General Relativity

Installment I.

August 31, 2003

Prologue

Consider a bead that slides freely on a stationary wire. Imagine that there is no friction, no air resistance, in this highly idealized world consisting of only the wire and the bead. This world is not quantum mechanical either.

If we let the bead be, it will sit there in a fixed place. If the wire is straight, and we give the bead an impulse, the bead will slide along at constant velocity. That is, the bead resists changes of its state of motion, as Galileo and some of his contemporaries realized it would more than three hundred years ago and as Newton implied in his first law of motion.

Now what happens to the motion of the bead as it moves along the wire if it comes to a place where the wire bends? Since we have assumed that the wire is stationary, we must conclude that the bead will stay on the wire — it corners nicely. Why? What keeps the bead on the wire? From our three-dimensional perspective, the bead has changed its velocity and so, if it is obeying Newtons' s second law, a force must have been exerted. This force, we normally say, is exerted by the wire on the bead. If I tell you the shape of the wire, you should be able to figure out what that force must be.

But there is a simpler way to think about this problem. We may assume that the bead lives in a world that is defined by the wire and, even it if were sentient, it would know nothing of our three-dimensional existence. In that formulation of the bead problem, we simply postulate that the bead is *constrained* to remain in the one-dimensional world defined by the wire. It is we, looking down on the bead and wire from our three-dimensional perspective, who worry about why the bead stays on the wire in order to be able to describe what is happening in our world. But may we not then wonder whether beings in higher dimensions than three are looking down on us (so to say), perhaps only in their imaginations, and deciding what the reason for our remaining in our own space may be? Is there a force constraining us to remain in our three-dimensional world? At any rate, we are seemingly confined to that world of three (or four, if we include time) dimensions and there is nothing to do about it for now. Like the bead on the wire, we move along in our own world, so here we begin by trying to determine what this world is like.

As most of us of have observed, the geometry of the nearby world is pretty nearly the geometry of Euclid. That is why the axioms of Euclidean geometry seem so natural to us and it is why many people regard geometry as a triumph of physics as well as of mathematical reasoning. But like the hypothetical wire, our world has bends in it, though they are hard to imagine for us living in this world. As objects such as the moon and the planets (and falling apples) move in our space, they follow paths that do not seem straight to us who think in terms of Euclidean geometry. To explain this behavior, Newton (and others) proposed that there is a gravitational force that forces them to behave in this way. We shall review briefly Newton's force and some of its features. However, here we are more concerned with the suggestion made by Einstein in the twentieth century that the planets are not driven by forces but that their seemingly bent paths are the analogues of straight lines in a curved geometry. The motions of the planets are like those of the bead on the bent wire wire or a ship circumnavigating the earth.

To see how Einstein's vision can be expressed in quantitative terms, we need to go beyond the geometry of Euclid. To learn what geometry we need for this, we shall go over classical mechanics, looking for clues. We shall also want to make the outcome compatible with the world as it is imagined in relativity, which we shall review as well. When we put it all together, we shall have made a semblance of Einstein's theory of gravity, a.k.a. general relativity. Along the way, we shall take a few digressions following the advice of Yogi Berra: "When you come to a fork in the road, take it."

So here goes.

1 Classical Mechanics

1.1 Newton's second law

Galileo knew that if a body in motion was not subjected to any external influences, it would keep on in its state of motion. This property of material objects is called inertia and it is expressed mathematically by Newton's first law of motion. Even more remarkably, Galileo stated what we may call the first statement of relativity in which he suggested that the laws of nature are not modified by uniform motion. This is what he wrote about it:

Shut yourself and a friend below deck in the largest room of a great ship, and have there some flies, butterflies, and similar small flying animals; take along also a large vessel of water with little fish inside it; fit up also a tall vase that shall drip water into another narrow-necked receptacle below. Now, with the ship at rest, observe diligently how those little flying animals go in all direction; you will see the fish wandering indifferently to every part of the vessel and the falling drops will enter into the receptacle placed below .... When you have observed these things, set the ship moving with any speed you like (so long as the motion is uniform and not variable); you will perceive not the slightest change in any of the things named, nor will you be able to determine whether the ship moves or stands still by events pertaining to your person.

You can read about this in "Galileo Galilei" by Ludovic Geymonat (McGraw-Hill 1965, pp.119-20). Galileo adds

And if you should ask me the reason for all these effects, I shall tell you now: Because the general motion of the ship is communicated to the air and everything else contained in it, and is contrary to their natural tendencies, but is indelibly conserved in them.

One of the most important of these laws of Nature is Newton's second law of motion. According to Newton, an external action or force is needed to change the state of motion of a body. A force produces an acceleration proportional to its strength and in the same direction as the force:

$$\mathbf{f} = m\mathbf{a} \tag{1}$$

where $\mathbf{f}$ is the force and $\mathbf{a}$ is the acceleration. The constant of proportionality, m, measures the body's resistance to the change of its state of motion and it is called the mass, or the inertia.

Of course, both Galileo and Newton were thinking that all this activity and its descriptions were taking place in a space that fulfills the axioms of Euclid to very good approximation. In such a space, we can install Cartesian coordinates with a conveniently chosen origin. To go from that origin to any point with coordinates (x, y, z) we make a displacement of a given amount and in a given direction. Such a displacement is traditionally indicated by an arrow of the right length and direction. This arrow is often called a vector and is denoted by $\mathbf{x}$. The association between a point in space, designated by its coordinates, (x, y, z), and a displacement vector, $\mathbf{x}$, is a wonderful convenience of Euclidean space that simplifies the expression of Newton's second law of motion.

Suppose we idealize the body obeying the law as just a point endowed with inertial mass. We call such a body a particle, from the Latin word *pars* meaning part with the addition of the diminutive ending *icle*. We may describe the particle's motion by associating to each point along the trajectory a vector $\mathbf{x}$ that varies with time, t. The dependence of the position $\mathbf{x}$ on time t describes the motion and we express this dependence by writing $\mathbf{x}(t)$. The information embodied in this description is what astronomers call an ephemeris, though in practice it can be provided only at a finite number of discrete times.

The quantity t is mysterious. St. Augustine said (roughly) "I know what time is until someone asks me to define it." Indeed, basic quantities such as mass, time and coordinates are exceedingly hard to define. The more elementary something is, the harder it is to define since there is nothing more elementary in terms of which to define or explain them. Such things are perhaps best defined by prescriptions for their determination. Thus, there is somewhere stored some well-conserved object used as a standard of mass and also some variable object that is used to define units of time. So we have prescriptions for measuring the basic quanitities even though it is not easy to know how to define them abstractly. One way to deal with these deep things is to suggest you will learn what these are when you read a more advanced treatment. Experience shows, however, that when you get there you will be assumed to have learned what these quantities are by reading elementary books. Such dilemmas are not unique to physics. In his book called "Money, " J.K. Galbraith says

It will be asked in this connection if a book on the history of money should not begin with some definition of what money really is. ... The precedents for such effort are not encouraging. Television interviewers with a reputation for penetrating thought regularly begin interviews with economists with the question: "Now tell me, just what is money anyway?" The answers are invariably incoherent.

Here we shall call on your experience of basic things and not try to go more deeply into them as we move onto more advanced topics.

2 Trajectories

The situation then is that we are relying on our daily experience to guide us in thinking of the motion of a particle in space and its trajectory. We still ought to make a more precise statement of what such a trajectory is. For this we call on the ideas of mathematicians, without striving for their rigor.

Consider a straight line lying in Euclidean space. Mathematicians call this one-dimensional space $\mathbf{R}$ since to every point on the line one can associate a real number, s (say), that increases monotonically and continuously along the line. Now if this line is rigid, we can pick it up and toss it into space. When we snap our finers, the line goes flaccid and develops bends, loops and similar complications as it flies out there. When we snap our fingers a second time, $\mathbf{R}$ goes rigid and stays put. So now we have a curve in space. Each point on the curve has carried with it its value of s and each point also has acquired coordinates (x, y, z) (assuming that the space has three dimensions). So we have automatically obtained an association between the coordinates in space and position along the curve. That is, the coordinates of the points along the curve are functions of s and we write this as $x(s), y(s), z(s)$, or $\mathbf{x}(\mathbf{s})$. The motion of our particle takes place along one such curve, or trajectory.

We could have associated the parameter s to points on the curve in a number of ways. If the curve is indeed a trajectory of our particle, we could complete the description of the motion by specifying the time t at which the particle passes through each point $\mathbf{x}(\mathbf{s})$. That is, we give $s(t)$ or $\mathbf{x}(\mathbf{s}(\mathbf{t}))$, written in the mildly inconsistent but suggestive way of some physicists as $\mathbf{x}(t)$. By assigning a value of t to each point along the curve, we have given a new parametrization of the curve, one that is significant in the description of the motion of the particle. The velocity of the particle is then the time derivative of the position.

