The Geometry of the Universe

Newton versus Einstein

gravitational force acting between the two objects (assuming they are both spherical) is

$$F = -\frac{GM_g m_g}{r^2}. \tag{3.1}$$

What is the acceleration that results from this gravitational force? Newton had something to say about that as well. Every object in the universe, said Newton, has a property that we may call the "inertial mass." Let the inertial mass of an object be m_i. Newton's second law of motion says that force and acceleration are related by the equation

$$F = m_i a. \tag{3.2}$$

The property of an object that determines how strongly it is pulled on by the force of gravity is equal to the property that determines its resistance to acceleration by *any* force, not just the force of gravity. The equality of gravitational mass and inertial mass is called the *equivalence principle*.

$$m_g = m_i.$$

(Figure 3.2), the equivalence principle permits two possible interpretations, with no way of distinguishing between them:

> (1) The box is static, or moving with a constant velocity, and the bear is being accelerated downward by a gravitational force.
>
> (2) The bear is static, or moving at a constant velocity, and the box is being accelerated upward by a non-gravitational force.

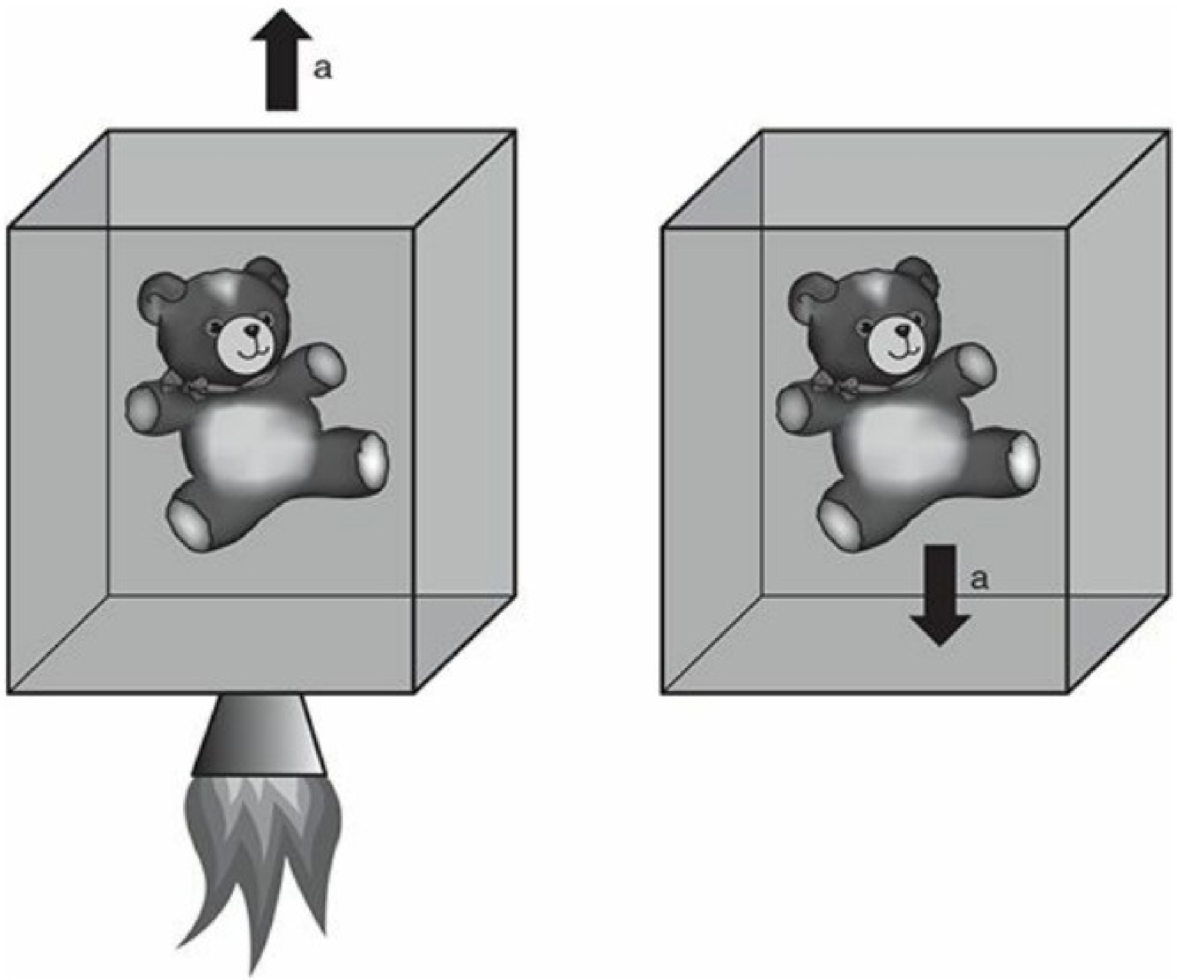


Figure 3.2 Equivalence principle (teddy bear version). The behavior of a bear in an accelerated box (left) is identical to that of a bear being accelerated by gravity (right).

