GEOMETRY AND PHYSICS: AN OVERVIEW

1 Geometry

Our theme is general relativity, seen not just as a theory of gravity, but as the expression of an approach to the world in which geometry and physics are united. The subsequent chapters will set out systematically the concepts and techniques needed; here we shall try to give an overview of the ideas that have led us to this approach, and to remind our readers of the basic facts of special relativity, with which we assume they have some acquaintance.

The laws of Euclidean geometry have since antiquity had a double status. On the one hand they were part of technology: empirical discoveries revealed when man started using the ruler and compass, which were developed and put to use in surveying and building. On the other hand they were for the Greeks examples of truths that could be discovered by pure reason, an innate capacity of the mind (as argued by Plato in the Meno, for example), which therefore did not depend for their validity on empirical measurements. This gave geometry an absolute status. Since the application of reason alone could derive these truths, they could not be other than what they were. Although, from Aristotle onwards, it was noted that there was a possibility of altering some of the axioms – notably the axiom about parallels not meeting and there being exactly one line parallel to a given line through a given point – without there being a logical contradiction, nonetheless such alterations were regarded as meaningless logical games. The fact that modifications in the axioms were manifest nonsense constituted a reductio ad absurdum proof of their absolute validity. In Kant's terminology, the laws of geometry were a priori truths which formed the necessary background conditions for the empirical world.

By the nineteenth century it was of course known that there was a sense in which non-Euclidean geometry was possible at a rather stronger level than that of a formal logical consistency: there exists what we now call a model for a non-Euclidean geometry. Namely, if one agrees that the term "straight line" denotes a great circle on a sphere, then it follows that the angles in a triangle add up to more than two right angles and parallels do not exist. But for most this was not held to violate the absolute status of Euclidean geometry. One used Euclidean geometry to prove the statement just given, and only escaped from its axioms by verbal trickery. A circle is not turned into a straight line by perversely calling it one, and it remains true that real straight lines obey the parallel axiom.

Two trends led to a shift in this consensus. First, mathematicians started to realise that non-Euclidean geometries were fun; they were interesting intellectual constructions in their own right and so a part of pure mathematics. Secondly, philosophers became increasingly suspicious of a priori truths. If an assertion was made about empirical objects, then the truth or falsity of that assertion had an empirical content. If by a "straight line" one means a real straight line drawn with a ruler or some other means then the properties of real straight lines can be examined empirically, and if it turns out that their properties are non-Euclidean (for instance, if very delicate measurements show that the angles of a triangle add up to more than two right angles), this only shows that real "straight lines" are not appropriately represented by the Euclidean ideal. And in such circumstances, we should have to enlist the help of mathematicians to provide us with a more appropriate non-Euclidean geometry.

A vital contributor to this development was Poincaré, who examined the meaning of geometrical terms in an empirical context. If we are to discover the geometry of the world, empirically then we must first decide on our procedure for making measurements. We must give a concrete specification of how we are going to draw a straight line, how we are going to measure a length and so on. In the case of length, for example, we must fix on an appropriate sort of measuring rod that is to be carried around from place to place in order to determine the ratios of different intervals in different places.

Poincaré gave the following example, which exactly illustrates the thinking that led to general relativity. Imagine a two-dimensional world confined to a disc in Euclidean space, inhabited by two-dimensional beings with measuring rods. And suppose that these beings and their rods are affected by a mysterious influence that causes them to shrink as they approach the edge of their world. Then for a certain simple shrinking law, the geometry which these creatures will determine by their empirical measurements will be non-Euclidean: the edge will be infinitely far away and, in this case, the angles of a triangle add up to less than two right angles and there are infinitely many parallels to a given line through a given point.

If we further suppose that these beings are strict empiricists, then they will conclude that their geometry is non-Euclidean. Any dissident non-empiricist among them might propose the alternative explanation that they live on a Euclidean disc affected by a shrinkage-influence. But the issue could only be decided if they could communicate with the God-like three-dimensional beings who could see their world from the outside and reveal to them that the non-empiricist view fitted better with the three-dimensional world.

