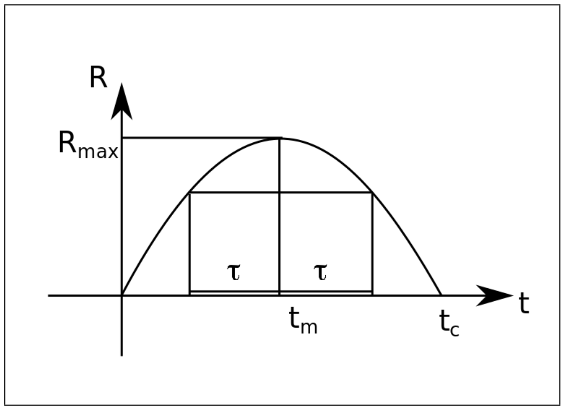
where t_m is the time of reaching R_{max}.

From (15.15) we have

$$\dot{R} = \begin{cases} +\sqrt{f(R)}, & t < t_m, \\ -\sqrt{f(R)}, & t > t_m. \end{cases} \tag{15.17}$$

Introduce τ_1, τ_2 such that $t = t_m - \tau_1$ for $t < t_m$ and $t = t_m + \tau_2$ for $t > t_m$, ($\tau_1, \tau_2 > 0$). Hence,

$$\frac{dR}{d\tau_1} = -\frac{dR}{dt} = -\sqrt{f\left(R(t_m - \tau_1)\right)},$$
$$\frac{dR}{d\tau_2} = +\frac{dR}{dt} = -\sqrt{f\left(R(t_m + \tau_2)\right)}.$$



Now we connect τ_1 and τ_2 via the condition $R(t_m - \tau_1) = R(t_m + \tau_2)$. Hence,

$$\begin{aligned} \frac{dR}{d\tau_1} &= \frac{dR}{d\tau_2}, \\ \Rightarrow \frac{d\tau_1}{dR} &= \frac{d\tau_2}{dR}, \\ \Rightarrow \tau_1 - \tau_2 &= \text{const.} \end{aligned}$$

Since $\tau_1 = \tau_2 = 0$ when $R = R_{max}$ we have $\tau_1 = \tau_2 \equiv \tau$. Hence,

$$R(t_m - \tau) = R(t_m + \tau)$$

indeed.

Integrating equation (15.15) for $t < t_m$ one finds

$$R = \alpha^2 \sin^2 x, \tag{15.18}$$

$$\alpha^2 \left(x - \frac{1}{2} \sin 2x \right) = ct. \tag{15.19}$$

(We used the IC $R(0) = 0$ again. The derivation is very similar to that of the open Universe case.) It is easy to show that (15.18) and (15.19) describe the evolution for $t > t_m$ as well.

Time of turning: From (15.18) we have that

$$R = R_{max} \quad \text{when} \quad x = x_m = \frac{\pi}{2}.$$

Substitute this into (15.19) to get

$$\begin{aligned} \alpha^2 \left(\frac{\pi}{2} - 0 \right) &= ct_m, \\ \Rightarrow t_m &= \frac{\pi}{2} \frac{\alpha^2}{c}. \end{aligned} \tag{15.20}$$

Time of collapse: From (15.18) we have that

$$R = 0 \quad \text{when} \quad x = 0, \pi.$$

From (15.19)

$$
\begin{aligned}
x = \pi \Rightarrow \alpha^2 (\pi - 0) &= ct_c, \\
\Rightarrow t_c &= \pi \frac{\alpha^2}{c}.
\end{aligned}
\tag{15.21}
$$

As expected from the symmetry, $t_c = 2t_m$.

Section 16

The Hubble law and the expansion of the Universe.

16.1 The observations

In 1929 American astronomer Edwin Hubble discovered that the emission lines in spectra at distant galaxies are systematically shifted towards the red part of the electromagnetic spectrum. Denote as λ_e the wavelength of the electromagnetic wave (photon) as measured at the time of emission at the source in the local inertial frame co-moving with the source and as λ_r the wavelength of this wave (photon) as measured in the local inertial frame co-moving with the telescope at the time of its arrival. Then according to Hubble $\lambda_r > \lambda_e$. The relative value of the change,

$$z = \frac{\lambda_r - \lambda_e}{\lambda_e} > 0 \tag{16.1}$$

is called the cosmological redshift (parameter of the source). He also discovered that z is proportional to the distance l to the galaxy:

$$z = \frac{H_0}{c} l, \tag{16.2}$$

where H_0 is called the Hubble constant. What are the theoretical interpretations of this result.

The change of wavelength (frequency) can be caused by the motion of the source relative to the observer – this is known as the *Doppler effect*. In particular, for electromagnetic waves, Special Relativity gives

$$z = \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}} - 1, \tag{16.3}$$

where $v = dl/dt$ is the speed of the source of light relative to the observer. Notice that $z > 0$ for $v > 0$ and $z < 0$ for $v < 0$. For $v \ll c$, this reduces to the Newtonian result

$$z \simeq \frac{v}{c}, \tag{16.4}$$

which allows to rewrite (16.2) as

$$v = H_0 l\,. \tag{16.5}$$

The typical velocity of galaxies observed by Hubble is $v \approx 100$km/s, $v/c = 10^{-4} \ll 1$ and the Newtonian approximation suffices. Their typical distance $l \approx$ few Mpc (1Mpc $= 10^6$pc and 1pc $\approx 3 \times 10^{18}$cm). So the most convenient unit for H_0 is (km/s)Mpc^{-1}. Based on the observations,

$$H_0 = 100h \text{ km/s Mpc}^{-1}\,, \tag{16.6}$$

where $h \approx 1$ stands for the uncertainty due to observational errors (different observational groups used to claim rather different values for h).

If we adopt the Doppler effect as the reason behind the Hubble's results, the other galaxies move away from us with speed which increases with the distance from us. This put us in the centre of the Universe once more and hence this is not a suitable explanation. The modern explanation is based on the expansion of the Universe itself instead. To see the difference, imagine bugs sitting quietly on a balloon. When the balloon is inflated the distance between the bugs grows, even if they do not move. To any such bug, other bugs appear to be moving away from it, with speed increasing with the distance.

16.2 The modern interpretation

Let us see how Relativistic Cosmology explains and in fact generalises Hubble's law in details. The Robertson-Walker metric is:

$$ds^2 = -c^2dt^2 + R^2(t)\left[\frac{d\chi^2}{1-k\chi^2} + \chi^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right].$$

Using the coordinate system where the observer is at the origin ($\chi = 0$), consider a photon moving radially towards the observer. Since for such a photon $d\theta = d\phi = 0$, along its world-line

$$ds^2 = -c^2dt^2 + R^2(t)\frac{d\chi^2}{1-k\chi^2}. \tag{16.7}$$

Since for any photon $ds^2 = 0$,

$$c^2dt^2 = R^2(t)\frac{d\chi^2}{1-k\chi^2} \tag{16.8}$$

and

$$\Rightarrow c\,dt = -R(t)\frac{d\chi}{\sqrt{1-k\chi^2}}. \tag{16.9}$$

(the sign "-" is determined by the direction of motion, $d\chi < 0$ for $dt > 0$.) Hence

$$c\int_{t_e}^{t_o}\frac{dt}{R(t)} = -\int_{\chi_e}^{0}\frac{d\chi}{\sqrt{1-k\chi^2}}, \tag{16.10}$$

where t_e is the time of the emission of the photon, t_o is the time of the observation and χ_e is the coordinate of the source, which is assumed to at rest in space (χ_e =const).

Consider a second photon, emitted at time $t_e + dt_e$ and received at time $t_o + dt_o$. Then

$$c\int_{t_e+dt_e}^{t_o+dt_o}\frac{dt}{R(t)} = -\int_{\chi_e}^{0}\frac{d\chi}{\sqrt{1-k\chi^2}}. \tag{16.11}$$

From (16.10) and (16.11)

$$\int_{t_e}^{t_o}\frac{dt}{R(t)} = \int_{t_e+dt_e}^{t_o+dt_o}\frac{dt}{R(t)} = \int_{t_e}^{t_o}\frac{dt}{R(t)} + \frac{dt_o}{R(t_o)} - \frac{dt_e}{R(t_e)}, \tag{16.12}$$

$$\Rightarrow \boxed{\frac{dt_o}{dt_e} = \frac{R(t_o)}{R(t_e)}.} \tag{16.13}$$

The time interval separating these two photons at the observer is different from that at the source.