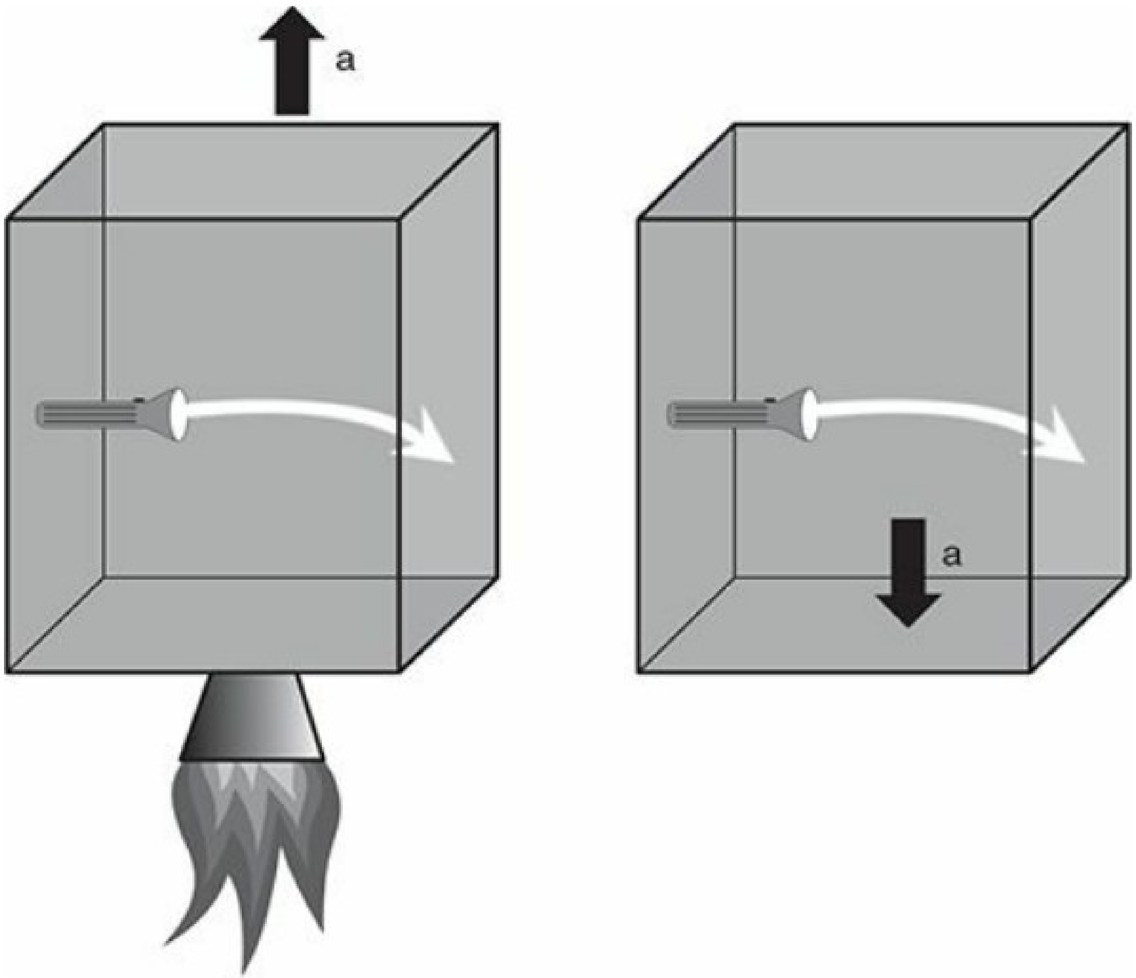
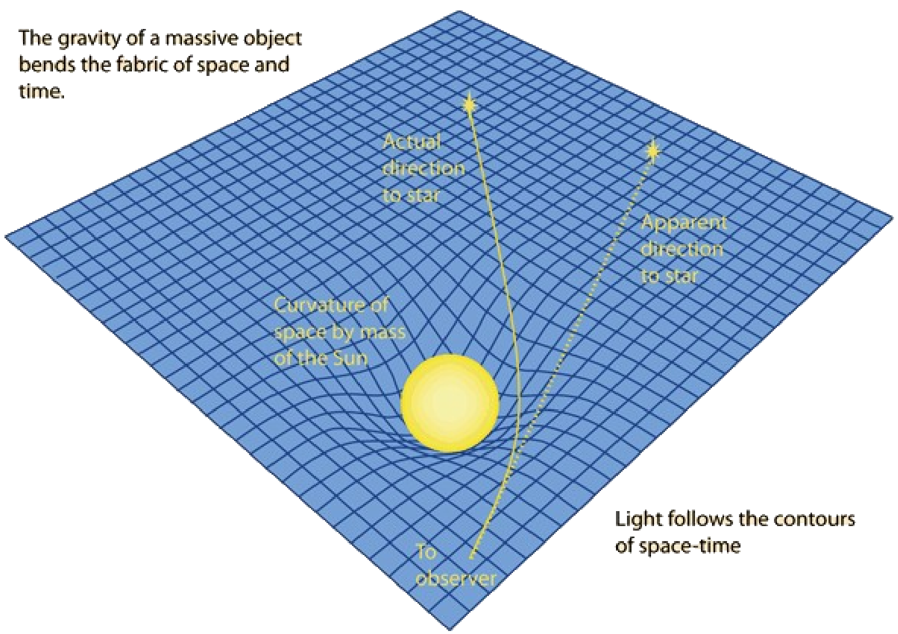