For us, the evidence is that the empirical geometry of space-time is not Euclidean. The non-empirical option is always open to us: we could hold that space-time is "really" Euclidean, but shrinkage-forces distort the situation. But in the absence of a revelation from five-dimensional beings outside our world we would have no way of knowing what the "real" Euclidean geometry is; no way of knowing, for example, which, if any, of the infinitely many families of curves that could be drawn so as to obey the axioms for Euclidean straight lines, were straight lines as seen from outside our world. So in practice we have little choice but to accept that the appropriate geometry for describing our world is not a Euclidean one.

Before moving on to space-time, one qualifying remark is needed. The non-Euclidean geometry of which we have been speaking is actually still too like Euclidean geometry to be an appropriate tool for general relativity. In non-Euclidean geometry it is assumed that the elements of Euclidean geometry (points, lines, planes, distance, angle, etc.) are all defined, and that all results are to be deduced from a set of axioms that specifies how these elements intersect (such as, the axiom that there is exactly one straight line through any pair of points). What will be needed for general relativity is differential geometry, whose axioms make no statements about lines and planes as a whole, but only about their limiting behaviour when very small regions are considered. The result is not Euclidean, and it is a sort of geometry!

2 Special relativity

Special relativity is a framework for physics in which space and time have a fundamentally different structure from that which obtains in Newtonian physics, although the geometrical concepts remain basically Euclidean. General relativity is a generalisation of this framework in the same way that differential geometry is a generalisation of Euclidean geometry, and so we need to review some of the fundamental ideas of special relativity, referring the reader to one of the many available textbooks for details.

Newtonian physics rests on one of the most fascinating paradoxes in the history of science. Its key fact, which forced the break with the Aristotelian approach to dynamics, was the realisation by Galileo that a uniform movement is indistinguishable from a state of rest, from the point of view of an observer undergoing the movement. To use Galileo's own example, a person enclosed in a ship sailing steadily across a smooth sea could not perform any local experiment, confined to the cabin of the ship, which would enable him to determine whether or not he was moving through the water. This meant that acceleration, not velocity, was the prime kinetic quantity, and the laws of dynamics had to be laws determining acceleration. But the concept which enabled Newton to give the definitive form to those laws was his idea of absolute space, thought of as a direct manifestation of God. Motion was simply motion from one point in absolute space to another, and to describe it all that was necessary was to set up a Cartesian coordinate system in absolute space. The theory that enshrined the equivalence of rest and motion was based on a mathematical description in which rest and motion (relative to absolute space) were absolutely distinct.

The empiricist argument acts in Newtonian physics in the same way as in geometry. Newton's laws themselves ensure that the absolute space in terms of which they are formulated is an unobservable chimaera. All we can actually observe is not absolute place but relative place. I am in the same place now relative to the earth as I was ten minutes ago, but several thousand kilometres further on relative to the sun. With the benefit of hindsight we can say that absolute space, as an array of places that preserve their identity throughout time enabling us to say that I am now in the same place as before, is a myth. The only reliable background in which to situate events is space-time, the set of all places-at-particular-times. My writing this sentence occupies a different place-and-time as my writing the first sentence of this chapter; but I cannot meaningfully say that it occupies a different place, in an absolute sense. Thus the idea of space-time is inherent in the laws of Newtonian dynamics.

In Newtonian physics the two events just referred to can be said to occupy a different time, though not meaningfully a different place. This is because a time-ordering is a basic part of the assumed structure of Newtonian space-time. For any two events A and B it is possible to say either that A is before B, or B is before A, or they are simultaneous. Consistency in this order requires that simultaneity is an equivalence relation: space-time is divided up into equivalence classes of mutually simultaneous events, with each class constituting the universe at a given time.

In special relativity this concept of simultaneity is abandoned. Each observer has his own idea of simultaneity, but different observers disagree (just as, in Newtonian physics, each has his idea of "in the same place"). Thus in Newtonian dynamics (or better, in view of Newton's own predelictions, Galilean dynamics) there is absolute time but no absolute space, whereas in special relativity there is neither.