Instead of two photons we could consider two successive crests of a light wave with the period T_e at the source and T_r at the observer. Provided the period is reasonably short (so that R does not change much during one period) we can replace dt_o with T_r and dt_e with T_e and obtain

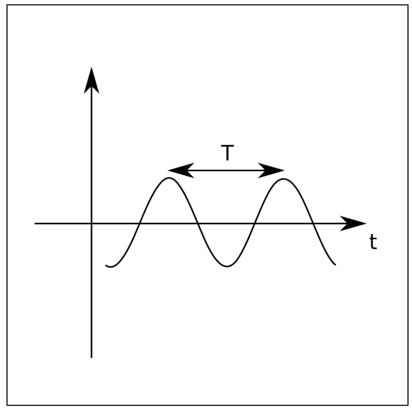


Figure 16.1: Two crests of a monochromatic wave.

$$\frac{T_r}{T_e} = \frac{R(t_o)}{R(t_e)}. \tag{16.14}$$

For electromagnetic waves $T = \lambda/c$. Hence

$$\frac{\lambda_r}{\lambda_e} = \frac{R(t_o)}{R(t_e)}.$$

The wavelengths differ! The corresponding redshift is

$$z = \frac{\lambda_r - \lambda_e}{\lambda_e} = \frac{\lambda_r}{\lambda_e} - 1 = \frac{R(t_o)}{R(t_e)} - 1. \tag{16.15}$$

Since Hubble observed $z > 0$ this implies

$$R(t_o) > R(t_e).$$

Thus, our Universe is expanding! The equation

$$\boxed{\frac{R(t_o)}{R(t_e)} = 1 + z} \tag{16.16}$$

is called the *generalised Hubble law*. At first sight, it does not look like the original Hubble law (eq.16.2), but for $z \ll 1$ it reduces to one. To show this, we first notice that $z \ll 1$ implies

$$R(t_o) \simeq R(t_e)\,.$$

This allows us to replace $R(t)$ with $R(t_0)$ in Eq.(16.10) and obtain

$$c\int_{t_e}^{t_o} \frac{dt}{R(t_o)} \simeq \int_0^{\chi_e} \frac{d\chi}{\sqrt{1-k\chi^2}},$$

which gives us

$$c\,(t_o - t_e) \simeq R(t_o) \int_0^{\chi_e} \frac{d\chi}{\sqrt{1-k\chi^2}}\,.$$

Since on the right side we have nothing else but the current distance l to the source, we have obtained that

$$l \simeq c\,(t_o - t_e)\,. \tag{16.17}$$

Form the Maclaurin expansion for $R(t)$ about $t = t_0$, we have

$$\begin{aligned}
R(t_e) &\simeq R(t_o) + \dot{R}(t_o)\,(t_e - t_o) \\
&= R(t_o)\left(1 + \frac{\dot{R}(t_o)}{R(t_o)}\,(t_e - t_o)\right) \\
&\simeq R(t_o)\left(1 - \frac{\dot{R}(t_o)}{R(t_o)}\frac{l}{c}\right) \\
&= R(t_o)\left(1 - A_0\frac{l}{c}\right),
\end{aligned}$$

where $A_0 = \dot{R}(t_o)/R(t_o)$. Substitute this into equation (16.16):

$$1 + z = \frac{R(t_o)}{R(t_e)} = \frac{1}{1 - A_0\frac{l}{c}}.$$

Since $R(t_e)/R(t_o) \approx 1$, we have $A_0\frac{l}{c} \ll 1$ and using the first two terms of the Maclaurin expansion

$$\frac{1}{1-x} = 1 + x + \dots$$

we obtain

$$1 + z \simeq 1 + \frac{A_0}{c}l\,,$$

and hence

$$z \simeq \frac{A_0}{c}l\,. \tag{16.18}$$

Comparing this with the original Hubble result

$$z = \frac{H_0}{c}l,$$

we conclude that the generalised Hubble law yields the original one with

$$\boxed{H_0 = \frac{\dot{R}(t_o)}{R(t_o)}.} \tag{16.19}$$

in the limit of small redshift. *The Hubble constant measures the rate of the current expansion of the Universe!*

16.3 The Hubble constant and the age of the Universe.

Suppose that $R \propto t^a$ (or $R = At^a$). Then

$$\dot{R} = \frac{a}{t}R$$

and

$$\frac{\dot{R}}{R} = \frac{a}{t} = H(t).$$

Hence

$$H_0 \equiv H(t_0) = \frac{a}{t_0}$$

and the current age of the Universe

$$\boxed{t_0 = \frac{a}{H_0}.} \tag{16.20}$$

Thus, *the Hubble constant is a good indicator of the age of the Universe.*

In the Friedmann's model of *flat Universe*, $a = \frac{2}{3}$ and

$$t_0 = \frac{2}{3}\frac{1}{H_0}\,. \tag{16.21}$$

Let us estimate the age of the Universe.

$$H_0 = 100h\frac{\text{km/s}}{\text{Mpc}} \approx \frac{100h \times 10^5 \text{cm/s}}{10^6 \times 3 \times 10^{18} \text{cm}} = \frac{h}{3 \times 10^{17}} s^{-1}\,.$$

Hence

$$\frac{1}{H_0} \approx 3 \times 10^{17} h^{-1} s \approx 10^{10} h^{-1} \text{years}, \tag{16.22}$$

($1yr \approx 3 \times 10^7 s$). *Thus, the typical age of the Universe spans billions of years*.

Section 17

The Big Bang.

All three Friedmann models give $R = 0$ in the past. This hints that Universe did not exist forever but was born some time ago from a point (a singularity). Close to the $R = 0$ moment, the physical conditions become so extreme that the assumption of matter being cold ($P \ll \rho c^2$), used in Friedmans models, fails. Yet, this does not stop solutions from reaching the singularity in the past.

Indeed, for cold matter ($P \ll \rho c^2$) we have

$$\rho \propto \frac{1}{R^3} \to \infty \quad \text{as} \quad R \to 0 \tag{17.1}$$

(see Section 15). Thus, mass-energy density diverges. In fact, the gas pressure and temperature also diverge in this limit. Indeed

$$P \propto \rho^{\Gamma}, \tag{17.2}$$

where $\Gamma \approx 5/3$ for a non-relativistic gas and

$$\frac{P}{\rho c^2} \propto \rho^{\Gamma-1} \to \infty \quad \text{as} \quad R \to 0\,. \tag{17.3}$$

Gas with $P \gg 1$ becomes relativistic and instead of $\Gamma = 5/3$ we need to use $\Gamma = 4/3$. Moreover, one has to use the radiational EOS with $P = \rho c^2/3$ instead of $P = 0$ of cold matter. However, this *does not* lead to qualitatively different results as the initial singularity is still there.

Thus the cosmological solutions of GR equations suggest that *the Universe was created in one big explosion*, which is referred to as the *Big Bang*. Just like many normal explosions produce a powerful flash of light, a lot of light was produced during the Big Bang. Initially, this light was trapped inside dense matter but as the Universe had expanded beyond certain size and become transparent it decoupled from matter and its photons have been freely moving through space ever since, unaffected by anything but the expansion of the Universe (see the Hubble law). This radiation was discovered in 1965 by Penzias and Wilson (Nobel prize 1968) as the cosmic microwave background (CMB) radiation. It has a typical *"black body spectrum"*, which peaks at microwaves. Its temperature is $T \approx 2.7K$. It appears to come from all directions *isotropically* (almost isotropically, as there are tiny fluctuations). When these photons decoupled there tenperature was $T \approx 3000^{o}$K. This is determined mainly by the atomic physics, which is very well known indeed. For the black body spectrum, the mean energy of photons is proportional to temperature. Thus, now the CMB photons are about a thousand times less energetic than at the time of decoupling. As shown below, this cooling is a result of the very substantial expansion of the Universe since the time of decoupling.

The photon's energy is given by the Planck equation

$$\boxed{E = h\nu,} \tag{17.4}$$

where h is a universal constant, known as the Planck constant, and ν is the photon frequency,

$$\nu = \frac{c}{\lambda}. \tag{17.5}$$

Using the generalised Hubble law (16.16),

$$\frac{E_r}{E_d} = \frac{\nu_r}{\nu_d} = \frac{\lambda_d}{\lambda_r} = \frac{R(t_d)}{R(t_0)}, \tag{17.6}$$

where suffix d refers to the parameters at the time of decoupling and suffix r to the current parameters. Since $E_r/E_d \approx 1/1000$, the Universe has expanded by the factor of one thousand since the decoupling.

Section 18

The critical density.