Let $\mathbf{x}_0$ be the position of the particle at time t_0 and $\mathbf{x}_1 = \mathbf{x}_0 + \delta\mathbf{x}$ be the position at time $t_1 = t_0 + \delta t$. Draw a straight line through $\mathbf{x}_0$ and $\mathbf{x}_1$. When δt approaches zero, this line becomes the tangent line to the curve at the point $\mathbf{x}$. The instantaneous velocity of the particle at t_0 is along this line and it is given by

$$\mathbf{v}(t) = \lim_{\delta t \to 0} \frac{\delta \mathbf{x}}{\delta t} \tag{2}$$

The right side of this expression is the derivative of position with respect to time and we write the velocity vector as

$$\mathbf{v}(t) = \frac{d\mathbf{x}}{dt}. \tag{3}$$

(We shall be looking at vectors more carefully, but let us for now continue in this casual manner meant to provide an intuitively helpful entrée to the kinematics of particle motion.) The velocity of the particle is then the tangent vector to the trajectory when the points along it are labeled by the time at which the particle passes through them. Naturally, we need to figure out how to do that labelling, namely how to deal with the dynamics of the particle, but that too will come later.

Let us think back now to the illustration of the bead on the hoop. The curve we have been talking about is the wire that the bead slides along; this wire is the world that the bead inhabits. The vector that we have just introduced to describe the velocity of the bead does not lie in that world but in the tangent line to the bead's world, or wire. In Euclidean space, we do not suffer such an inconvenience, but for someone living in the wire, the velocity vector cannot be part of heir world — it is confined to the tangent line. The tangent line at any point $\mathbf{x}$ on the trajectory is called the tangent space of the wire and is denoted (in mathematical language) by $\mathbf{T_x}$. The tangent space is the vector space inhabited by the velocity of the bead and there is such a tangent space at every point on the wire.

As we shall see, in Einstein's description of our world, we give up the Euclidean vision of our geometry and so we must give up the luxury of drawing vectors, such as velocities, in the space itself. As for the wire, we shall need to introduce a tangent space analogous to $\mathbf{T_x}$. This is already evident when we think of the velocity of a ship sailing across the Pacific Ocean. We shall not only need a vector space at each point in our world, we shall also have to devise a suitable vector calculus for dealing with dynamical issues that arise. For now though let us close this section by stating that we can compute the derivative of $\mathbf{v}$ just as we computed the derivative of $\mathbf{x}$. In this way, we get the acceleration,

$$\mathbf{a} = \dot{\mathbf{v}} = \ddot{\mathbf{x}} \tag{4}$$

Newton's second law is then written as

$$\mathbf{f} = m\ddot{\mathbf{x}} \tag{5}$$

That part of the force that keeps the bead on the wire is neither in the wire nor in $\mathbf{T_x}$. That is why we said that it is of no real concern to the bead, even though we do worry about it ourselves.

3 Work and Energy

Another convenience of working with Cartesian coordinates in Euclidean space is that the square of the length of the position vector $\mathbf{x}$ is the sum of the squares of the coordinates. We call this quantity the scalar product of $\mathbf{x}$ with itself. This product is denoted by a dot, so it is also called the dot product:

$$\mathbf{x} \cdot \mathbf{x} = x^2 + y^2 + z^2 \tag{6}$$

This is sometimes written in the suggestive form $\mathbf{x} \cdot \mathbf{x} = \mathbf{x}^2$. Similarly, if we denote the components of $\mathbf{v}$ by (u, v, w) we can write the dot product of the velocity with itself as

$$\mathbf{v}^2 = u^2 + v^2 + w^2 \tag{7}$$

Now if we take the dot product of $\mathbf{v} = \dot{\mathbf{x}}$ with the law of motion (5), we find that

$$m(\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z}) = \mathbf{v} \cdot \mathbf{f} \tag{8}$$

The quantity on the right of this equation is called the rate of working of the force $\mathbf{f}$. The quantity on the left is the rate of change (or time derivative) of $\frac{1}{2}m(\dot{x})^2 + \frac{1}{2}m(\dot{y})^2 + \frac{1}{2}m(\dot{z})^2$, the kinetic energy. So we have

$$\mathbf{f} \cdot \dot{\mathbf{v}} = \frac{1}{2}m\frac{d}{dt}(\mathbf{v})^2 \,. \tag{9}$$

The rate of working of $\mathbf{f}$ is the rate of change of the kinetic energy of the particle.

We know that certain forces, such as Newton's force of gravity, can be expressed as the derivative (or gradient) of a function of space and time called the potential. Without discussing what there is about $\mathbf{f}$ that lets us do this, let us just assume that we may, for now, write that

$$\mathbf{f} = -\nabla\phi \tag{10}$$

where the function ϕ is assigned a value at every point in the space for every time; we write this as $\phi(\mathbf{x}, t)$. The minus sign in (10) is included by convention and, though it is not necessary, its absence would make some people uncomfortable.

When we introduce (10) into (9), we can integrate the resulting equation and we find that

$$\frac{1}{2}m\mathbf{v}^2 + \phi = E \tag{11}$$

where E is a constant of integration called the energy. The fact that E is constant expresses expresses the law of conservation of energy for this simple example of Newton's second law. It says that the total energy, kinetic energy ($\frac{1}{2}m\mathbf{v}^2$) plus potential energy ($\phi(\mathbf{x}, t)$) is a constant.

If we keep to the simple case where the force is the gradient of a potential, we may write Newton's second law as

$$m\dot{\mathbf{v}} + \nabla\phi = \mathbf{0}\,. \tag{12}$$

When the two terms in this equation are together like that on the same side of the equal sign, we feel the temptation to think they have something in common. Some people even call $m\dot{\mathbf{v}}$ a force — the intertial force. We may then define a total force as

$$\mathbf{F} = m\dot{\mathbf{v}} + \nabla\phi\,. \tag{13}$$

It is sometimes useful to go beyond Galileian relativity and transform ourselves into a coordinate system moving so that $\mathbf{F} = \mathbf{0}$ and to make Newton's second law look like a version of the first law. This is a hint of what is coming, but let us not get ahead of our story.

But here is a last thought on these matters before going on to the next section. Suppose we use (11) to define the energy without worrying about where that expression for E came from. Then we would not know that E is supposed to be a constant and we might innocently ask how it changes in time. That is, we might compute

$$\frac{dE}{dt} = m\mathbf{v}\cdot\dot{\mathbf{v}} + \mathbf{v}\cdot\nabla\phi\,. \tag{14}$$

(This was done by noting that $d\phi/dt = \mathbf{v}\cdot\nabla\phi$.) Then (14) may be written as

$$\frac{dE}{dt} = \mathbf{v}\cdot\mathbf{F}. \tag{15}$$

So if we want to conserve energy ($dE/dt = 0$) we need only to require that the total force, $\mathbf{F}$, be perpendicular to the velocity. This is a weaker condition than the demand that Newton's second law be satisfied and is merely the demand that the total force does no work. In our example of a bead on the wire, the velocity of the bead is tangent to the wire. For energy to be conserved, the total force must therefore be perpendicular to the wire, as is the force needed to keep the bead on the wire. Thus, keeping an object in its space does not interfere with the conservation of energy, which is a lesson to think about.

4 More Mechanics

We have seen that the sum of the kinetic energy and the potential energy is conserved in the special case we looked at. We might ask if there is anything interesting about the difference between the two kinds of energy in similarly simple circumstances. That is, is there anything about

$$L = \frac{1}{2}m\mathbf{v}^2 - \phi(\mathbf{x}, t). \tag{16}$$

that we ought to know? This quantity is called the Lagrangian (or Lagrangean) after Luigi Lagrange, as he is known in Turin, his birthplace. The derivative of L is

$$\frac{dL}{dt} = m\mathbf{v} \cdot \dot{\mathbf{v}} - \mathbf{v} \cdot \nabla\phi\,. \tag{17}$$

There does not seem to be much to say about this. It however turns out that what is interesting about L is not so much its dervative as its integral.

In our three-dimensional space, with no wire to guide it, a particle must choose its route to go from P_1 to P_2. In the simplest vision of mechanics, this choice is dictated by the forces acting. Remarkably, this is just the route that gives the minimum value of A among possible trajectories between the two points. A trajectory is specified by giving the coordinates locating the particle at every instant t between t_1, the time of departure from P_1, and t_2, the arrival time at P_2. Let us designate one such trajectory as $\mathbf{x}(t)$, where $\mathbf{x}(t_1) = P_1$ and $\mathbf{x}(t_2) = P_2$. The action is then defined as

$$A = \int_{t_1}^{t_2} L dt, \tag{18}$$

which is taken between any two points in space, P_1 and P_2.