Figure 3.3 Equivalence principle (photon version). The path followed by a light beam in an accelerated box (left) is identical to the path followed by a light beam being accelerated by gravity (right). [The deflection shown is greatly exaggerated for the sake of visualization. The actual deflection will be ~ 2×10^{-14} m if the box is 2 meters across.]

The presence of mass, in Einstein's view, causes space to be curved. In fact, in the fully developed theory of general relativity, mass and energy (which Newton thought of as two separate entities) are interchangeable, via the famous equation $E = mc^2$. Moreover, space and time (which Newton thought of as two separate entities) form a four-dimensional spacetime. A more accurate summary of Einstein's viewpoint, therefore, is that the presence of mass-energy causes spacetime to be curved. We now have a third way of thinking about the motion of the teddy bear in the box:

> (3) No forces are acting on the bear; it is simply following a *geodesic* in curved spacetime.

If you take two points in an N-dimensional space or spacetime, a geodesic is defined as the locally shortest path between them.



The Way of Newton:

Mass tells gravity how to exert a force ($F = -GMm/r^2$),

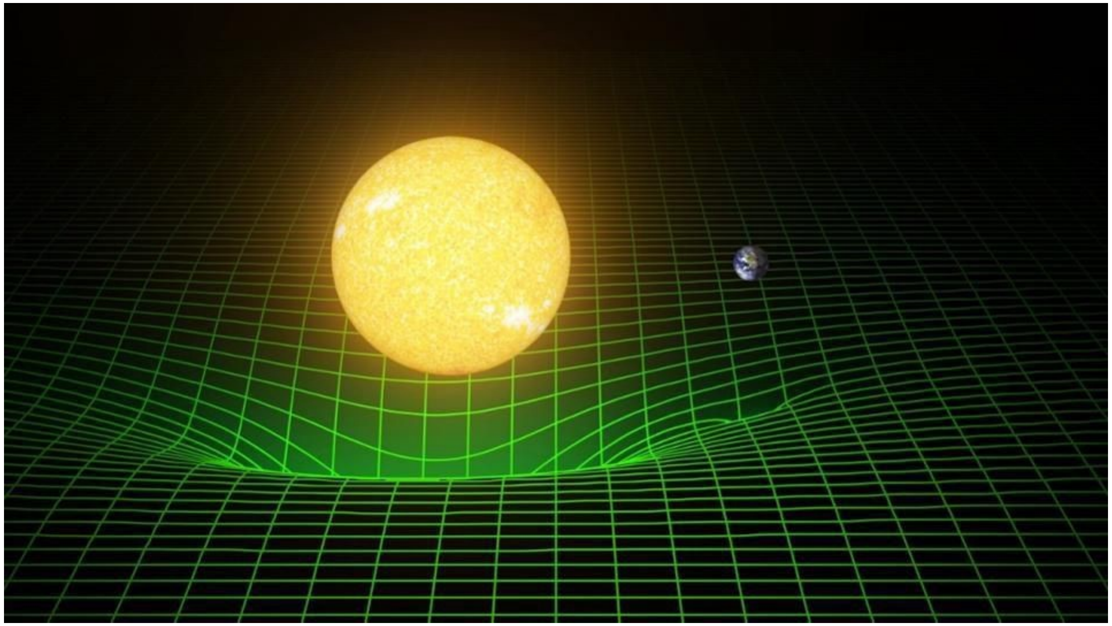
Force tells mass how to accelerate ($F = ma$).

The (General) Way of Einstein:

Mass-energy tells spacetime how to curve,

Curved spacetime tells mass-energy how to move.[2]

Einstein's description of gravity gives a natural explanation for the equivalence principle. In the Newtonian description of gravity, the equality of the gravitational mass and the inertial mass is a remarkable coincidence. However, in Einstein's theory of general relativity, curvature is a property of spacetime itself. It then follows automatically that the gravitational acceleration of an object should be independent of mass and composition – it's just following a geodesic, which is dictated by the geometry of spacetime.