Consider the Friedmann equation with $\Lambda = 0$,

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{R^2}. \tag{18.1}$$

Since $\dot{R}/R = H(t)$, the Hubble constant at time t, Eq.(18.1) can be written as

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{R^2},$$

and hence

$$\rho - \frac{3}{8\pi G}H^2 = k\frac{3c^2}{8\pi G R^2}.$$

Using the *critical density*

$$\boxed{\rho_c = \frac{3H^2}{8\pi G}.} \tag{18.2}$$

this result can be presented as

$$\rho - \rho_c = k\frac{3c^2}{8\pi G R^2}. \tag{18.3}$$

Hence

$$\boxed{k = \begin{cases} 1, & \text{if } \rho > \rho_c, \\ 0, & \text{if } \rho = \rho_c, \\ -1, & \text{if } \rho < \rho_c. \end{cases}} \tag{18.4}$$

From this it is clear that *the flat Universe ($k = 0$) is a very special (singular) case*. We should expect the Universe to be either closed or open. The current critical density

$$\rho_{c_0} = \frac{3H_0^2}{8\pi G} \approx 1.9h^2 \times 10^{-29}\,\frac{\text{gram}}{\text{cm}^3}$$

is very small. From the astronomical prospective, a more appropriate unit is galaxies/Mpc^3. Taking the typical galaxy mass of 10^{11} solar masses, we obtain

$$\rho_{c_0} \approx 2.78h^2\,\frac{\text{galaxies}}{\text{Mpc}^3}.$$

From the astronomical observations of *visible matter*

$$\rho_{0,vis} \approx 0.05\rho_c,$$

which implies an open Universe. However, this is not the whole story as there are other components (in addition to the visible matter) in the Universe, which we will discuss later.

Another useful parameter is the so-called *critical parameter* Ω

$$\boxed{\Omega = \frac{\rho}{\rho_c}.} \tag{18.5}$$

Using Ω, one can rewrite (18.3) as

$$\boxed{1 - \Omega = \frac{kc^2}{R^2H^2}\,.} \tag{18.6}$$

Section 19

”Real” distances to sources with known redshift.

19.1 General results

What is the current distance from the origin ($\chi = 0$) to the point with the comoving coordinate $\chi = \chi_e$? The RW-metric at *present time* is:

$$ds^2 = -c^2 dt^2 + dl^2,$$

where

$$dl^2 = R_0^2 \left(\frac{d\chi^2}{1-k\chi^2} + \chi^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right).$$

Along the radial direction $d\theta = d\phi = 0$ and

$$dl = R_0 \frac{d\chi}{\sqrt{1-k\chi^2}}.$$

Hence the distance is

$$\boxed{r = R_0 \int_0^{\chi_e} \frac{d\chi}{\sqrt{1-k\chi^2}}\,.} \tag{19.1}$$

This gives us

$$r = \begin{cases} R_0 \arcsin(\chi_e) & \text{if} \quad k = 1\,, \\ R_0 \chi_e & \text{if} \quad k = 0\,, \\ R_0 \mathrm{arcsinh}(\chi_e) & \text{if} \quad k = -1\,. \end{cases} \tag{19.2}$$

In astronomical observations we do not measure χ_e and hence the above question is of theoretical nature. More practical question would be “*What is the distance from us to a remote galaxy with the redshift z?*” In order to answer this question, one needs to know not only the current geometry of the Universe but also the history of its expansion prior to the time of observation.

Along the world-lines of photons, streaming towards us from the source, $ds^2 = 0$ and hence

$$c^2 dt^2 = R^2(t) \frac{d\chi^2}{1-k\chi^2},$$

where t spans from the time of emission t_e to the time of observation t_o. Hence

$$\Rightarrow \quad c\,dt = -R(t)\frac{d\chi}{\sqrt{1-k\chi^2}},$$

and

$$c\int_{t_e}^{t_o}\frac{dt}{R(t)} = -\int_{\chi_e}^{0}\frac{d\chi}{\sqrt{1-k\chi^2}}. \tag{19.3}$$

Substituting this into (19.1), we obtain the neat result

$$\boxed{r(z) = cR_0\int_{t_e}^{t_o}\frac{dt}{R(t)}}\,. \tag{19.4}$$

In order to proceed, we need to know $R(t)$, which is different for different Friedmann models.

19.2 The case of Friedmann's flat Universe

Consider the Friedmann's model of flat Universe, ($k = 0, P = 0, \Lambda = 0$), which gives $R(t)$ as an explicit function. Namely,

$$R(t) = R_0\left(\frac{t}{t_o}\right)^{2/3}. \tag{19.5}$$

Then (19.4) reads

$$r(z) = ct_o{}^{2/3}\int_{t_e}^{t_o} t^{-2/3}\,dt = 3ct_o\left(1-\left(\frac{t_e}{t_o}\right)^{1/3}\right). \tag{19.6}$$

From (19.5)

$$\frac{t_e}{t_o} = \left(\frac{R(t_e)}{R(t_o)}\right)^{3/2} = (1+z)^{-3/2}\,. \tag{19.7}$$

Substitute this into (19.6) to obtain

$$\boxed{r(z) = 3ct_o\left(1-(1+z)^{-1/2}\right).} \tag{19.8}$$

Finally, we can substitute $t_o = 2/3H_0$ (see Section 16) and obtain the equation

$$\boxed{r(z) = \frac{2c}{H_0}\left(1-(1+z)^{-1/2}\right),} \tag{19.9}$$

which contains only observable parameters. (Other models of the Universe give other solutions for $r(z)$.) For $z \ll 1$ this reduces to the original Hubble law,

$$r(z) \simeq \frac{c}{H_0}z.$$

However, in general, $r(z)$ grows slower with z. In fact,

$$r(z) \to \frac{2c}{H_0} \quad \text{as} \quad z \to +\infty,$$

which tells us that in order to be seen a source must be located the distance $r < r_{h_o}$, where

$$\boxed{r_{h_o} = \frac{2c}{H_0}.} \tag{19.10}$$

Since in this model the Universe is infinite and homogeneous, there are sources beyond this distance but they cannot be observed. The surface $r = r_{h_o}$ is called the (current) *cosmological horizon* of the Universe.

Section 20

The cosmological horizon.

Equation (19.7) gives us a hint on the nature of the cosmological horizon:

$$t_e = t_0(1+z)^{-3/2}.$$

From this we see that $t_e \to 0$ as $z \to \infty$. Thus $z = \infty$ corresponds to emission produced at the time of Big Bang! Since the speed of light is finite, a photon can travel only a finite distance since the Big Bang and hence there must be a limit to how far we can see. Sources beyond the cosmological horizon exist but they cannot be seen by us simply because *their emission has not reached us yet*. Hence we need to answer the question:

How far can a photon travel during the life-time of the Universe?

This time we will use the RW-metric with the origin at the point of the emission. Along the photon's world-line

$$\begin{aligned} c^2 dt^2 &= R^2(t)\frac{d\chi^2}{1-k\chi^2}, \\ \Rightarrow c\,dt &= R(t)\frac{d\chi}{\sqrt{1-k\chi^2}}, \end{aligned} \tag{20.1}$$

(now $d\chi > 0$ for $dt > 0$, hence the sign "+" in (20.1)),

$$\Rightarrow \int_0^t \frac{c\,dt}{R(t)} = \int_0^{\chi(t)} \frac{d\chi}{\sqrt{1-k\chi^2}}. \tag{20.2}$$

Here $\chi(t)$ is the coordinate of the photon at time t. The distance from the origin at this time is

$$r_h(t) = R(t)\int_0^{\chi(t)} \frac{d\chi}{\sqrt{1-k\chi^2}}. \tag{20.3}$$

Using (20.2) this can be written as

$$\boxed{r_h(t) = c\,R(t)\int_0^t \frac{dt}{R(t)}.} \tag{20.4}$$

As an example, consider the *Friedmann's model of flat Universe*. In this model,

$$R(t) = R_0\left(\frac{t}{t_o}\right)^{2/3},$$

and (20.4) yields

$$r_h(t) = c\,t^{2/3}\int_0^t t^{-2/3}\,dt = 3ct^{2/3}\left[t^{1/3}\right]_0^t = 3ct.$$

Since in this model the current value of the Hubble "constant" is related to time via $t = \frac{2}{3}\frac{1}{H(t)}$ we finally obtain

$$\boxed{r_h(t) = \frac{2c}{H(t)}\,.} \tag{20.5}$$

This agrees with the radius of the cosmological horizon at present time, which was discovered in the previous section.

Two important comments:

1. Notice that equation (20.5) implies $r_h(t) > ct$. This is because not only the photons make their way through the Universe, but the Universe itself is expanding constantly, increasing distances between the points of its space.