To think about how to find the minimum of A, let us imagine a space in which each point corresponds to a particular trajectory from P_1 to P_2. To every point in that space, we may associate a value of A, which is then a function defined on the space of trajectories. This is a rather abstract notion. A trajectory is itself a function of time, so that A is a function of functions. Such things are called functionals.

For any particular trajectory $\mathbf{x}(t)$, we can evaluate A. And for a trajectory close to that one, say $\mathbf{x}(t) + \delta\mathbf{x}(t)$, we can also make that evaluation and obtain a the new value, $A + \delta A$. Here we are assuming that at every time t, $\delta\mathbf{x}(t)$ is a small separation between corresponding points on the two trajectories. We may then construct the difference

$$\delta A = A[\mathbf{x}(t) + \delta\mathbf{x}(t)] - A[\mathbf{x}(t)] \tag{19}$$

To evaluate δA approximately, we note that, along the perturbed trajectory, we have

$$L = \frac{1}{2}m[\dot{\mathbf{x}}(t) + \delta\dot{\mathbf{x}}(t)]^2 - \phi(\mathbf{x} + \delta\mathbf{x}, \mathbf{t}) \tag{20}$$

Then, for small $|\delta\mathbf{x}|$, we can estimate that,

$$L \approx \frac{1}{2}m\dot{\mathbf{x}}(t) \cdot \delta\dot{\mathbf{x}}(t) - (\nabla\phi) \cdot \delta\mathbf{x}(t) \tag{21}$$

where the last term comes from taking the leading term in the Taylor expansion for ϕ. We see then that

$$\delta A = \int_{t_1}^{t_2} \left[\frac{1}{2} m\dot{\mathbf{x}}(t) \cdot \delta\dot{\mathbf{x}}(t) - \delta\mathbf{x} \cdot \nabla\phi\right] dt \tag{22}$$

We integrate the first term on the right by parts and note that there is no contribution from the end points since all trajectories go from P_1 to P_2. Then we find

$$\delta A = -\int_{t_1}^{t_2} [\ddot{\mathbf{x}} + \nabla\phi] \cdot \delta\mathbf{x} dt$$

As for functions, functionals like A are stationary at their extrema, their minima and maxima. If we want A to be an extremum for the preferred trajectory, we should expect δA to be zero on that trajectory. We want $\delta A = 0$ on the correct trajectory and this should be true for any $\delta\mathbf{x}$. This latter stipulation ensures that we must have

$$\ddot{\mathbf{x}} + \nabla\phi = \mathbf{0} \tag{23}$$

This is Newton's law of motion when the force is $\mathbf{f} = \nabla\phi$.

Notice that all that we have really determined is that A is stationary on the correct path (as given by Newton), but this is remarkable enough. More importantly, we have seen an example of how to find the condition for a functional to have a stationary value. This is an important process that it will be useful for us to know and that is really why we have taken the trouble to look at this example in a bit of detail.

5 The Equations of Lagrange

When we obtain Newton's equation of motion by demanding that the action be an extremal when evaluated on the trajectory that the Newton's law requires we are doing more than noting an interesting side line. We are discovering a way to derive governing equations in general once we state thee Lagrangian approach in general terms. We have, in other words, uncovered a powerful means to derive laws of physics. We shall do a once over lightly of this procedure in this section even though it is not essential for learning relativity. You should either just glance it over or skip it for now as it is not essential for following the rest of the rest of these notes.

In (16) we see that the Lagrangian is an explicit function of both $\mathbf{x}$ and $\dot{\mathbf{x}}$. It will also be a function of t if ϕ is. So let us we write the Lagrangian as $L(\mathbf{x}, \dot{\mathbf{x}}, t)$. The action is, as before,

$$A[\mathbf{x}(t)] = \int_{t_1}^{t_2} L(\mathbf{x}, \dot{\mathbf{x}}, t) dt \tag{24}$$

where the integral is taken along the trajectory implied by $\mathbf{x}(t)$.

Again, we take the difference between the integral over the trajectory $\mathbf{x}(t)$ and a neighboring one $\mathbf{x}(t) + \delta\mathbf{x}(t)$ to obtain

$$\delta A[\mathbf{x}(t)] = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial \mathbf{x}} \cdot \delta\mathbf{x} + \frac{\partial L}{\partial \dot{\mathbf{x}}} \cdot \delta\dot{\mathbf{x}}\right] dt \tag{25}$$

This time, we need to do some interpreting of the terms in this formula. For example, for brevity, we have written

$$\frac{\partial L}{\partial \mathbf{x}} \cdot \delta \mathbf{x} = \frac{\partial L}{\partial x} \cdot \delta x + \frac{\partial L}{\partial y} \cdot \delta y + \frac{\partial L}{\partial z} \cdot \delta z \tag{26}$$

That is, by $\partial L/\partial \mathbf{x}$ we mean what in ordinary vector calculus we would write as ∇L. We do not use the more familiar notation since we would not be able to write a similar thing for $\partial/\partial \dot{\mathbf{x}}$ in the abbreviation

$$\frac{\partial L}{\partial \dot{\mathbf{x}}} \cdot \delta \dot{\mathbf{x}} = \frac{\partial L}{\partial \dot{x}} \cdot \delta \dot{x} + \frac{\partial L}{\partial \dot{y}} \cdot \delta \dot{y} + \frac{\partial L}{\partial \dot{z}} \cdot \delta \dot{z} \tag{27}$$

We may once more integrate by parts to obtain

$$\delta A[\mathbf{x}(t)] = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{x}}} \right] \cdot \delta \mathbf{x} dt \tag{28}$$

For the trajectory that makes A smallest we require that $\delta A = 0$. This condition is met when the integral in (28) is zero, for small but arbitrary $\delta \mathbf{x}$. This happens when

$$\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{x}}} \right) = 0$$

This is called the Euler-Lagrange equation.

These ideas are the beginnings of a rich version of mechanics in a general sense. The basic idea of what is called Lagrangian mechanics is that we have a quantity, $L(\mathbf{x}, \dot{\mathbf{x}}, t)$, that depends on local properties of the trajectory. From L we define an action, $A = \int L dt$, where the integral is taken along a particular trajectory. The value of the action is a function of the trajectory and, in nature, the trajectory for which the functional derivative vanishes is the one followed. The condition for this is written as $\delta A/\delta \mathbf{x} = 0$ to stress that the functional derivative is intended. This gives rise to the equation of motion

$$\frac{\partial L}{\partial \mathbf{x}} - \frac{d\mathbf{p}}{dt} = \mathbf{0} \tag{29}$$

where $\mathbf{p} = \partial L/\partial \dot{\mathbf{x}}$ is identified with the momentum. As you must already suspect from this brief glimpse of advanced mechanics, there is a lot of method in this madness and we shall try to relate it to what is to come.

General Relativity

Installment II.

September 8, 2003

1 The Story So Far

From Galileo we know that physical behavior is unaffected by uniform motion. The image of a vehicle moving at constant velocity is often used to exemplify this physical law. Galileo liked ships for this; Einstein took trains for the mental voyage, which he called a gedanken (=thought) experiment.

Observer G is on the ground, not moving, and watching a train going by in heir positive x direction. (We pretend that the earth does not move.) A second observer, call heir T, is going by in a box car. T sets up a rectangular coordinate system with heir x' axis lined up with G's x axis. These being classical physicists, so we assume, with them, that they both have the same times in their two systems. That is,

$$t' = t, \tag{1}$$

which is very nearly true in daily life. This would once have been thought so obvious as to not need mention. Yet, the very fact that we feel the need to mention in a discussion of this kind is already a departure from the classical imagery.

At time t', T observes a fly in heir coordinate system at x'. At time $t = t'$, G observes the fly at coordinate x in heir system. What is the relation between the two measurements of position? (There are some minor effects that might complicate the study of this question but we may neglect them in a thought experiment. One is the tiny travel time of light between the two systems. Can you think of any others?) Back in the good old days (before March 14, 1879, say), we would have said that

$$x' = x + vt \tag{2}$$

Let $f(x)$ be a force in the x direction in G's system. Then Newton's second law says that, for an object of mass m,

$$m\frac{d^2x}{dt^2} = f(x). \tag{3}$$

We may use (2) to see what form the law takes in T's system. From (2) we see that $f(x) = f(x' - vt)$. We introduce a new function f' with the property that $f'(x') = f(x' - vt) = f(x)$. The new function is the force at the point x but expressed in the primed coordinate. A function with this property is called a (Galilean) scalar. We then see that

$$m \frac{d^2 x'}{dt'^2} = f'(x'), \tag{4}$$

The form of the law is unchanged by uniform motion as just Galileo wrote This invariance of the form of the law (sometimes called covariance) under the transformation given by (1)-(2) is not true for Maxwell's equations of electromagnetism, which are covariant under a different transformation called the Lorentz transformation. Let us pause to look briefly at this transformation even though it is a bit early for this.