There remain some absolute structures, however. In both theories there exists a privileged class of observers, called inertial observers, characterised by the fact that, like Galileo's observer in the cabin of a ship, they do not feel any of the characteristic effects of acceleration. By contrast a non-inertial observer, who is necessarily accelerating relative to any inertial observer, feels apparent *inertial forces*, such as centrifugal and Coriolis forces, attributable to the acceleration. (Recall that if one is in a car that is accelerating forwards, then one feels a force pushing one backwards.)

The two different structures on space-time, Galilean and special relativistic, are expressed through the transformation laws relating the coordinates which different inertial observers place on space-time, using the same standard measuring techniques. (A coordinate system constructed by a standard procedure is called a *frame of reference* or simply a *frame*, and the coordinates of an inertial observer are called an *inertial frame*.) To a large extent these laws of transformation between inertial frames are the same. If x, y, z, t and x', y', z', t' are the three coordinates of space and one of time defined by two different inertial observers O and O', respectively, then they are related by some combination of the following (where we write $\vec{r}$ for (x, y, z) and $\vec{r}'$ for (x', y', z'), and $\vec{a}$ and b are a vector and a number depending on the two observers chosen):

(i)	a relative rotation of the space coordinates	$\vec{r} = A\vec{r}'$
(ii)	a displacement of the space coordinates	$\vec{r} = \vec{r}' + \vec{a}$
(iii)	a displacement of the time coordinates	$t = t' + b$
(iv)	a boost (see below)	

The difference comes in the form of the boost. In the Galilean case it is given by

$$t = t'$$

$$x = x' + vt'$$

$$y = y'$$

$$z = z'$$

while in the case of special relativity it is given by

$$ct = \frac{ct' + (v/c)x'}{\sqrt{1 - v^2/c^2}}$$
$$x = \frac{x' + vt'}{\sqrt{1 - v^2/c^2}}$$
$$y = y'$$
$$z = z'$$

where v is a number depending on the observers and c is an absolute constant of nature with the dimensions of velocity. We note the crucial distinction that in the Galilean case the time coordinate is unchanged, so that the observers agree on the definition of simultaneity. The above form shows that c has the role of a conversion constant (like Boltzmann's constant in thermodynamics) needed to relate the quantities t and x which have different physical dimensions when normal units are used.

It is a consequence of these transformations that if a signal or particle is moving with the speed c according to one observer, then it is moving with the same speed relative to the other. (As Fraser has expressed it, relativity replaces Newton's absolute rest by an absolute speed.) It turns out that light propagates with just such a speed, and so has the same speed according to all observers. This phenomenon, which runs quite contrary to Galilean thinking, was a consequence of Maxwell's description of electromagnetism, and it provided the motivation for the development of special relativity. The transformation laws can in fact be deduced from the constancy of the velocity of light, together with a few plausible symmetry assumptions. Experimentally this was checked to a high accuracy by the Michaelson-Morley and Kennedy-Thorndyke experiments.

3 General relativity

The single most important step in the development of general gelativity from special relativity is a change in the application of the idea of inertial observers, a change forced upon one by the distinctive properties of gravitation. As was already known to Galileo, a gravitational field accelerates all bodies in the same way, independently of their mass and internal constitution. A simple Newtonian argument clarifies this. The property of a body whereby it resists any attempt to change its state of motion is its *inertial mass* m_i. This enters the Newtonian equation of motion

$$\vec{F} = m_i \vec{w} \tag{3.1}$$

where $\vec{F}$ is the total force acting on the body and $\vec{w}$ is the acceleration. On the other hand the property which determines a body's response to a gravitational field is the *gravitational mass* m_g which enters the law for the gravitational force $\vec{F}_g$ acting on the body:

$$\vec{F}_g = m_g \vec{g} \tag{3.2}$$

where $\vec{g}$ is the gravitational field intensity, generated according to the inverse square law by a gravitating body. The constancy of $\vec{w}$ implies, from (3.1) and (3.2), that

$$m_i / m_g = \kappa \tag{3.3}$$

a constant, independent of the body chosen. By choosing the units of $\vec{g}$ suitably we can arrange that $\kappa = 1$, leading to $\vec{w} = \vec{g}$, independently of the body's mass.