2. The existence of the cosmological horizon poses the causality paradox of Friedmann's cosmology: *How can the Universe be uniform if it consists of causally disconnected parts?*

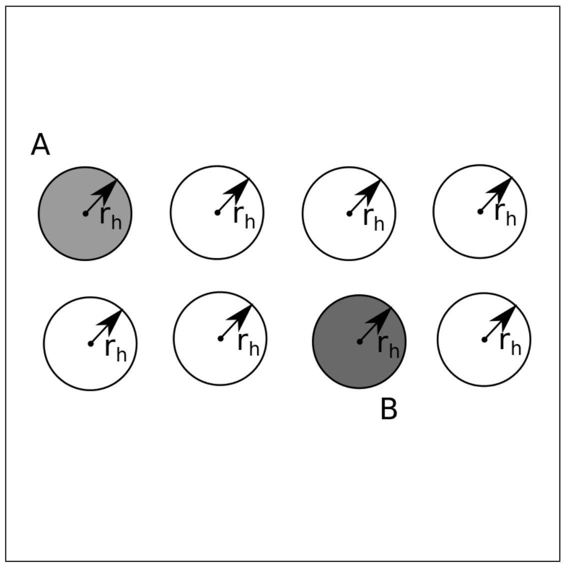


Figure 20.1:

Indeed, in order for the part A to adjust its state to the state of part B it has "to know" what the state of B is. However, as one can see in the figure, the corresponding information has not had enough time yet to travel all the way from B to A.

Section 21

The standard bar method of exploring geometry of the Universe.

21.1 The ”actual” angular size.

Consider a bar of the length l at the distance $r \gg l$ from the observer. Suppose this bar is perpendicular to the line of sight, the geodesic connecting the observer with the centre of the bar. The angular size $\alpha \ll 1$ of this bar is the angle between the geodesics connecting the observer with the end points of the bar. In Euclidean geometry,

$$l = 2r\tan(\alpha/2) \approx r\alpha\,.$$

(This result is exact for an arc of a circle centred on the observer.) Thus, one can find the distance to the bar as

$$r = \frac{l}{\alpha}\,. \tag{21.1}$$

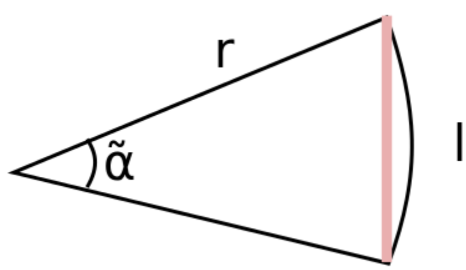


Figure 21.1: Angular and linear sizes of an object.

However, for Riemann’s hypersphere and Lobachevsky’s hyperbolic space the results differ. To see this, we can just do the calculations for an arc. The starting point is the RW-metric in the generalised spherical coordinates $\{r,\theta,\phi\}$ (see Sections 6 and 7):

$$dl^2 = \begin{cases} dr^2 + R^2\sin^2(r/R)\left(d\theta^2 + \sin^2\theta d\phi^2\right), & k=1,\\ dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), & k=0,\\ dr^2 + R^2\sinh^2(r/R)\left(d\theta^2 + \sin^2\theta d\phi^2\right), & k=-1. \end{cases} \tag{21.2}$$

To simplify the calculations, we can choose such a coordinate system that its origin is at the observer and the arc is aligned with a θ coordinate line. With such a choice, $dr = 0$, $d\phi = 0$, and

$$dl = dl_\theta = \begin{cases} R\sin(r/R)d\theta\,, & k=1\,, \\ r\,d\theta\,, & k=0\,, \\ R\sinh(r/R)d\theta, & k=-1\,, \end{cases} \tag{21.3}$$

along the arc. Integration along the arc gives

$$\boxed{l = \begin{cases} R\sin(r/R)\Delta\theta, & k=1, \\ r\Delta\theta, & k=0, \\ R\sinh(r/R)\Delta\theta, & k=-1\,. \end{cases}} \tag{21.4}$$

This result hints at how in principle one may try to determine which of the three geometries describes the real world. Just measure angular sizes of standard bars placed at various distances from us and check which of the equations in (21.4) fits the data. In fact, this would work for a static Universe. However, things are a bit more involved for an expanding Universe.

21.2 The observed angular size.

In an expanding Universe, the observed angular size θ_{ob} will be the same as the actual one at the time of observation, θ, only if the "bar" participates in the expansion of the Universe, meaning that is its length grows like R. For a solid bar, this is not the case as the atomic forces strongly depend on the distance between atoms and hence do not allow it to change. Similarly, the gravitational attraction keeps the mean distance between stars in a stellar cluster, and hence the cluster size fixed in time. The same applies to other gravitationally bound systems.

However, *the observed angular size always equals to the actual one at the emission time* of the photons which are received during the observation. This is illustrated in the figure below where it is assumed that the bar's length remains unchanged during the expansion of the Universe.

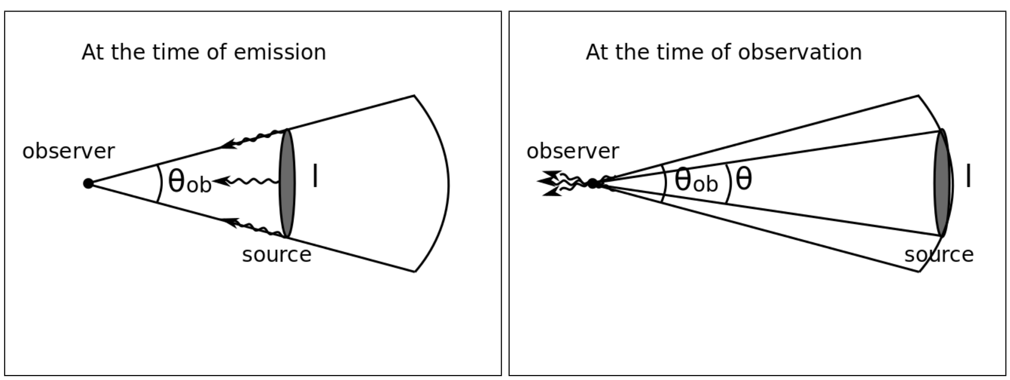


Figure 21.2:

Hence,

$$l = \begin{cases} R(t_e)\sin(r(t_e)/R(t_e))\theta_{ob}, \\ r(t_e)\theta_{ob}, \\ R(t_e)\sinh(r(t_e)/R(t_e))\theta_{ob}. \end{cases} \tag{21.5}$$

21.3 The standard bar method.

Observations of objects with known linear sizes allow us to explore the geometry of the Universe. To show this, consider the example of Friedmann's flat Universe. Hence

$$\theta_{ob} = \frac{l}{r(t_e)}\,. \tag{21.6}$$

Moreover, in this model

$$r(t_0) = \frac{2c}{H_0}\left(1-(1+z)^{-1/2}\right),$$

(see Sec. 19). Since

$$r(t_e) = r(t_0)\left(\frac{R(t_e)}{R(t_0)}\right) = r(t_0)\,(1+z)^{-1}\,,$$

we obtain

$$r(t_e) = \frac{2c}{H_0}\frac{\left(1-(1+z)^{-1/2}\right)}{1+z}\,. \tag{21.7}$$

and finally,

$$\boxed{\theta_{ob}(z) = l\frac{H_0}{2c}\frac{1+z}{\left(1-(1+z)^{-\frac{1}{2}}\right)}.} \tag{21.8}$$

Both θ_{ob} and z are observable parameters. Hence from observations of some astronomical objects of fixed l one can determine the function $\theta_{ob}(z)$ directly and then check if it is consistent with that of Eq.21.8. If it is not then the Friedmann's model of flat Universe is not a suitable one.

Naively, one would expect θ_{ob} to decrease with distance, and hence z. However, using (21.8) we find

$$\theta_{ob} = \begin{cases} (lH_0/c)(1/z), & z \ll 1, \\ (lH_0/2c)\,z, & z \gg 1. \end{cases} \tag{21.9}$$

These asymptotes suggest that the angular size is minimum at some $z_{\min}$. In fact, simple analysis yields $z_{\min} = 1.25$.

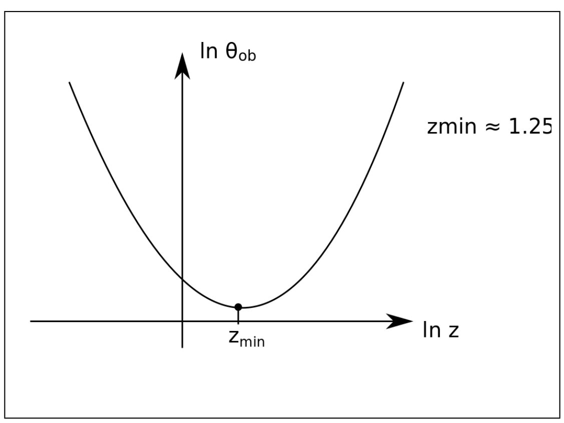


Figure 21.3:

For $z < z_{\min}$ the angular size of an object of fixed linear size decreases with distance (as expected) but *for $z > z_{\min}$ it increases*. Equation 21.7 provides us with a simple explanation to this paradoxic result. As z increases, the distance to the source at the time of emission first grows but then begins to decrease. Thus, in the case of large z, sources with larger z are closer to the Sun and hence have larger real angular size at the time of emission.