The Geometry of the Universe

Describing Curvature

In developing his theory of general relativity, Einstein faced multiple challenges. Ultimately, he wanted a mathematical formula (called a *field equation*) that relates the curvature of spacetime to its mass-energy density, similar to the way in which Poisson's equation relates the gravitational potential of space to its mass density. En route to this ultimate goal, however, Einstein needed a way of mathematically describing curvature. Since picturing the curvature of a four-dimensional spacetime is difficult, let's start by considering ways of describing the curvature of two-dimensional spaces, and then extend what we have learned to higher dimensions.

$$\alpha + \beta + \gamma = \pi,$$

distance $d\ell$ between points (x, y) and $(x + dx, y + dy)$ is given by

$$d\ell^2 = dx^2 + dy^2.$$

$$x = r \cos \theta, y = r \sin \theta.$$

$$d\ell^2 = dr^2 + r^2 d\theta^2.$$

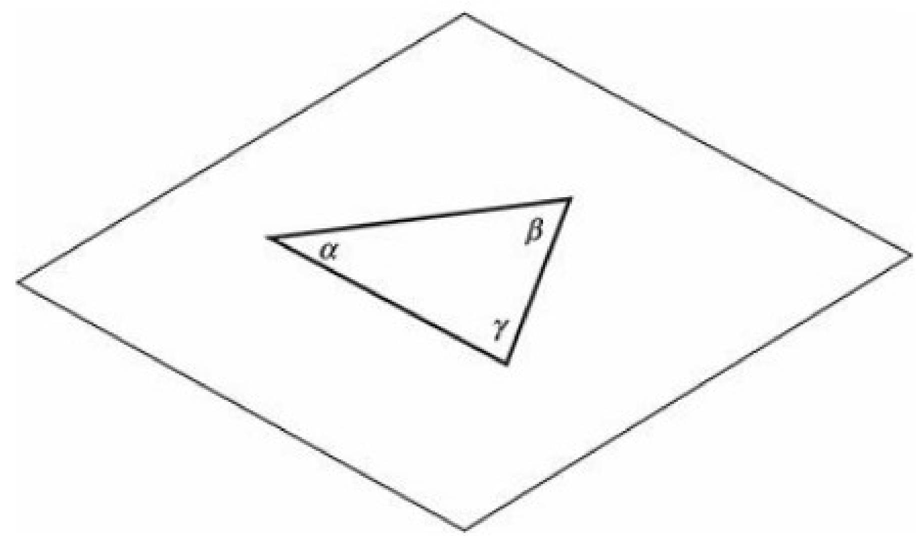
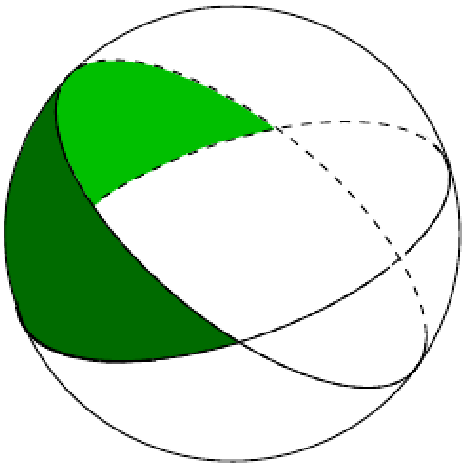


Figure 3.4 A Euclidean, or flat, two-dimensional space.

Diangles

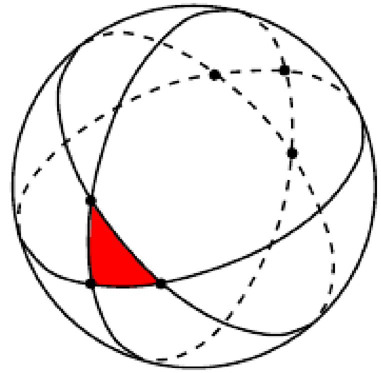
Any two distinct great circles intersect in two points which are negatives of each other.



The angle between two great circles at an intersection point is the angle between their respective planes.

A region bounded by two great circles is called a diangle.

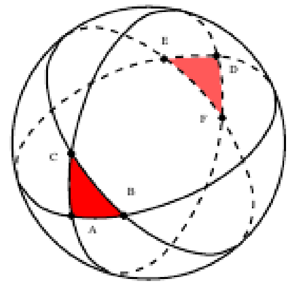
The angle at both the vertices are equal. Both diangles bounded by two great circles are congruent to each other.