Relation (3.3) has been experimentally verified to a very high accuracy (at best 10^{-12}) for a large range of masses, from a neutron (Witteborn and Fairbank, 1967), to an apple (Roll et al., 1964; Braginsky and Panov, 1971; etc.) to a planet (Williams et al., 1976).

The phenomenon is seen as even more remarkable when we realise that, according to special relativity, a (negative) contribution to the inertial mass of the body is its binding energy – something quite different in kind from the rest mass of the constituent particles. It might be understandable if the gravitational mass were proportional to the number of particles, so that each particle had a sort of unit gravitational charge, which would in turn be approximately equal to the inertial mass. But the accuracy of the experiments rules this out: it is the inertial mass, in all its various forms, and nothing else, that is related to the gravitational mass.

The gravitational field is not the only place in nature where an acceleration is independent of the nature of the body involved: the same thing happens with so-called inertial forces, the forces felt by an observer who is not inertial to which we have already referred in defining an inertial observer. In the case of uniform acceleration, for example, the form of the inertial forces can be derived immediately from Newton's laws of motion. If $\vec{F}$ denotes a "real" (i.e. not just inertial) force acting on an object, $\vec{r}$ denotes its position vector in an inertial frame and $\vec{r}'$ its position in a uniformly accelerating frame, then the two frames will be related by

$$\vec{r}' = \vec{r} - \frac{1}{2}\vec{w}t^2$$

and so, while Newton's law of motion in the inertial frame reads

$$m\ddot{\vec{r}} = \vec{F}$$

if we transform to the accelerating frame by substituting for $\vec{r}$ in terms of $\vec{r}'$ we have

$$m\ddot{\vec{r}}' = \vec{F} - m\vec{w}.$$

Thus in the accelerating frame every object is subjected to a "fictitious force" that is proportional to its mass (just as the gravitational force was proportional to the mass) in addition to any "real" force acting on it. A calculation would yield the same result in the case of the fictitious force due to a rotational acceleration (the combined Coriolis and centrifugal force). This proportionality to the mass, which arises automatically simply because we have created the force by transferring a term from the $m\ddot{\vec{r}}$ of Newton's equations from one side of the equation to the other, is a distinctive feature of the fictitious forces, distinguishing them from, say, the electrostatic force which depends on the charge rather than the mass.

The similarity between the gravitational force and fictitious forces makes it possible to annul the effect of the gravitational force by transforming to a coordinate system that is accelerating with precisely the gravitational acceleration. This is termed a *freely falling coordinate system* or *frame*. An example is the coordinate system defined by observers in a satellite in orbit, in free-fall under gravity, where objects float freely with no gravitational effects. So the question naturally arises, whether gravity itself is a fictitious force.

For the fictitious forces of acceleration and rotation it is possible to transform back to the original inertial frame in which the fictitious force is annulled everywhere. This is not the case with the gravitational force because the direction of the gravitational acceleration varies from place to place. A coordinate system that is freely falling for observers in Italy will be accelerating in the wrong direction for observers in Australia. So we could only regard gravity as a fictitious force if we were prepared to localise the idea of a coordinate system. We can seek to annul gravity, not everywhere but only sufficiently close to some particular freely falling observer. How close this is depends on the accuracy with which our measurements are to be carried out.

But to implement even this restricted idea of reducing gravity to a fictitious force, it is essential that the effect of gravity is precisely equivalent to the effect of an acceleration, so that by passing to a (local) freely falling frame, gravity can be cancelled exactly. The assertion that this is indeed the case is called the *principle of equivalence*. It entails two ideas: as far as mechanical effects are concerned, the gravitational force must be exactly proportional to the inertial mass (equation 3.3); and any non-mechanical effects of acceleration (for example, electromagnetic effects) must be exactly the same for gravity.

The electromagnetic effect of gravitation (manifested in the red-shift of electromagnetic radiation) has been verified to one per cent in laboratory experiments (Pound and Rebka, 1960). So, combining this with the overwhelmingly impressive verification of the mechanical equivalence, we shall here take the principle to be verified.

If we are prepared to work with a frame that is only local, then, by taking freely falling frames as our standard inertial frame, gravity can be abolished near any chosen observer, and the gravita-