The role of "standard bar" can be played by galaxies, clusters of galaxies etc. But the fluctuations of the Cosmic Microwave Background have been the most useful so far. They are located at huge distances corresponding to $z \simeq 1000$ and their predicted angular scale is very sensitive to parameters of cosmological models. *The best fit to the CMB observations is given by models with*

$$\boxed{\Omega_0 \simeq 1\,,} \tag{21.10}$$

ıwhich hints the our Universe may be flat ($k = 0$). Since the estimates based on the mass of visible matter give much smaller critical parameter, $\Omega_{0,\mathrm{vm}} \simeq 0.02h^{-2}$, this tells us that *in addition to the visible matter, and radiation, there must be some invisible components in the Universe, which account for most of its mass!*

In Cosmology, the parameter

$$r_{\mathrm{ang}} = \frac{l}{\theta_{ob}} \tag{21.11}$$

is called the *angular size distance*. This would be a real distance in a static Universe with Euclidean geometry of space. In reality, it reasonably approximates the real distance only for very close sources, $z \ll 1$.

Section 22

The standard candle method.

22.1 The basic idea.

Consider a source of electromagnetic radiation (photons). Denote as $d\mathscr{E}_e$ the amount of energy emitted by the source during the time dt_e as measured in the source frame. Then

$$\boxed{L = \frac{d\mathscr{E}_e}{dt_e}} \tag{22.1}$$

is called the *source luminosity*. This in not a directly observable parameter. What we measure directly is the source brightness (or energy flux density)

$$\boxed{S = \frac{d\tilde{\mathscr{E}}_r}{dt_r\, dA}} \tag{22.2}$$

where dA is the surface element at the observer's location normal to the direction to the source, and $d\tilde{\mathscr{E}}_r$ is the amount of energy (emitted by the source) which crosses the surface element during the time dt_r. What is the exact connection between S at the time of observation and L at the time of the emission of the observed radiation?

Consider a sphere, which is centred on the source has the radius r equal to the distance to the observer. Denote as $d\mathscr{E}_r$ the energy flowing across the sphere during the time dt_r When the source emission is isotropic, S is constant over the sphere and hence

$$d\mathscr{E}_r = A\frac{d\tilde{\mathscr{E}}_r}{dA} = SAdt_r.$$

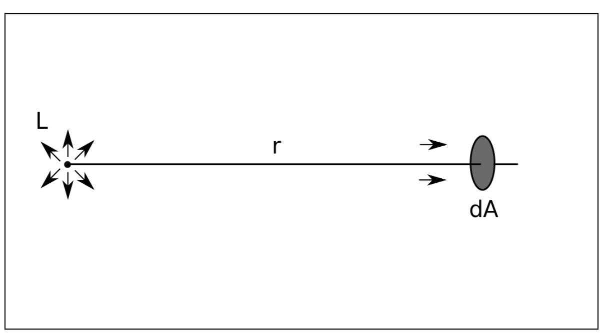


Figure 22.1:

In a transparent non-expanding Universe, the energy $d\mathscr{E}_r = d\mathscr{E}_e$ would cross the sphere during the time $dt_r = dt_e$ and hence

$$L = SA\,. \tag{22.3}$$

In a Universe with Euclidean geometry, we would also have $A = 4\pi r^2$ and hence

$$\boxed{r = \left(\frac{L}{4\pi S}\right)^{1/2}.} \tag{22.4}$$

This shows how to deduce the distance to a source with known luminosity based on its observed brightness in such a simple Universe. In a Universe with more complicated dynamics and geometry, the distance given by Eq.22.4 is not the same as the actual distance. It is called "the luminosity distance".

22.2 The standard candle method in modern cosmology.

In an expanding Universe with non-Euclidean geometry, three new features emerge and modify Eq.22.4.

(a) The energy is not conserved. The energy of photons decreases as they travel across the Universe (see Section 17):

$$E_r = E_e \cdot (1+z)^{-1}\,.$$

Because of this cosmological redshift, when the photons emitted during the time dt_e will cross the sphere of radius r they will be carrying not the energy $d\mathscr{E}_e$ but only

$$d\mathscr{E}_r = (1+z)^{-1} d\mathscr{E}_e\,. \tag{22.5}$$

(b) Photons emitted during the time interval dt_e at the source are received during the time $dt_r \neq dt_e$. At the observer (see equation (16.13)):

$$dt_r = dt_e \left(\frac{R(t_o)}{R(t_e)}\right) = dt_e(1+z). \tag{22.6}$$

Combining the last two results, we obtain

$$\boxed{L = \frac{d\mathscr{E}_e}{dt_e} = (1+z)^2 \frac{d\mathscr{E}_r}{dt_r} = (1+z)^2 SA} \tag{22.7}$$

or

$$\boxed{S = \frac{(1+z)^{-2}}{A} L\,.} \tag{22.8}$$

(c) The area A is not given by the Euclidean formula.

For the metric forms (21.2)

$$dA = \begin{cases} R^2 \sin^2(r/R)\sin(\theta)d\theta d\phi\,, & k=1\,,\\ r^2\sin(\theta)d\theta d\phi\,, & k=0\,,\\ R^2\sinh^2(r/R)\sin(\theta)d\theta d\phi\,, & k=-1\,, \end{cases} \tag{22.9}$$

Hence,

$$A = \begin{cases} 4\pi R^2\sin^2(r/R)\,, & k=1\,,\\ 4\pi r^2\,, & k=0\,,\\ 4\pi R^2\sinh^2(r/R)\,, & k=-1\,, \end{cases} \tag{22.10}$$

As we have already seen in Sec.19 , cosmological models provide $r(z)$ and hence $A(z)$. Thus, for each cosmological model Eq.(22.8) predicts $S(z)$ for sources with the same luminosity. This opens the possibility to test these models observationally.

The example of flat Universe. As discussed in the previous section, the observations of CMB fluctuations strongly suggests $k=0$, and hence this model deserves special attention here. Fortunately, the involved calculations are particularly simple in this case. In the Friedmann's model of flat Universe,

$$r(z) = \frac{2c}{H_0}\left(1-(1+z)^{-1/2}\right),$$

(see equation (19.9)) and hence

$$S(z) = L\frac{H_0^2}{16\pi c^2}\left[(1+z)\left(1-(1+z)^{-1/2}\right)\right]^{-2}. \tag{22.11}$$

This is known as the Hubble Relation.

Thus, provided there are good standard candles in the Universe, they should exhibit S-z dependence in agreement with Eq.22.11. Astrophysical studies suggest that supernovae of a particular type (type Ia) are excellent standard candles producing the same luminosity at the peak brightness. Mass fully automatic searches for type-Ia supernovae in distant galaxies were started in the 1990's and have delivered large data sets which could be used to test cosmological models. *To the great surprise of cosmologists, the derived S-z dependence did not agree with the result (22.11).*

Section 23

Accelerating Universe

23.1 The deceleration parameter

The supernovae observations tell us that there is something wrong with the theoretical models of Friedmann's cosmology. To see what exactly is wrong, one may opt not to use the theoretical solutions for $R(t)$ but try instead to deduce it directly from the observations. For example, one can try to fit the observations with the truncated power series for $R(t)$ about the present time

$$R(t) = R_0 + \dot{R}_0(t-t_o) + \frac{1}{2}\ddot{R}_0(t-t_o)^2 .$$

To this end it is convenient to rearrange the expansion as

$$R(t) = R_0\left(1 + \frac{\dot{R}_0}{R_0}(t-t_o) + \frac{1}{2}\frac{\ddot{R}_0}{R_0}(t-t_o)^2\right). \tag{23.1}$$

The latter form is better as all the terms in the brackets are now dimensionless and hence do not depend on the choice of physical units. $\dot{R}_0/R_0 = H_0$ is the Hubble "constant" at $t = t_o$. Introducing the *deceleration parameter*

$$\boxed{q_0 = -\frac{\ddot{R}_0}{R_0 H_0^2},} \tag{23.2}$$

one can rewrite Eq.(23.1) as

$$\boxed{R(t) = R_0\left[1 + H_0(t-t_o) - \frac{q_0}{2}\left(H_0(t-t_o)\right)^2\right].} \tag{23.3}$$

The fact that $\ddot{R} < 0$ in all Friedmann's models is the historical reason behind the minus sign in the definition of q_0 and its name. This truncated expansion is reasonably accurate when $H_0(t-t_o) \ll 1$ (the error scales like $(H_0(t-t_o))^3$).