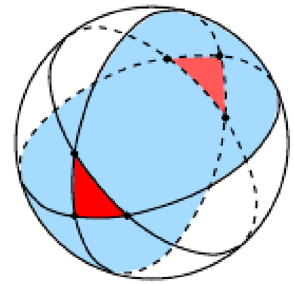
Spherical Triangle

Girard's Theorem

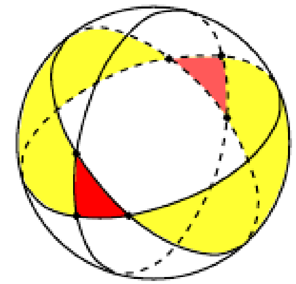
The area of a spherical triangle with angles α, β and γ is $\alpha+\beta+\gamma-\pi$.



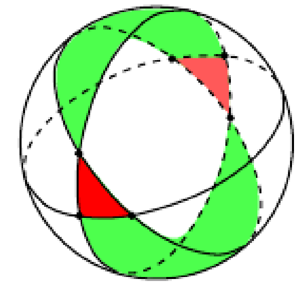
$\triangle ABC$



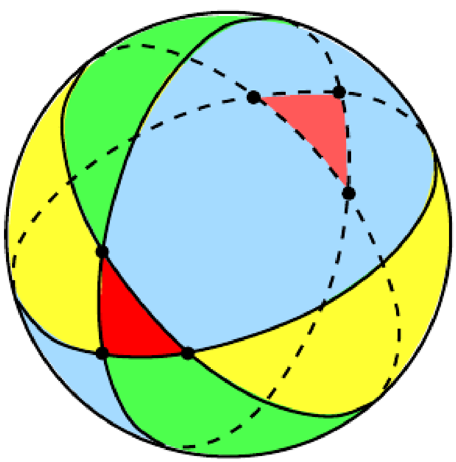
R_{AD}



R_{BE}



R_{CF}



Let R_{AD}, R_{BE} and R_{CF} denote pairs of diangles as shown. Then $\triangle ABC$ and $\triangle DEF$ each gets counted in every diangle.

$R_{AD} \cup R_{BE} \cup R_{CF} = S^2$, $\mathrm{Area}(\triangle ABC) = \mathrm{Area}(\triangle DEF) = X$.

$$\begin{aligned} \mathrm{Area}(S^2) &= \mathrm{Area}(R_{AD}) + \mathrm{Area}(R_{BE}) + \mathrm{Area}(R_{CF}) - 4X \\ 4\pi &= 4\alpha + 4\beta + 4\gamma - 4X \\ X &= \alpha + \beta + \gamma - \pi \end{aligned}$$

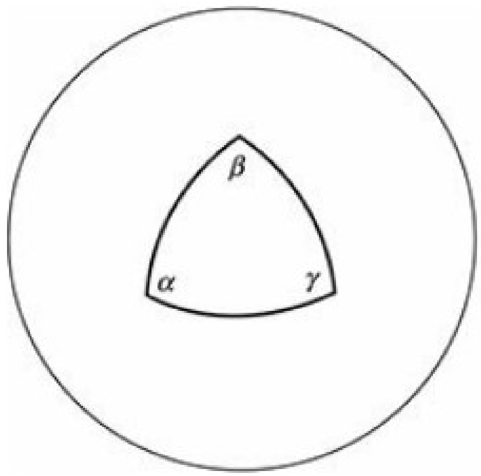


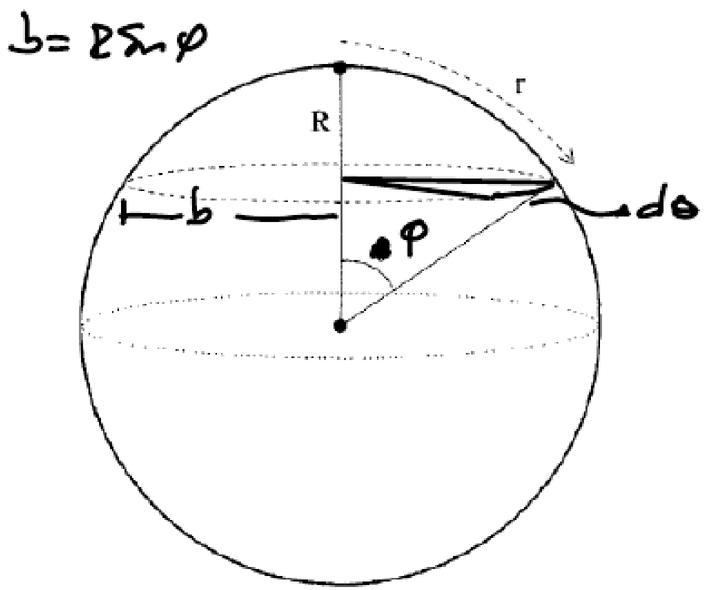
Figure 3.5 A positively curved two-dimensional space.

$$\alpha + \beta + \gamma = \pi + A/R^2, \tag{3.25}$$

where A is the area of the triangle, and R is the radius of the sphere. All spaces in which $\alpha + \beta + \gamma > \pi$ are called positively curved spaces. The surface of a sphere is a special variety of positively curved space; it has curvature that is both homogeneous and isotropic. That is, no matter where you draw a triangle on the surface of a sphere, or how you orient it, it must always satisfy Equation 3.25, with the radius R being the same everywhere and in all directions. For brevity, we can describe a space where the curvature is homogeneous and isotropic as having "uniform curvature." Thus, the surface of a sphere can be described as a two-dimensional space with uniform positive curvature.