2 The Lorentz Transformation

Since we need to specify both position and time to identify an event such as a meeting with someone, we should not be surprised that the time of the event would be treated on the same footing as the place. If we keep on thinking for now of a space of one dimension we then have to give simply t and x to make a date. But if the person we are meeting is not on the uniformly moving train with us, we might want to tell hem t' and x', which we could find from the Galilean transformation laws, (1) and (2). As we saw, the form of Newton's laws of motion are not modified by this transformation and, as these are quite good laws, we might happily adopt the Galilean transformation in physics. But Lorentz (and Poincaré) found that invariance of form under that transformation did not apply to Maxwell's equations (which we shall write down at some point). Suffice it to say for now that Maxwell's equations do not change form when, instead of the Galilean transformation rules, we use the transformation

$$x' = \gamma(x + vt) \tag{5}$$

$$t' = \gamma(t + \frac{vx}{c^2}) \tag{6}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{7}$$

and c is the speed of light in a vacuum, that is, 3×10^{10} cm/sec.

This difference in preferences for describing how coordinates in spacetime change when we go into a moving frame seems to pose a problem of choice, but it is not so grievous as it might first appear. Light goes $300,000$ km/sec — several circumferences of the earth — in one second. In our daily experience, including our studies of planetary motion, nothing is going that fast. If $|v| << c$, then $\gamma \approx 1$ and (5)-(6) are practically indistinguishable from (1) and (2) so all is well many aspects of daily life, including most mechanical problems. However, when things are moving at high enough speeds, the two transformation laws differ markedly and there then is a real choice to be made: which of the two laws of transformation is the right one? Or is there something more fundamental in our thinking that has gone agleigh? Perhaps the physical laws do not remain the same when we go into a uniformaly moving system?

In 1905, Einstein argued for the validity of the Lorentz transformation and, to deal with the resulting dilemma, he showed how the laws of mechanics needed to be modified when the bodies in question move at high speed. We will discuss all that and look further into these transformation laws after we have studied some vector analysis by way of preparation. This will be our standard procedure — a little physics interspersed with a little mathematical discussion. This is the sort of trick used by novelists in which one jumps from place to place and protagonist to protagonist in successive chapters till, at the end, it all comes nicely together, with any luck. Such weaving together of several strands in the narrative is the plan we shall follow here. But please take a few moments now to answer the following question.

You know how to find x' and t' given x and t from (5)-(6) or (5)-(6). Suppose you are given x' and t'. What are the transformation laws that tell you x and t for both the Galilean and Lorentzian transformations? Do these inverse transformations sound reasonable? Explain that (to yourselves).

But now, let us come to the real reason that we have opened the issue of tranformation laws at this time. For this we need to look at Figure 1, a plot of the coordinate system in spacetime that we shall often use. In daily life, we use different units for space and for time, even though we now want to consider time as a coordinate in our world, our own spacetime. This turns out to be more convenient for visualizing what is happening in relativistic matters if we use the same units for space and time. That is why some people use light years, for example. So let us fix the problem using c, which appears to be a constant of Nature and has the dimension of length/time. We can use either x/c as our space coordinate, which then is expressed in time units (as in light years), or we can take ct to mark time. Either way, the value of c becomes unity in the new units, so that takes care of that, until you need to express things in everyday units (And even that is not difficult.)

Because $c = 1$ in these natural units, Figure 1 looks quite simple. The t and x axes are those of observer G. They are at right angles to each other since nothing indicates that things need to be otherwise and G wisely chose them this way. The choice of t for the vertical axis is the general preference among relativists and is adopted here. There are two lines at 45° in the figure and these represent light particles, or photons, one moving to the right at speed c ($= 1$) and one to the left at speed c. (Quantum mechanics is being ignored here y there are no worries about how we can know where the photon is when we also know its precise velocity.) Had we not adjusted the units, those two photon paths through spacetime — their worldlines — would be indistinguishable from the t axis on the scale of this plot. Note also that the worldline of the peripatetic T on heir train is shown in the figure going to the right at a respectable clip.

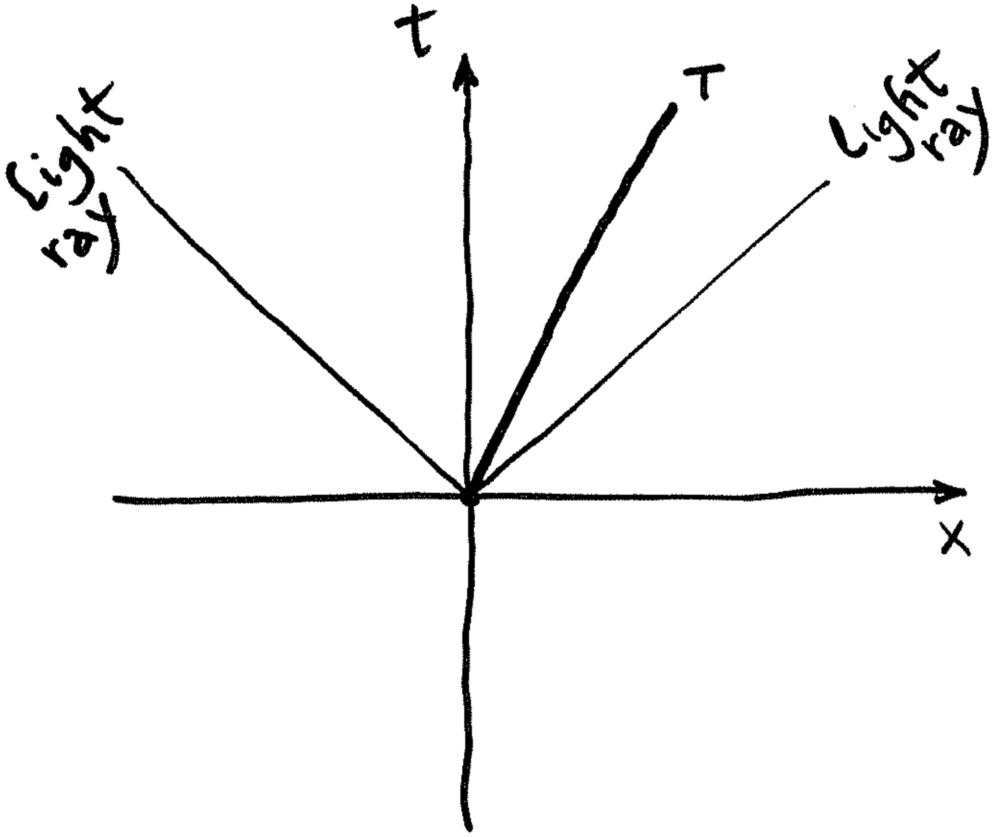

Figure 1.

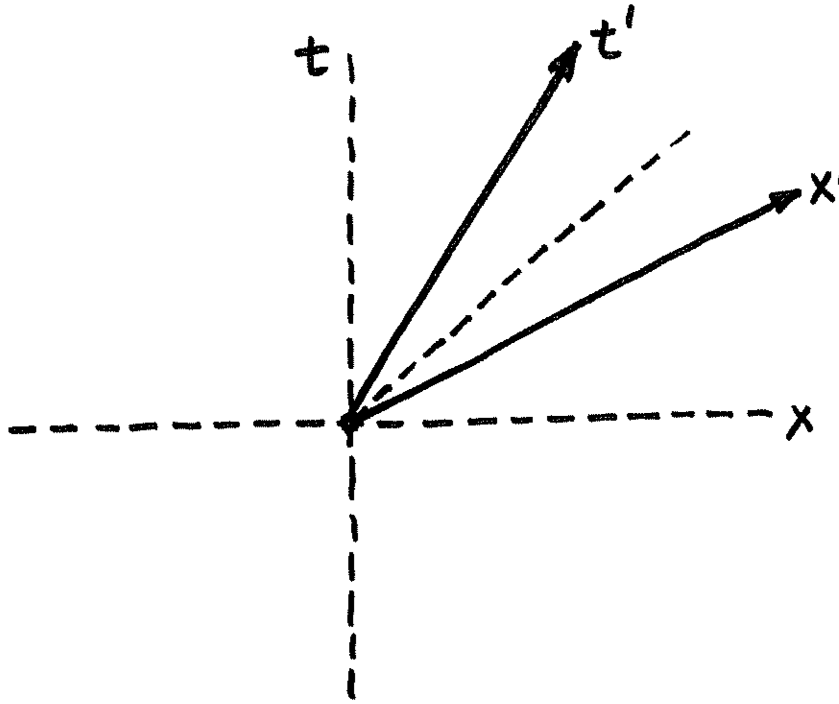

Figure 2.