Using eq.(23.3) instead of the theoretical prediction in eq.(19.4), one obtains $r(z, H_0, q_0)$ which replaces $r(z)$ based on the theoretical models. In this approach, the parameters H_0 and q_0 are to be found from fitting the observations. This has been done and it turns out that the supernovae observations can be fitted very well. However, the required deceleration parameter must be negative,

$$q_0 \approx -0.6 .$$

Contrary to the predictions of all Friedmann's models, the expansion of our Universe is apparently speeding up!

23.2 Return of the Cosmological constant?

In order to see one possible solution to this crisis, let us return to the acceleration equation with the Cosmological constant retained in it,

$$\boxed{\left(\frac{\ddot{R}}{R}\right) = -\frac{4\pi G}{3}\left(\rho + 3\frac{P}{c^2}\right) + \frac{\Lambda}{3},} \tag{23.4}$$

(see Section 13). Since both for matter and radiation $\rho, P \geq 0$, in models with $\Lambda = 0$ we have $\ddot{R} < 0$. The Universe is decelerating. To have an accelerating Universe we must have $\Lambda > 0$. Thus one possible solution is to return to the modified gravity of the Einstein equation with the Cosmological constant.

23.3 Revised critical density

With $\Lambda \neq 0$ the Friedmann equation reads as

$$\left(\frac{\dot{R}}{R}\right)^2 + \frac{kc^2}{R^2} = \frac{8\pi}{3}G\rho + \frac{\Lambda}{3}, \tag{23.5}$$

or, since $\dot{R}/R = H$, as

$$H^2 + \frac{kc^2}{R^2} = \frac{8\pi}{3}G\rho + \frac{\Lambda}{3}.$$

Hence

$$\begin{aligned} 1 + \frac{kc^2}{R^2H^2} &= \frac{8\pi G\rho}{3H^2} + \frac{\Lambda}{3H^2} = \\ &= \frac{\rho}{\rho_c} + \frac{\Lambda}{3H^2}, \end{aligned} \tag{23.6}$$

where ρ_c is the "old" critical density (see §18). Thus $\rho/\rho_c = 1$ does not imply $k = 0$ (flat Universe) any more and we have to modify the definition of the critical parameter Ω if we want to retain the property

$$\boxed{\Omega = \begin{cases} > 1, & \text{for the closed Universe } (k > 0), \\ 1, & \text{for the flat Universe } (k = 0), \\ < 1, & \text{for the open Universe } (k < 0). \end{cases}} \tag{23.7}$$

One obvious solution is to define it via

$$\boxed{\Omega = \Omega_m + \Omega_r + \Omega_\Lambda,} \tag{23.8}$$

where

$$\boxed{\Omega_m = \frac{\rho_m}{\rho_c}, \quad \Omega_r = \frac{\rho_r}{\rho_c} \quad \text{and} \quad \Omega_\Lambda = \frac{\Lambda}{3H^2}.} \tag{23.9}$$

Here we have splitted the contributions of the cold matter, ρ_m, and the hot matter (or radiation), ρ_r, into the total mass-energy density $\rho = \rho_m + \rho_r$ and hence their contributions to the critical parameter.

(The value of Ω at present time t_0 is denoted as Ω_0.)

Section 24

Standard Cosmological Model

24.1 The key dynamic equation

The Standard Cosmological Model is the one based on the GR equations with the Cosmological constant. For example, the Friedmann equation reads

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3}G(\rho_r + \rho_m) - \frac{kc^2}{R^2} + \frac{\Lambda}{3}, \tag{24.1}$$

where we separated the total mass-energy density into the matter and radiation contributions. Solving the fluid equation for matter and radiation separately we obtain

$$\rho_m = \rho_{m,0}\left(\frac{R}{R_0}\right)^{-3} = \frac{\rho_{m,0}}{a^3} \tag{24.2}$$

and

$$\rho_r = \rho_{r,0}\left(\frac{R}{R_0}\right)^{-4} = \frac{\rho_{r,0}}{a^4}, \tag{24.3}$$

where we introduce the renormalised scaling factor

$$a = \frac{R}{R_0}$$

and $\rho_{m,0}$ and $\rho_{r,0}$ are the current (at $a = 1$) densities. Since we do not know the conditions at the beginning of the Universe but know its current state it makes sense to set the initial conditions at the present time.

Replacing R with a and substituting the results for ρ_r and ρ_m, we can write the Friedmann equation as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\left(\frac{\rho_{r,0}}{a^4} + \frac{\rho_{m,0}}{a^3}\right) - \frac{kc^2}{R_0^2a^2} + \frac{\Lambda}{3}. \tag{24.4}$$

Using the definitions of the critical parameters, this equation can also be written as

$$\dot{a}^2 = H_0^2\left(\frac{\Omega_{r,0}}{a^2} + \frac{\Omega_{m,0}}{a} + (1-\Omega_0) + \Omega_{\Lambda,0}a^2\right). \tag{24.5}$$

This equation determines the evolution of a in the Standard Cosmology. It has a number of parameters, the Hubble constant H_0 and the criticality parameters $\Omega_{r,0}$, $\Omega_{m,0}$, $\Omega_{\Lambda,0}$ and Ω_0. These parameters are to be found via fitting the model to the observational data. The initial condition is $a(t_0) = 1$, where t_0 is quite arbitrary at this stage and can actually be set to zero. Later, if out solution indicates some special time t_* which makes more sense to consider as the zero time, we simply replace t with $t - t_*$.

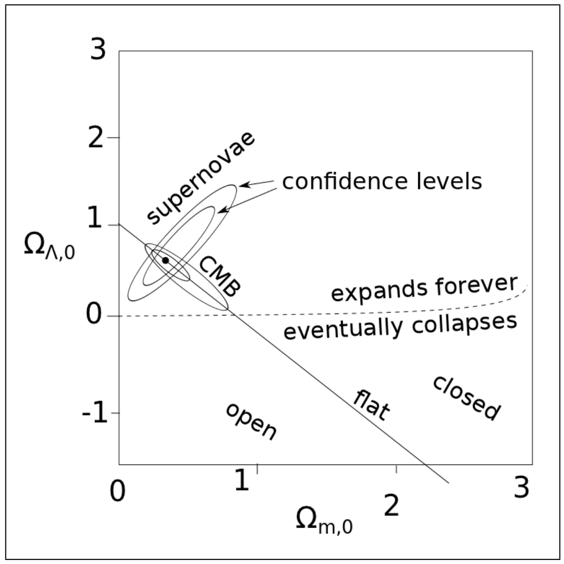


Figure 24.1: Confidence levels for the cosmological parameters from CMB and supernova observations.

24.2 Parameters of the Standard Model

When the Cosmological constant is retained, the acceleration equation reads

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right) + \frac{\Lambda}{3}. \tag{24.6}$$

The astronomical observations tell us that the radiation makes only a very small contribution to the energy density of the Universe at the present state of its evolution. Hence, $\rho_{m,0} \gg \rho_{r,0}, P_{r,0}$ and

$$\frac{\ddot{R}(t_0)}{R_0} \approx -\frac{4\pi G}{3}\rho_{m,0} + \frac{\Lambda}{3}.$$

Substituting this expression into the definition of the deceleration parameter we find

$$q_0 = \frac{\Omega_{m,0}}{2} - \Omega_{\Lambda,0}. \tag{24.7}$$

From the supernovae observations (the standard candle method) $q_0 \approx -0.6$ and from the observations of CMB fluctuations (the standard bar method) $\Omega_0 \approx 1.0$. Hence,

$$\frac{\Omega_{m,0}}{2} - \Omega_{\Lambda,0} \approx -0.6$$

and

$$\Omega_{m,0} + \Omega_{\Lambda,0} \approx 1.0\,,$$

where we ingnored $\Omega r, 0 \ll \Omega_{m,0}$. Solving these for $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ we find

$$\Omega_{m,0} \approx 0.27, \quad \Omega_{\Lambda,0} \approx 0.73. \tag{24.8}$$

Based on the astronomical observations,

$$\Omega_{r,0} \approx 8 \times 10^{-5}\,, \tag{24.9}$$

with the dominant contribution coming from the CMB radiation. Because it so small, we did not take it into account in the above calculations of $\Omega_{r,0}$ and $\Omega_{\Lambda,0}$. Finally,

$$H_0 \approx 72\,\mathrm{km/s\,Mpc^{-1}}\,.$$

24.3 Reconstructing the past and predicting the future

Equation (24.5) is rather complicated and in general has to be integrated numerically. However, we do not need to resort to numerical integration to reach the following important conclusions.