On the surface of a sphere, we can set up a polar coordinate system by picking a pair of antipodal points to be the “north pole” and “south pole” and by picking a geodesic from the north to the south pole to be the “prime meridian.” If r is the distance from the north pole, and θ is the azimuthal angle measured relative to the prime meridian, then the distance $d\ell$ between a point (r, θ) and another nearby point $(r + dr, \theta + d\theta)$ is given by the relation

$$d\ell^2 = dr^2 + R^2 \sin^2(r/R) d\theta^2. \tag{3.26}$$



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

Note that the surface of a sphere has a finite area, equal to $4\pi R^2$, and a maximum possible distance between points. (In a non-Euclidean space, the distance between two points is defined as the length of the geodesic connecting them.) The distance between antipodal points, at the maximum possible separation, is $\ell_{max} = \pi R$. By contrast, a plane has infinite area, and has no upper limit on the possible distance between points.[4]

In addition to flat spaces and positively curved spaces, there exist negatively curved spaces. An example of a negatively curved two-dimensional space is the hyperboloid, or saddle shape, shown in Figure 3.6. For illustrative purposes, it would be useful to show you a surface of uniform negative curvature, just as the surface of a sphere has uniform positive curvature. Unfortunately, the mathematician David Hilbert proved that a two-dimensional surface of uniform negative curvature cannot be constructed in a three-dimensional Euclidean space. The saddle shape illustrated in Figure 3.6 has uniform curvature only in the central region, near the "seat" of the saddle.

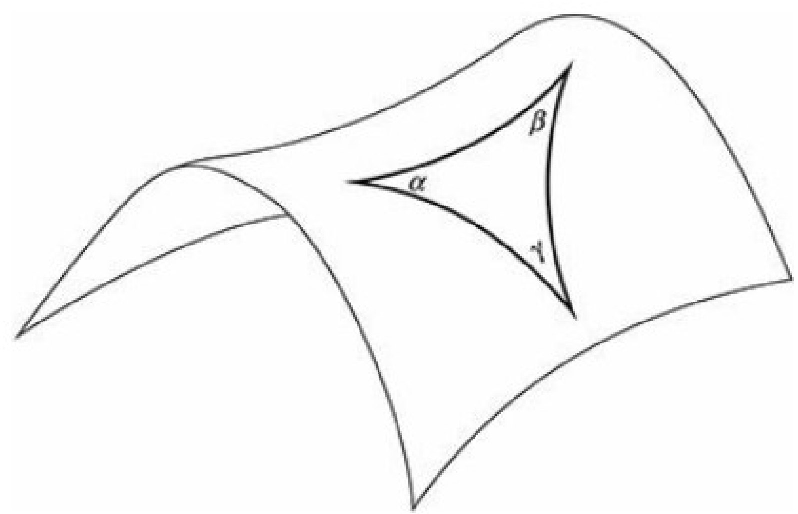


Figure 3.6 A negatively curved two-dimensional space.

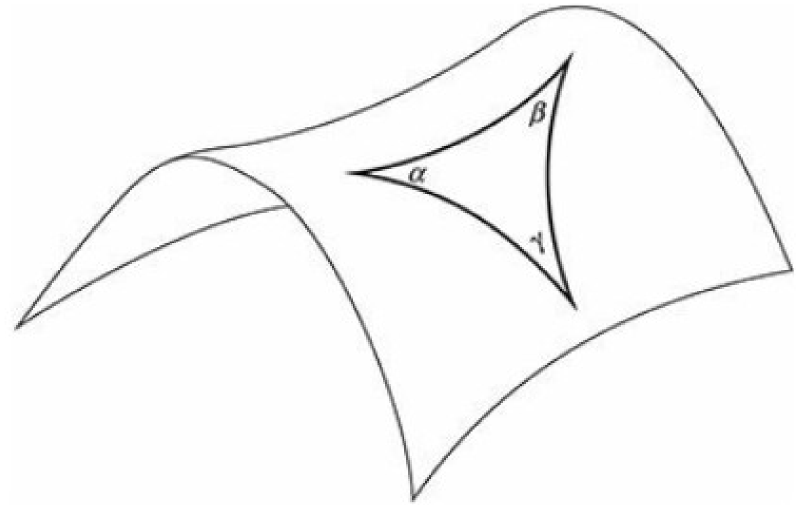


Figure 3.6 A negatively curved two-dimensional space.

$$\alpha + \beta + \gamma = \pi - A/R^2, \tag{3.27}$$

where A is the area of the triangle.