In the coordinate system shown, G is not moving in space but is moving in time, that is, vertically in this coordinate system. How shall we draw the coordinate system of T? Naturally, we shall assume that like G, T in heir reference frame is moving in the direction of heir time axis. That means that the world line of T coincides with the t' axis. But where shall we place the x' axis?

We could look at the Lorentz transformation for guidance in drawing the x' axis, but let us try working from Figure 1 first. There are two features that seem significant there. The t and x axes are orthogonal to each other and the worldlines of light particles bisect the angles between the axes because light is traveling with respect to G at the speed c (= 1, in natural units). The orthogonality is just a matter of convenience, but the speed of light is apparently a constant of nature and it represents some important physics. So let us assume that light moves at speed $c = 1$ for observer T as well as for G. If you ask why, you will not be alone. One answer is that this is what follows from the Lorentz transformation. It also follows from some closely related assumptions made by Einstein in 1905, but in fact he was just trying to rationalize some features of electromagnetic theory. So the real answer to the question is that the world is made that way and the miracle is that Maxwell figured out the right equations for electromagnetism. It is no wonder that, on studying Maxwell's equations, Ludwig Boltzmann wrote "Was it a god who wrote these lines?' But as we shall see, these features of space are really an expression of its geometric structure.

If the speed of the photons is to be $c = 1$ in system T, the light ray must bisect the angle between the t' and x' axes (as you should verify). This is shown in Figure 2 where the coordinate axes of T are drawn in the system of G. What is striking is that the new axes are not orthogonal. This shows at least one case where oblique coordinates arise in a somewhat natural way and, as you can imagine, they can be forced on us for other reasons when we get into complicated geometrical situations. Let us then consider vector analysis in oblique coordinate systems before going on with the physics problems. In doing that we shall begin to learn the vectorial notation that will be needed in studying relativity, so buckle down for a longish haul.

3 Vector Analysis

3.1 Arrows

For now, let us stay in Euclidean space. The number of dimensions can be anything, if it is not too large, but the illustrations we use will be mainly in the (Euclidean) plane, to which we are confined by inadequate skill at drawing. For our present purposes, a vector may be represented as an arrow in the space. This convenience is going to be denied us once we get the hang of the notation and go to more general spaces, so it might be well to prepare for this eventuality by beginning to think about vectors more abstractly. A reading of the book on vector spaces by Halmos, for example, would make for nice bedtime reading for the serious student.

To denote vectors, let us use boldface Latin letters. As to pictures, the most vivid illustration of a vector is an arrow representing a displacement; the length of the arrow is the displacement distance and the arrow points in the direction of the displacement. Suppose we are taken from point P to point Q by a displacement $\mathbf{a}$ and then to a third point (that shall remain nameless) by another displacement $\mathbf{b}$. The total displacement is by a vector $\mathbf{c}$ which is the sum of $\mathbf{a}$ and $\mathbf{b}$ as illustrated in Figure 3 by the parallelogram law. What could be simpler? Certainly not the abstract statement of all this, which we do not attempt here. What you see though is what a luxury it is to draw the vector in the space in which we are using it.

Subtraction of a vector, as in $\mathbf{a}-\mathbf{b}$, is effected by adding the negative of the subtracted vector. The negative of a vector is obtained by reversing the direction of the arrow representing it. This is done algebraically by introducing the notion of multiplication of a vector by a real number. So suppose that $|\mathbf{a}|$ is the length of $\mathbf{a}$. Then, if q is a positive real number, $q\mathbf{a}$ is a vector in the same direction as $\mathbf{a}$ with length $q|\mathbf{a}|$. If q is a negative real number, $q\mathbf{a}$ is a vector in the direction opposite to that of $\mathbf{a}$ and with length $-q|\mathbf{a}|$.

This summary would be inadequate if you did not know all this already. So if you do not know it all quite well, immediate remediation is needed. This can be obtained by reading an elementary book on vectors or asking a helpful friend or instructor.

3.2 Notation

If we are in a space of dimension n, we need n linearly independent vectors to make a basis for representing any vector in the space. For our purposes, by saying two vectors are linearly independent we mean that they are not colinear. Let us denote these non-colinear basis vectors as $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. A concise way to write this is by means of what is called the *range convention*. In that case we simply denote the collection of basis vectors as $\mathbf{e}_i$ where it is understood that i ranges over the values $1, 2, \ldots, n$. If then $\mathbf{e}_i$ is a basis, we may represent any vector $\mathbf{a}$ as a linear combination of the $\mathbf{e}_i$. That is, we may write that

$$\mathbf{a} = a^1\mathbf{e}_1 + a^2\mathbf{e}_2 + \ldots + a^n\mathbf{e}_n \quad (8)$$

The numbers $a^1, a^2, \ldots, a^n$ are called the components of $\mathbf{a}$ and we may write them as a^i in view of the range convention. (It is understood that $i = 1, 2, \ldots, n$.)

People new to this way of writing things often ask why they are written this way. Many are especially bothered by the fact that the index on the components is written as a superscript where it may be confused with an exponent. If you happen to know two people with the same given names, do you not manage to keep things straight? I myself know of a few different people called Engelbert, but I know which is which. Anyway, it is all a matter of

convention. (You may recall this exchange from the 'Hunting of the Snark,' where there is much wisdom to be found,
What's the good of Mercator's
North Poles and equators,
Tropical zones and meridian lines?
So the Bellman would cry.
And the crew would reply,
They are merely conventional signs.)

Another way to write (8) is

$$\mathbf{a} = \sum_{i=1}^{n} a^i \mathbf{e}_i \ , \tag{9}$$

which has a nice feel to it. But an even more convenient version of this is provided by the *summation convention* that states: whenever a suffix is repeated with one index up and one down it is understood that this suffix is to be summed over its full range. Thus, on adopting this convention, we may write (9) as

$$\mathbf{a} = a^i \mathbf{e}_i \ . \tag{10}$$

Once you have used this notation for a while, it will become second nature to you.

3.3 Inner Product

As we are in Euclidean space, we may use such old standards as the inner, or dot, product. If θ is the angle between two vectors $\mathbf{a}$ and $\mathbf{b}$, then we may write

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta . \tag{11}$$

The dot product of two orthogonal vectors is then zero and the dot product of a vector with itself is the square of its length. All this is the situation in the wondrous space of Euclid. Of course, this plot will thicken, but let us see how things look when we use the notation that we have just introduced.

Assume that we have a basis $\mathbf{e}_i$ and two vectors $\mathbf{a} = a^i \mathbf{e}_i$ and $\mathbf{b} = b^j \mathbf{e}_j$. (Here i and j are dummy indices since they are summed over.) Then

$$\mathbf{a} \cdot \mathbf{b} = a^i b^j \mathbf{e}_i \cdot \mathbf{e}_j \ . \tag{12}$$

If the coordinate axes *were* chosen to be mutually orthogonal and the basis vectors had unit length, we would have

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \tag{13}$$

where δ_{ij} is called the Kronecker delta; $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$. The Kronecker delta is our first instance of an object with two indices; both of them are subject to the range convention and the summation convention. Also, this delta is symmetric in its indices: $\delta_{ij} = \delta ji$. We may then write

$$\mathbf{a} \cdot \mathbf{b} = a^i b^j \delta_{ij} \tag{14}$$

where a double sum is implied.

Let

$$a_i = a^j \delta_{ji} \tag{15}$$

On the left of this definition we have one lower free index, i. On the right, we have the *same* free index and the dummy index j (repeated). The grammar is that the same free indices must always appear on each side of an equality and this rule is very helpful in locating errors. So is the taboo on having and index appear more than twice in an expression. The process indicated in (15) is called lowering the index and it gives us a new set of numbers a_i with the index downstairs. Of course, because of the simple nature of the delta, the two sets of components (for that is what they are) are equal for the same index but this is a feature of this special case — Euclidean space with orthonormal basis vectors.

If you find this notation new, it may be helpful to relate it to one you may have seen before. (Even if you have not seen it before, it may be helpful.) Let us represent a^i as a vertical column of numbers, $a^1, a^2, \ldots$. And let us represent a_i as a horizontal row of numbers $a_1, a_2, \ldots$. In other words, our vector $\mathbf{a}$ has two kinds of components, though the distinction is not really needed (except perhaps conceptually) when we deal with rectangular Cartesian coordinates and it is usually not even mentioned in that case. Indeed, the two sets of components are in fact equal at this stage and we may write, for $n = 2$, that

$$a^i = \begin{pmatrix} a^1 \\ a^2 \end{pmatrix} ; \qquad a_i = \begin{pmatrix} a_1 & a_2 \end{pmatrix} . \tag{16}$$

Then, for example,

$$\mathbf{a} \cdot \mathbf{b} = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} a^1 \\ a^2 \end{pmatrix} = a_1 b^1 + a_2 b^2 . \tag{17}$$

This form of expressing the dot product is not quite the same as the one we shall be using since here the order of the vectors matters in this notation. (Those who know quantum mechanics will see an analogy also to the bra and ket notation of that subject.)