- The curvature term, $(1-\Omega_0) = -kc^2/R_0^2H_0^2$, does not vary with a. It it already rather small at present, $|1-\Omega_0| \ll 1$, and hence it is totally insignificant in the past and in the future, where other terms become much larger. *Thus, the Universe is effectively flat – it does not matter much if* $k = 0,\ -1$ *or* $+1$. Its evolution is almost the same for all these three choices. This is in great contrast to the Friedmann solutions which are qualitatively different for the three choices of k.
- The matter and radiation terms, $\Omega_{r,0}/a^2$ and $\Omega_{m,0}/a$ respectively, grow as $a \to 0$. Thus, they dominate in the past (when $a \ll 1$).
- Since the radiation term grows faster than the matter one as $a \to 0$, in the past there should be a transition from the radiation-dominated phase to the matter-dominated phase. This occurs when
$$\frac{\Omega_{m,0}}{a} = \frac{\Omega_{r,0}}{a^2}\,,$$
which yields
$$a_{eq} = \frac{\Omega_{r,0}}{\Omega_{m,0}} \approx 3\times 10^{-4}\,.$$
This is a much smaller and hence much younger Universe.

  Based on the densities corresponding to a_{eq}, one can show that the Universe must be opaque at this time. Photons cannot propagate freely but get absorbed and emitted again at very high rate. Matter and radiation are tightly coupled. They decouple only at $a_{dc} \approx 10^{-3}$. For $a < a_{eq}$ particles of matter become relativistically hot and hence described by the same equation of state as the radiation, $P = \rho c^2/3$. We may no longer differentiate between them in the Friedmann equations.
- During the radiation-dominated phase, Eq.(24.5) reduces to
$$\dot{a} = H_0\frac{\sqrt{\Omega_{r,0}}}{a}\,. \tag{24.10}$$
Its solution is
$$a^2 = 2H_0\sqrt{\Omega_{r,0}}t + \text{const}\,.$$
One can see that it implies the existence of such t_* that $a(t_*) = 0$, and hence Big Bang is still a feature of Standard Cosmology. Shifting the origin of time via $t \to t - t_*$, so that the Universe is created at $t = 0$, we obtain
$$a \propto t^{1/2}\,. \tag{24.11}$$
- As $\Omega_{\Lambda,0}a^2$ is the only term on the right side of eq.(24.5) which grows with a, it will dominate the future evolution of the Universe. The transition from matter dominated Universe to Λ-dominated Universe occurs when
$$\frac{\Omega_{m,0}}{a} = \Omega_{\Lambda,0}a^2\,,$$
which yields
$$a_\Lambda = \left(\frac{\Omega_{m,0}}{\Omega_{\Lambda,0}}\right)^{1/3} \approx 0.72,.$$
This is only in the relatively recent past, in the Cosmological sense, and hence at present the Universe is in the transition to the epoch of Λ-domination.
- Ignoring all other terms, we obtain
$$\dot{a} = \alpha a\,,$$
where $\alpha = H_0\Omega_{\Lambda,0}^{1/2}$. Its solution is
$$a = A\exp(\alpha t)\,. \tag{24.12}$$
Hence an exponential expansion in the future (when $a \gg 1$)!

Section 25

Dark matter and dark energy

25.1 Dark matter

According to the astronomical observations, all the visible matter in the Universe (stars, planets, gas, dust etc.) contribute only a small fraction to $\Omega_{m,0}$. Namely

$$\Omega_{vis,0} = \frac{\rho_{vis,0}}{\rho_c} \approx 0.05.$$

The rest of the matter must be in invisible form,

$$\Omega_{inv,0} = \Omega_{m,0} - \Omega_{vis,0} \approx 0.22.$$

Where is this invisible or *dark matter*? How can it be detected?

1. Via its gravitational interaction with visible matter.

2. Via bending of light (gravitational lensing, warping of space).

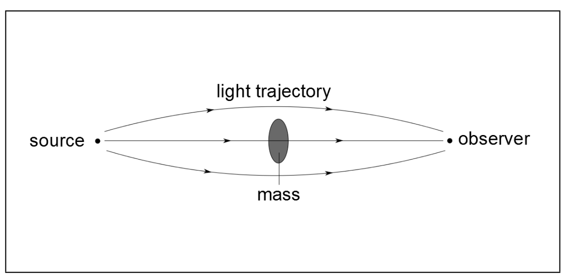
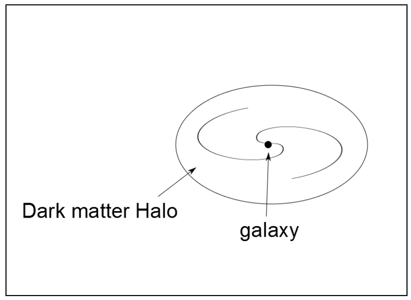


Figure 25.1: The gravitational lens. The higher the mass, the stronger the bending.

The astronomical observations tell us that some dark matter is present inside galaxies in the form of an extended halo. But most of it is in the form of haloes in clusters of galaxies, with

$$\Omega_{halo,0} \approx 0.2 - 0.3.$$

This is very close to $\Omega_{m,0}$, which is found in cosmological models! What is the nature of dark matter? Most likely it is made of exotic particles, which do not interact electromagnetically and hence do not produce light. Such particles are predicted in modern physical theories of matter. Some, like neutrinos, have been detected experimentally (There are too few neutrinos in the Universe to explain the dark matter and other types of particles are needed.). “Hunting” for dark matter particles in laboratory is a “cool” current activity.



25.2 Dark energy

Einstein's Λ-term is a part of the modified gravity law (13.4). Thus, modified gravity (the Cosmological Constant) is one possible explanation of the observations. Alternatively, the gravity law may still keep its original form without the Λ-term, but in addition to matter and radiation the Universe hosts something else which has more or less the same effect on its evolution as the Cosmological Constant.

Let us inspect the Friedmann equations again and check if their Λ terms can be reshaped as contributions due to some "fluid".

The Friedmann equations with Λ terms:

$$\left(\frac{\dot{R}}{R}\right)^2 + \frac{kc^2}{R^2} = \frac{8\pi}{3}G\rho + \frac{\Lambda}{3}, \tag{25.1}$$

$$\left(\frac{\ddot{R}}{R}\right) = -\frac{4\pi G}{3}\left(\rho + 3\frac{P}{c^2}\right) + \frac{\Lambda}{3}. \tag{25.2}$$

Introduce ρ_Λ and P_Λ such that

$$\frac{\Lambda}{3} = \frac{8\pi G}{3}\rho_\Lambda\,, \tag{25.3}$$

and

$$\frac{\Lambda}{3} = -\frac{4\pi G}{3}\left(\rho_\Lambda + 3\frac{P_\Lambda}{c^2}\right). \tag{25.4}$$

Using these we can write (25.1) and (25.2) as

$$\left(\frac{\dot{R}}{R}\right)^2 + \frac{kc^2}{R^2} = \frac{8\pi G}{3}\left(\rho + \rho_\Lambda\right), \tag{25.5}$$

$$\left(\frac{\ddot{R}}{R}\right) = -\frac{4\pi G}{3}\left(\rho + \rho_\Lambda + \frac{3}{c^2}\left(P + P_\Lambda\right)\right). \tag{25.6}$$

These are the same as the Friedmann equations with $\Lambda = 0$ but with two fluids, one with parameters ρ, P and the other with ρ_Λ, P_Λ. The second fluid is very peculiar. From (25.3) and (25.4) we find

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}, \tag{25.7}$$

$$P_\Lambda = -\rho_\Lambda c^2. \tag{25.8}$$

This can neither be matter no radiation simply because $P_\Lambda < 0$. The pressure is negative!

The nature of this "fluid" is still unknown. It is speculated that this could be an additional fundamental force field (in addition to the electromagnetic, strong (nuclear), weak and gravitational force), which is normally even weaker than gravity. This hypothetical field is romantically called *quintessence* or the fifth element. (The four fundamental elements of existence in classical philosophy are the earth, air, water and fire.)

If quintessence is the correct explanation then

1. Gravity is described by the original Einstein's equation of gravity which does not have the Λ-term;
2. Quintessence "mimics" the Λ-terms in cosmological models;
3. This does not have to be a perfect mimicry, e.g. the mass-energy density of the quintessence may be a function of time and the equation of state does not have to be exactly $P = -\rho c^2$.