On a surface of uniform negative curvature, we can set up a polar coordinate system by choosing some point as the pole, and some geodesic leading away from the pole as the prime meridian. If r is the distance from the pole, and θ is the azimuthal angle measured relative to the prime meridian, then the distance $d\ell$ between a point (r, θ) and a nearby point $(r + dr, \theta + d\theta)$ is given by

$$d\ell^2 = dr^2 + R^2 \sinh^2(r/R) d\theta^2. \tag{3.28}$$

A surface of uniform negative curvature has infinite area, and has no upper limit on the possible distance between points.

$$d\ell^2 = dr^2 + r^2 d\theta^2.$$

$$d\ell^2 = dr^2 + R^2 \sin^2(r/R)[d\theta^2 + \sin^2\theta d\phi^2].$$

$$d\ell^2 = dr^2 + R^2 \sinh^2(r/R)[d\theta^2 + \sin^2\theta d\phi^2].$$

$$d\ell^2 = dr^2 + S_\kappa(r)^2 d\Omega^2,$$

$$d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$$

$$S_\kappa(r) = \begin{cases} R\sin(r/R) & (\kappa = +1) \\ r & (\kappa = 0) \\ R\sinh(r/R) & (\kappa = -1). \end{cases}$$

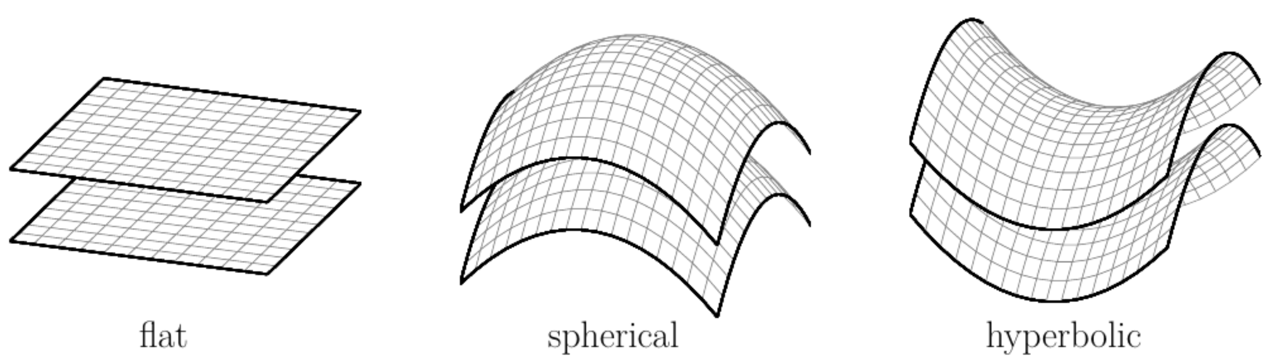
In the limit $r \ll R$, $S_\kappa \approx r$, regardless of the value of κ. When space is flat, or negatively curved, S_κ increases monotonically with r, with $S_\kappa \to \infty$ as $r \to \infty$. By contrast, when space is positively curved, S_κ increases to a maximum of $S_{max} = R$ at $r/R = \pi/2$, then decreases again to 0 at $r/R = \pi$, the antipodal point to the origin.

The coordinate system (r, θ, ϕ) is not the only possible system. For instance, if we switch the radial coordinate from r to $x \equiv S_\kappa(r)$, the metric for a homogeneous, isotropic, three-dimensional space can be written in the form

$$d\ell^2 = \frac{dx^2}{1 - \kappa x^2/R^2} + x^2 d\Omega^2.$$

Table 4.1 A summary of possible geometries

Curvature	Geometry	Angles of triangle	Circumference of circle	Type of universe
$k > 0$	Spherical	$> 180°$	$c < 2\pi r$	Closed
$k = 0$	Flat	$180°$	$c = 2\pi r$	Flat
$k < 0$	Hyperbolic	$< 180°$	$c > 2\pi r$	Open



The spacetime of the universe can be foliated into flat, spherical (positively-curved) or hyperbolic (negatively-curved) spatial hypersurfaces.

The Robertson–Walker Metric

$$ds^2 = -c^2dt^2 + dr^2 + r^2d\Omega^2. \tag{3.37}$$

The metric given in Equation 3.37 is called the *Minkowski metric*, and the spacetime that it describes is called Minkowski spacetime. Note, from comparison with Equation 3.33, that the spatial component of Minkowski spacetime is Euclidean, or flat.