3.4 Oblique Coordinates

When we go through the same exercise that we just performed, but with oblique coordinate axes, some new aspects of vector analysis emerge. In Figure 4, we have a sketch of a two-dimensional situation with two coordinates whose axes are at an oblique angle to one another. Also shown is an arrow representing a vector $\mathbf{a}$ and two short arrows along the coordinate axes indicating the basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. We wish to write $\mathbf{a} = a^i \mathbf{e}_i$ but what is the geometric significance of the components? That is, how do we project the vector $\mathbf{a}$ onto the coordinate axes? In the figure, the standard projection is shown in which lines from the head of the arrow are drawn parallel to the axes and these cut the axes to give us the two vectors $a^1\mathbf{e}_1$ and $a^2\mathbf{e}_2$.

We may also compute the inner product of two vectors, $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{a}\cdot\mathbf{b} = a^i b^j \mathbf{e}_i \cdot \mathbf{e}_j \ . \tag{18}$$

Now let

$$g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j \ . \tag{19}$$

Then the inner product is written as

$$\mathbf{a}\cdot\mathbf{b} = a^i b^j g_{ij} \tag{20}$$

We see that, in the case of oblique coordinates, the situation resembles that with rectangular coordinates but that the Kronecker delta is replaced by g_{ij} (which is the same as g_{ji}). In this relatively simple situation, we can write δ_{ij} and g_{ij} as two-dimensional arrays of numbers, that is, as matrices. The latter case is a touch more complicated, but it is still quite computable and you should compute the individual values of g_{ij} if you haven't already visualized them. More than this, we may also lower the index on a^i:

$$a_j = a^i g_{ij} \ . \tag{21}$$

Here, the indices were written a little differently than before to give you an idea how things work, but the rules for their manipulation are the same. What is really different now is that the components a_i are not the same as the a^i and this points to an alternative mode of projection onto coordinate axes in the case of oblique coordinates.

First let us observe that we can write g_{ij} as an array as in

$$g_{ij} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \ . \tag{22}$$

In the present case, we have $g_{11} = g_{22} = 1$ and $g_{12} = g_{21} = \cos\vartheta$ where ϑ is the angle between the (oblique) coordinate axes. A little trigonometry then reveals that if we drop perpendiculars from the head of the arrow of $\mathbf{a}$ onto the coordinate axes we obtain segments on the axes whose lengths are given by $a_i = g_{ij}a^j$. This alternate way of generating components is not so meaningful geometrically. A more fruitful approach is obtained by introducing the so-called reciprocal basis.

We first introduce the reciprocal matrix g^{ij} defined so that

$$g_{ij}g^{jk} = \delta_i^{\ k} \tag{23}$$

where

$$\delta_i^{\ k} = \begin{cases} 1 & \text{for} \quad i = j \\ 2 & \text{for} \quad i \neq j \end{cases} \ . \tag{24}$$

Now we note that

$$g^{ij}a_j = g^{ij}g_{jk}a^k = \delta_k^{\ i}a^k = a^i \ . \tag{25}$$

Just as we used g_{ij} to lower an index, we can use its reciprocal, g^{ij}, to raise an index.

The main aim just now is to get the hang of these manipulations by poring over them. But do notice that

$$\mathbf{a} = a^i\mathbf{e}_i = \delta_i^{\ j}a^i\mathbf{e}_j = g_{i\ell}g^{\ell j}a^i\mathbf{e}_j = \left(g_{\ell i}a^i\right)\left(g^{\ell j}\mathbf{e}_j\right) \ . \tag{26}$$

This suggests that we may also write

$$\mathbf{a} = a_i\mathbf{e}^i \tag{27}$$

where

$$\mathbf{e}^i = g^{ij}\mathbf{e}_j \ . \tag{28}$$

There are as many of the $\mathbf{e}^i$ as of the $\mathbf{e}_j$ and they form what is known as the reciprocal basis since

$$\mathbf{e}^i \cdot \mathbf{e}^j = \left(g^{ik}\mathbf{e}_k\right) \cdot \left(g^{j\ell}\mathbf{e}_\ell\right) = g^{ik}g^{j\ell}g_{k\ell} = g^{ik}\delta_k^{\ j} = g^{ij} \ . \tag{29}$$

The last equality in this string is a definition really, but it demands to be made. We also note that

$$\mathbf{e}_i \cdot \mathbf{e}^j = g_{ik}\mathbf{e}^k \cdot \mathbf{e}^j = g_{ik}g^{kj} = \delta_i^{\ j} \ . \tag{30}$$

From this we conclude that $\mathbf{e}_i \perp \mathbf{e}^j$ for $i \neq j$ and that $\mathbf{e}_i \cdot \mathbf{e}^j = 1$ for $i = j$. The projection of $\mathbf{a}$ onto the reciprocal axes is illustrated in Figure 5. Two forms of projection onto oblique axes thus gives us two sets of bases (dual to each other, one sometimes says) with two sets of coordinates. All this prefigures some important geometry that we must try to appreciate on our way to the theory of gravity.

Before we go on though, there is one last point to be made about this notation. In the Euclidean space, the position vector of a point (the fly of the example) is

$$\mathbf{x} = x\mathbf{e}_1 + y\mathbf{e}_2 + \dots \ . \tag{31}$$

This expression does not take advantage of the range and summation conventions. So we let $x^1 = x, x^2 = y, \dots$ and now we can write

$$\mathbf{x} = x^i\mathbf{e}_i \ . \tag{32}$$

The square of the distance of the point from the origin is then

$$\mathbf{x}^2 = \mathbf{x} \cdot \mathbf{x} = g_{ij}x^ix^j = (x^1)^2 + (x^2)^2 + 2x^1x^2\cos\varphi \ , \tag{33}$$

as it ought to be.

General Relativity

Installment III.

September 15, 2003

1 Vector Analysis Continued

3.4 Outer Product When working in three-dimensional Euclidean space we may define an outer, or cross, product of two vectors $\mathbf{a}$ and $\mathbf{b}$ written as $\mathbf{a} \times \mathbf{b}$ such that

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| \, |\mathbf{b}| \cos \varphi \tag{1}$$

where φ is the angle between the two vectors when they are referred to the same origin. (Remember that we can slide vectors around at will in Euclidean space.) The product is taken as a vector in the line perpendicular to each of $\mathbf{a}$ and $\mathbf{b}$ and its direction is that given by the right hand screw rule when $\mathbf{a}$ is turned into $\mathbf{b}$. It follows that

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a} \; . \tag{2}$$

If $\mathbf{a} = a^i \mathbf{e}_i$ and $\mathbf{b} = b^j \mathbf{e}_j$, then

$$\mathbf{a} \times \mathbf{b} = a^i b^j \, \mathbf{e}_i \times \mathbf{e}_j \tag{3}$$

But $\mathbf{e}_i \times \mathbf{e}_j$ is perpendicular to each of $\mathbf{e}_i$ and $\mathbf{e}_j$ and so is $\mathbf{e}^k$ when $k \neq j$ and $k \neq i$. Therefore we may express $\mathbf{a} \times \mathbf{b}$ as a linear combination of the reciprocal basis vectors $\mathbf{e}^k$ This is convenient but it does convey a feeling that there is something that distinguishes cross products from conventional vectors.

The coefficients in the linear combination of reciprocal vectors that gives us the cross product need to be determined but we may write the combination in an abstract way as

$$\mathbf{e}_i \times \mathbf{e}_j = A_{ijk} \mathbf{e}^k \tag{4}$$

where the coefficients A_{ijk} are as yet unknown. (We need the three indices since one is summed on and two are free to keep track of the two vectors in the product.) We dot $\mathbf{e}_\ell$ into this formula and recall that $\mathbf{e}_\ell \cdot \mathbf{e}^k = \delta_\ell^{\cdot k}$. Since $A_{ijk} \delta_\ell^{\cdot k} = A_{jk\ell}$ we find that

$$A_{ij\ell} = \mathbf{e}_\ell \cdot \mathbf{e}_i \times \mathbf{e}_j \tag{5}$$

When we put this into (4), we get

$$\mathbf{e}_i \times \mathbf{e}_j = [\mathbf{e}_k\, \mathbf{e}_i\, \mathbf{e}_j]\mathbf{e}^k \tag{6}$$

where

$$[\mathbf{e}_k\, \mathbf{e}_i\, \mathbf{e}_j] = \mathbf{e}_k \cdot \mathbf{e}_i \times \mathbf{e}_j \tag{7}$$

The three vectors $\mathbf{e}_i$, with $i = 1, 2, 3$, form a parallelpiped whose volume is $[\mathbf{e}_k\, \mathbf{e}_i\, \mathbf{e}_j]$ and the value of the triple product is unchanged by cyclic permutation of the indices. A permutation that is odd (noncyclic) merely changes the sign. (This sort of thing is discussed in older books on the mathematics of physics such as Margenau and Murphy.)