Section 26

The phase of Inflation

26.1 The problems of Standard Cosmology

There are a number of puzzling issues in the Standard Cosmological model. Here we discuss only two of them, the flatness problem and the horizon problem. Other issues are a bit more complex and require a very good background in particle physics. They are all solved if initially the expansion of the Universe proceeded in different way to that of Standard Cosmological Model.

The flatness problem

According to the cosmological observations, the critical parameter Ω_0 is very close to unity. Why is this? Unless the Universe is exactly flat ($k = 0$), we should have $\Omega_0 \neq 1$. *Could it be that Ω naturally evolves towards unity as the Universe expands even in the models with $k = \pm 1$?* If yes then we have a simple explanation. If no then the observed $\Omega_0 \approx 1$ is a mystery.

From (23.6)

$$|1-\Omega| = \frac{|k|c^2}{H^2R^2} = \frac{c^2}{\dot{R}^2}. \tag{26.1}$$

for $k = \pm 1$. Thus the evolution of Ω is determined by the evolution of $\dot{R}$. The problem is that in the Standard Model $\dot{R} \propto t^{-1/2}$ during the initial radiation-dominated phase and then as $\dot{R} \propto t^{-1/3}$ during the matter-dominated phase, which continued almost until the current epoch. Thus $|1-\Omega|$ has been increasing and hence Ω moving away from unity. That is the opposite to what is needed.

The horizon problem

The Standard Cosmological model assumes that the Universe is uniform, which is in agreement with the astronomical observations. But why is it uniform?

Well, the uniformity could imply an efficient mechanism of erasing inhomogeneities in the early Universe. However, no matter what actual physical processes might be responsible for this, *they would involve communication between different parts of the now observable Universe*. The highest communication speed is the speed of light. *Hence in the past, the part of the Universe which corresponds to currently observed Universe had to be smaller than the contemporary cosmological horizon.* That part of the Universe has the same comoving radius χ as the "currently visible Universe". Its metric radius r increases only due to the expansion of the Universe, and hence $r \propto R$. If galaxies were undestractable, their total number inside the sphere would not change in time. A good analogy is a circle painted on the surface of an expanding air balloon. In contrast, the cosmological horizon is a spherical light front. Its comoving radius increases with time and the number of galaxies inside of it grows.

At present, we can observe the Universe up to the CMB background, which appears to be very uniform. Thus, it is the comoving radius of the CMB, which defines the size of the currently visible Universe. The remarkable isotropy of the CMB emission in the sky tells us that the whole of the visible Universe was uniform already at the time of decoupling, t_d (the time when the CMB photons, which are reaching the Earth now, were produced). Hence, already at $t = t_d$ the comoving radius of the Cosmological horizon had to exceed the comoving radius of the currently visible Universe. Obviously, the same conclusion applies to the metric radii.

The metric radius of the Cosmological horizon is given by the equation

$$r_h(t) = cR(t)\int_0^t \frac{dt}{R(t)}\,, \tag{26.2}$$

(see Eq.20.4). Since in the Standard Cosmological Model the Universe is radiation-dominated for all $t < t_d$, we can substitute $R \propto t^{1/2}$ (see Eq.24.11) and hence obtain

$$r_h(t) = 2ct\,. \tag{26.3}$$

Now let us consider a source of the metric size $r_h(t_d) = 2ct_d$, located at the edge of the currently visible Universe at $t = t_d$, and calculate its angular size on the sky, θ_h. Since for $t_d < t < t_0$, the Universe is matter-dominated and flat (the curvature term is small), we can use (21.8) which yields

$$\theta_h = r_h(t_d)\frac{H_0}{2c}\frac{1+z}{\left(1-(1+z)^{-\frac{1}{2}}\right)} \approx t_d H_0 z\,, \tag{26.4}$$

where we used the fact that the source redshift is the same as that of the CMB radiation and hence $z \gg 1$. t_d can also be estimated using the expansion law of the matter-dominated phase. From Equations (19.7) and (16.21), it follows that

$$t_d = t_0(1+z)^{-3/2} \approx t_0 z^{-3/2} = \frac{2}{3H_0}z^{-3/2}\,. \tag{26.5}$$

Hence

$$\theta_h = \frac{2}{3\sqrt{z}}\,. \tag{26.6}$$

For CMB background $z \approx 10^3$ and (26.6) yields $\theta_h \approx 0.02\text{rad} \approx 1^o$. Hence, at t_d the radius of the cosmological horizon is aprroximately fifty times smaller than the radius of the currently visible Universe at the same time. Obviously, it was even smaller for $t < t_d$.

Thus, in the Standart Cosmological Model there is no time to establish causal-connectivity across the visible Universe and no physical process can erase its initial inhomogeneities. Mysteriously, the Universe had to be uniform from the start.

26.2 Inflation

Since for sufficiently small $R(a)$ we enter the conditions not accessible to the modern physics, weird things may occur there. Suppose that prior to the radiation-dominated phase there was another phase, which we will call Inflation, when the Universe was expanding exponentially

$$R = R_* e^{\alpha t} \text{for} 0 < t < t_i\,, \tag{26.7}$$

where $t_i \ll t_d$ In this case, all the problems of Standard Cosmology can be resolved[1].

Solution to the flatness problem

During Inflation,

$$\dot{R} \propto e^{\alpha t}$$

[1]Note that R does not vanish at $t = 0$, which opens the possibility of yet another phase preceding Inflation.

and hence

$$|1-\Omega| = \frac{c^2}{\dot{R}^2} \propto e^{-2\alpha t} . \tag{26.8}$$

Thus even if initially Ω is very large or very small, it becomes close to unity for sufficiently large t, when $\alpha t \gg 1$.

Solution to the horizon problem

During Inflation, the cosmological horizon grows as

$$r_h(t) = cR(t) \int_0^t \frac{dt}{R(t)} = \frac{c}{\alpha}(e^{\alpha t} - 1) . \tag{26.9}$$

Provided $\alpha t \gg 1$ by the end of Inflation ($t = t_i$) the horizon radius reaches

$$r_h(t_i) = \frac{c}{\alpha} e^{\alpha t_i} . \tag{26.10}$$

During the subsequent radiation-dominated phase, the horizon radius continues to increase, but this additional increase is still limited by $2ct_d$. As we have just seen, this is about a fifty times below the contemporary size of the currently visible Universe. Thus, r_h must exceed the size of the currently visible Universe already by $t = t_i$

$$r_h(t_i) > r_u(t_i) ,$$

where $r_u(t)$ is the size of the visible Universe. Since during the radiation-dominated phase $R \propto t^{1/2}$, we have $r_u(t) \propto t^{1/2}$ and hence

$$r_u(t_i) = \left(\frac{t_i}{t_d}\right)^{1/2} r_u(t_d) .$$

Moreover, $r_u(t_d) \approx 100ct_d$ (see Eq.26.6). Collecting all these results, we find the condition

$$r_h(t_i) > 10^2 ct_d \left(\frac{t_i}{t_d}\right)^{1/2} . \tag{26.11}$$

Let us show that this condition can be satisfied. Using eq.(26.10), it can be written as

$$e^x > 10^2 x \left(\frac{t_d}{t_i}\right)^{1/2} , \tag{26.12}$$

where $x = \alpha t_i$, or

$$x > \ln 100 + \ln x + \frac{1}{2} \ln\left(\frac{t_d}{t_i}\right) . \tag{26.13}$$

Since $\ln x$ is a slowly growing function, this condition is always satisfied by sufficiently large x. The typical duration of the inflation phase discussed in modern Cosmology is very small, $t_i = 10^{-34}$s being a representative value. Using (26.5), we estimate $t_d = 10^{13}$s and hence $t_d/t_i = 10^{47}$. Substituting this into Eq.26.13, we obtain

$$x > 59 + \ln x$$

or $x > 63$. Thus, during Inflation the Universe increases its size by more then $e^{63} \simeq 10^{27}$ times!

Inflaton

What would make the Universe to expand exponentially? We have already seen in Section 24 that an exponential expansion is expected in the Standard Cosmology in the near future of the Universe, where the Cosmological-constant term will dominate all other terms of the Friedmann equation (see Eq.24.12). Just like this term is associated with the Quintessence field, the, theorists associate Inflation with another field which may have existed in the very early Universe and call it "Inflaton". Its equation of state is very close to $P_i = -\rho_i c^2$, leading to only very slowly changing mass-energy density during Inflation. Inflation ends when Inflaton particles become unstable and decay into other particles, which satisfy the EOS of radiation $P = \rho c^2/3$.

The theory of Inflation successfully resolves many other problems of the Friedmann Cosmology, which we can not discuss within the limited scope of this module. In fact it remains a very hot topical issue in modern Cosmology and particle physics.