The Minkowski metric of Equation 3.37 applies only within the context of special relativity. With no gravity present, Minkowski spacetime is flat and static. When gravity is added, however, the permissible spacetimes are more interesting. In the 1930s, the physicists Howard Robertson and Arthur Walker asked, "What form can the metric of spacetime assume if the universe is spatially homogeneous and isotropic at all time, and if distances are allowed to expand or contract as a function of time?" The metric they derived (independently of each other) is called the *Robertson–Walker metric*.[6] It can be written in the form

$$ds^2 = -c^2dt^2 + a(t)^2\left[dr^2 + S_\kappa(r)^2d\Omega^2\right],$$

$$ds^2 = dt^2 - a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\right)$$

The time variable t in the Robertson–Walker metric is the cosmological proper time, called the *cosmic time* for short, and is the time measured by an observer who sees the universe expanding uniformly around him or her. The spatial variables (r, θ, ϕ) are called the *comoving coordinates* of a point in space; if the expansion of the universe is perfectly homogeneous and isotropic, then the comoving coordinates of any point remain constant with time.

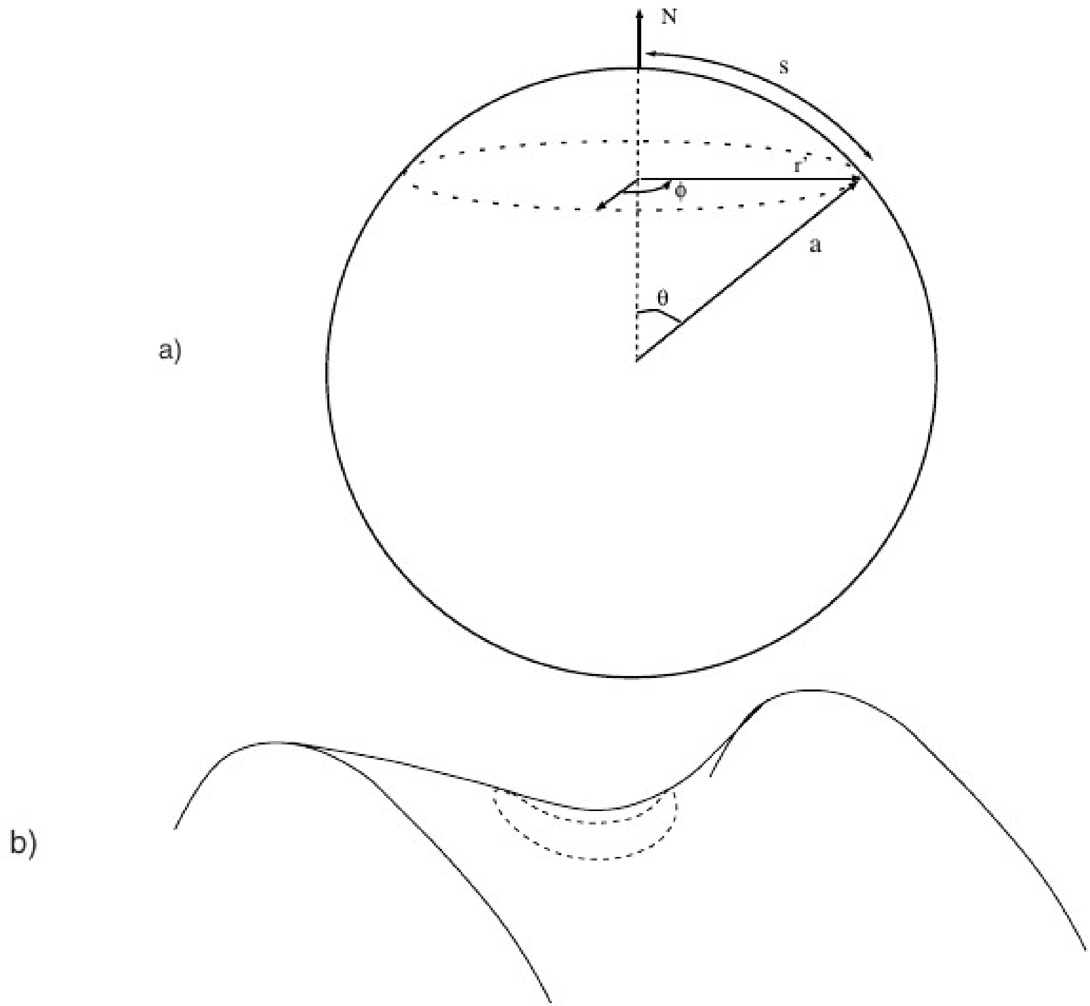


Fig. 3.5. (a) The circumference $2\pi r'$ and radius s of a circle drawn out on a sphere have a ratio less than 2π. (b) The circumference and radius of a circle drawn out on a saddle surface have a ratio greater than 2π.