Let us introduce $v = [\mathbf{e}_k\, \mathbf{e}_i\, \mathbf{e}_j]$. Then we may write (6) out as

$$\mathbf{e}^1 = \frac{\mathbf{e}_2 \times \mathbf{e}_3}{v}, \qquad \text{etc.} \tag{8}$$

where the other components are found by cyclic permutation. To write this in a convenient way we introduce the permutation symbol ϵ^{ijk} which equals zero if any two of the indices are the same, equals one if the order of the indices is a cyclic permutation of 123 and equals minus one if the indices are a cyclic permutation of 132. Then (8) is compressed into

$$\mathbf{e}^i = \epsilon^{ijk}\,(\mathbf{e}_j \times \mathbf{e}_k)/(2v)\ . \tag{9}$$

Let us now write (3) as

$$\mathbf{a} \times \mathbf{b} = a^2 b^3\, \mathbf{e}_2 \times \mathbf{e}_3 + a^3 b^2\, \mathbf{e}_3 \times \mathbf{e}_2 + ... = (a^2 b^3 - a^3 b^2)\, \mathbf{e}_2 \times \mathbf{e}_3 + ... \tag{10}$$

Now we write out all the terms and use (8) to obtain

$$[\mathbf{e}_1\, \mathbf{e}_2\, \mathbf{e}_3]\, \mathbf{a} \times \mathbf{b} = (a^2 b^3 - a^3 b^2)\, \mathbf{e}^1 + (a^3 b^1 - a^1 b^3)\, \mathbf{e}^2 + (a^1 b^2 - a^2 b^1)\, \mathbf{e}^3, \tag{11}$$

which begins to resemble the familiar way of writing out the cross product. To complete the story we need to look a bit more at $[\mathbf{e}_k\, \mathbf{e}_i\, \mathbf{e}_j]$. This is a bit of manipulation that you do not really need to get into, but I'll leave it as an exercise for now and sketch it in a later installment for those who may be interested.

2 Fictitious Forces

2.1 Relativity

The statement by Galileo (Installment I) about the goings on in the hold of a ship being indifferent to the imposition of uniform motion of the ship itself may have surprised people in 1600. Nowadays, we are not so surprised by it since we encounter the phenomenon readily, as in planes moving in tranquil air. But the implication of Galileo's imagery are striking since it tells us that if two observers are moving with respect to each other at contsant velocity, neither one being accelerated, each of them has the right to consider heirself to be at rest. (As we shall see, there are developments in modern cosmology that alter that vision, but by the time we get to them, we shall have seen how to absorb their implication into our thinking.) Therefore, each of the relatively moving observers might wish to claim that the correct statement of the laws of mechanics would be with respect to, that is, *relative* to, heir own frame of reference. This form of relativity is called special since it presumes that the two observers move with respect to each other at constant velocity. (In other languages, the adjective used is more like *restricted*, but in English, the word special is used.) We shall be generalizing those notions and by way of preparation, let us begin to see how things look to accelerated observers.

2.2 Rotating Frames

Suppose that we are in a frame rotating with angular speed Ω, as indeed we are. The rotation may be described about an axis and this permits us to speak of an angular velocity $\mathbf{\Omega}$, where this vector is aligned with the axis of rotation. We would like then to set up a coordinate system rotating with the system. If we take the earth as our example, we should note that the earth's rotation rate is not constant, though it does vary only very slowly.

Let us use the term inertial frame to signify a frame that is not rotating. This in itself may raise questions in your mind, but let us postpone the issue of *precisely* what is meant by the inertial frame and settle for the frame of someone fixed in space. We write the basis vectors in the inertial frame as $\mathbf{e}_i$ and vector $\mathbf{a}$ is then written as usual as $a^i\mathbf{e}_i$.

Let us denote the basis vectors in the rotating frame as $\hat{\mathbf{e}}_\mathbf{i}$. (You may wonder why we do not use a prime to denote this second set of quantities to be consistent with the notation used earlier. We'll go into that later.) The vector $\mathbf{a}$ itself is expressed as $\hat{a}^i\,\hat{\mathbf{e}}_\mathbf{i}$ where $\hat{a}^i$ are the components in the rotating frame.

In the inertial frame, the basis vectors are fixed, so that

$$\frac{d\mathbf{a}}{dt} = \frac{da^i}{dt}\mathbf{e}_i \tag{12}$$

By contrast, the derivative of $\mathbf{a}$ in the rotating frame is

$$\frac{d\mathbf{a}}{dt} = \frac{d\hat{a}^i}{dt}\hat{\mathbf{e}}_i + \hat{a}^i\frac{d\hat{\mathbf{e}}_\mathbf{i}}{dt} \tag{13}$$

The qualitative difference between the two statements is that the basis vectors in the rotating frame, as seen in the inertial frame from which we are presumed to start out, are varying. We need to decide how the basis vectors in the rotating frame are varying.

When a short time δt goes by, the system rotates by a small amount $\Omega\delta t$. This causes each of the basis vectors to turn by a small amount without any change in length. The change in each basis vector is represented by the addition of $\delta\hat{\mathbf{e}}_\mathbf{i}$. Since a rotation does not change the length of a vector, this slight change in basis vector must be perpendicular to the basis vector itself and to $\boldsymbol{\Omega}$. We then conclude that the direction of $\delta\hat{\mathbf{e}}_\mathbf{i}$ is in the direction of $\boldsymbol{\Omega}\times\hat{\mathbf{e}}_\mathbf{i}$. The sign of this correction is positive in the direction obtained by turning $\boldsymbol{\Omega}$ into $\hat{\mathbf{e}}_\mathbf{i}$; that is in fact how the direction of Ω is chosen. The amount of the correction is proportional to $\Omega\delta t$. It is also proportional to the component of $\hat{\mathbf{e}}_\mathbf{i}$ perpendicular to $\boldsymbol{\Omega}$, which varies with the sine of the angle between $\hat{\mathbf{e}}_\mathbf{i}$ and $\boldsymbol{\Omega}$ — just what we get from the cross product between the two vectors. Hence we have

$$\frac{d\hat{\mathbf{e}}_\mathbf{i}}{dt} = \boldsymbol{\Omega}\times\hat{\mathbf{e}}_\mathbf{i} \tag{14}$$

and so

$$\frac{d\mathbf{a}}{dt} = \frac{d\hat{a}^i}{dt}\hat{\mathbf{e}}_i + \boldsymbol{\Omega}\times\mathbf{a} \tag{15}$$

To clarify these things one often writes $D_\mathbf{I}$ to denote the time derivative in the inertial frame. That is, $D_\mathbf{I}$ is the time derivative holding the $\mathbf{e}_\mathbf{i}$ fixed:

$$D_\mathbf{I} = \frac{da^i}{dt}\mathbf{e}_i \tag{16}$$

Then, we write $D_\mathbf{R}$ for the time derivative in the rotating frame, that is, the derivative holding $\hat{\mathbf{e}}_\mathbf{i}$ fixed:

$$D_\mathbf{R} = \frac{d\hat{a}^i}{dt}\hat{\mathbf{e}}_\mathbf{i} \tag{17}$$

As we see from (15), we have

$$D_\mathbf{I}\mathbf{a} = D_\mathbf{R}\mathbf{a} + \boldsymbol{\Omega}\times\mathbf{a} \tag{18}$$

Another way that people have of saying this is that when you transform into a rotating reference frame you should transform the time derivative as in

$$\frac{d}{dt} \to \frac{d}{dt} + \boldsymbol{\Omega}\times \tag{19}$$

Either way you look at this, the first interesting question to ask is what happens to the derivative of the position vector, $\mathbf{x}$? We find that

$$D_\mathbf{I}\mathbf{x} = D_\mathbf{R}\mathbf{x} + \boldsymbol{\Omega}\times\mathbf{x} \tag{20}$$

This says that

$$\mathbf{v}_\mathbf{I} = \mathbf{v}_\mathbf{R} + \boldsymbol{\Omega}\times\mathbf{x} \tag{21}$$

where $\mathbf{v}_\mathbf{I}$ and $\mathbf{v}_\mathbf{R}$ are the velocities relative to the two frames. So this shows how the velocity transforms when we go to a rotating frame. Even if you are stationary in the rotating frame, you will have a velocity $\boldsymbol{\Omega}\times\mathbf{x}$ with respect to the inertial frame. Since that velocity is not a constant in time, you will feel an acceleration by virtue of being in the (accelerating!)

rotating frame. If you happen to be on the surface of the earth, this motion will carry you in a direction tangent to your latitude circle along with the rest of us.

Now we consider what happens to the acceleration felt in the inertial frame when we go into the rotating frame. The inertial acceleration is $D_{\mathbf{I}}\mathbf{v_I}$. If we put $\mathbf{a} = \mathbf{v_I}$ into (18) we have

$$D_{\mathbf{I}}\mathbf{v_I} = D_{\mathbf{R}}\mathbf{v_I} + \boldsymbol{\Omega} \times \mathbf{a} \tag{22}$$

When we put (21) into this, we get

$$D_{\mathbf{I}}\mathbf{v_I} = D_{\mathbf{R}}(\mathbf{v_R} + \boldsymbol{\Omega} \times \mathbf{x}) + \boldsymbol{\Omega} \times (\mathbf{v_R} + \boldsymbol{\Omega} \times \mathbf{x}) \tag{23}$$

The acceleration with respect to the rotating frame is $\mathbf{a_R} = D_{\mathbf{R}}\mathbf{v_R}$ so we may write that the inertial acceleration $\mathbf{a_I}$ is given by

$$\mathbf{a_I} = \mathbf{a_R} + 2\boldsymbol{\Omega} \times \mathbf{v_R} + \dot{\boldsymbol{\Omega}} \times \mathbf{x} + \boldsymbol{\Omega} \times \boldsymbol{\Omega} \times \mathbf{x} \tag{24}$$

where we have written $\dot{\boldsymbol{\Omega}}$ for $D_{\mathbf{R}}\boldsymbol{\Omega}$. In the triple product we simplify by noting that $\boldsymbol{\Omega} \times \mathbf{x}$ is perpendicular to $\boldsymbol{\Omega}$ and that it has magnitude $\Omega|\mathbf{x}|\sin\varphi$ where φ is the angle between $\boldsymbol{\Omega}$ and $\mathbf{x}$. (If you are at position $\mathbf{x}$ on earth, then φ is your colatitude.) Also, since $\boldsymbol{\Omega} \times \mathbf{x}$ is perpendicular to $\boldsymbol{\Omega}$, the triple product has magnitude $\Omega^2\varpi$ where $|\varpi|\,|\mathbf{x}|\sin\varphi$ in a notation that some astronomers like. If you chase down the right-hand rule twice you see that the triple product is a vector perpendicular to $\boldsymbol{\Omega}$ pointing away from where $\mathbf{x}$ is. So if the vector ϖ is perpendicular to the rotation axis and pointing toward $\mathbf{x}$, the acceleration formula is

$$\mathbf{a_I} = \mathbf{a_R} + 2\boldsymbol{\Omega} \times \mathbf{v_R} + \dot{\boldsymbol{\Omega}} \times \mathbf{x} - \Omega^2\varpi \tag{25}$$

(The ϖ in this formula should be in boldface, but I cannot make it happen.)

Suppose that the inertial observer feels no force; heir acceleration is then zero. Nevertheless, the observer in the rotating frame will feel an acceleration given by

$$\mathbf{a_R} = -2\boldsymbol{\Omega} \times \mathbf{v_R} - \dot{\boldsymbol{\Omega}} \times \mathbf{x} + \Omega^2\varpi \tag{26}$$

Judging from the form of Newton's second law, these terms on the right might be called forces. However since they arise as a result of going into an accelerating frame of reference, the are sometimes called fictitious forces, even though you will definitely feel them in a rotating frame, as you may know from having been a whirling platform of some sort. The third term on the right is called the centrifugal force and it comes from the inertial tendency to go in a straight line when this tendency is frustrated by whatever is keeping you in the rotating frame. (This was the message of the parable of the bead on the wire.)

The middle term on the right of (26) is called the Euler force. Since the earth's rotation is being slowed down as a result of tidal dissipation, there is such a force and it drives a very weak circulation in the earth's interior, according to some calculations. The first term on the right is called the Coriolis force and it plays a great role in meteorology and oceanography.

If you move over the surface of the earth, the Coriolis force is perpendicular to your velocity. This is like the Lorentz force felt by an electron moving through a magnetic field. When you look at weather maps you see that the contours of pressure are quite wavy. The Coriolis force has a lot to do with those waves. (Coriolis, by the way, was very interested in rotation and even wrote a book on billiards.)

For our purposes, the most striking thing in (26) is that if you want to think of it as a form of Newton's second law, you should multiply by the mass of the moving body. When you do that., all the terms on the right are also multiplied by the mass. So the mass cancels from this form of Newton's laws because fictitious forces are always proportional to the mass of the object suffering acceleration.

The acceleration produced by a fictitious force is independent of the mass of the object acted upon. It is therefore worth recalling that Galileo found that falling bodies had accelerations independent of their masses. Once you do that, you see what Einstein's notion of gravity was. All we need to do now is to see how he implemented it.

General Relativity

Installment IV.

September 17, 2003

1 Fictitious Forces Continued

1.1 An Expanding Medium

The universe is expanding (at least locally), so this is an especially interesting phenomenon for us. In this brief look, we shall study the classical version using Newton's gravitational force. A discussion of Newtonian cosmology can be found in H. Bondi's 'Cosmology.'

We select an origin of coordinates and consider an object — say a galaxy — at position $\mathbf{r}$ with respect to that origin. Assume that the expansion is uniform so that, apart from possible disturbances in that uniform expansion, all distances are continuously increased by a time-dependent scale factor $R(t)$. The position of the observed (or test) object is then

$$\mathbf{r} = R(t)\mathbf{x} \tag{1}$$

where $\mathbf{x}$ is the position of the galaxy in the expanding coordinate system.

We shall not proceed as formally as we did in the case of the rotating coordinate system since you have already seen how that goes. Let us just differentiate (1) with respect to time to obtain the velocity of the galaxy in the inertial frame as

$$\frac{d\mathbf{r}}{dt} = R(t)\frac{d\mathbf{x}}{dt} + \dot{R}\mathbf{x} \tag{2}$$

where the dot means time derivative and $\dot{R}$ is the rate of change of the scale factor. This can be rewritten as

$$\frac{d\mathbf{r}}{dt} = R(t)\frac{d\mathbf{x}}{dt} + H\mathbf{r} \tag{3}$$

where

$$H(t) = \frac{\dot{R}}{R}\ . \tag{4}$$

This relation between the velocities as seen in the two frames is the analogue of (21) in Installment III. In cosmology, the present value of $H(t)$ is called the Hubble constant. The quantity $H\mathbf{r}$ is the expansion velocity of the background medium at position $\mathbf{r}$ and it is called the Hubble velocity. The velocity of the galaxy relative to the expanding background medium, $R\dot{\mathbf{x}} = \dot{\mathbf{r}} - H\mathbf{r}$, is called the peculiar velocity. Relativists call $\dot{\mathbf{x}}$ the coordinate velocity to distinguish it from the (physical) peculiar velocity, $R\dot{\mathbf{x}}$. In that sense, an angular velocity may also be called a coordinate velocity.

When we differentiate (3), we get

$$\frac{d^2\mathbf{r}}{dt^2} = R\frac{d^2\mathbf{x}}{dt^2} + \dot{R}\frac{d\mathbf{x}}{dt} + \dot{H}\mathbf{r} + H\frac{d\mathbf{r}}{dt} \tag{5}$$

Now we use (3) and resort completely to the dot notation to obtain

$$\ddot{\mathbf{r}} = R\ddot{\mathbf{x}} + 2HR\dot{\mathbf{x}} + \dot{H}\mathbf{r} + H^2\mathbf{r} \tag{6}$$

The four terms on the right of this equation are (1) the acceleration in the expanding frame, (2) a drag term on the peculiar velocity analogous to the Coriolis term, (3) an acceleration like the Euler acceleration, and (4) something like the centrifugal term. However, in this case we have the simplification that

$$\dot{H} = \frac{\ddot{R}}{R} - H^2 \tag{7}$$

so that (6) becomes

$$\ddot{\mathbf{r}} = R\ddot{\mathbf{x}} + 2HR\dot{\mathbf{x}} + \frac{\ddot{R}}{R}\mathbf{r}\ . \tag{8}$$

A nice way to rewrite this is

$$\ddot{\mathbf{r}} = \frac{1}{R}\left(R^2\dot{\mathbf{x}}\right)^{\cdot} + \ddot{R}\mathbf{x}\ . \tag{9}$$

To complete this story, let specify $R(t)$ for the case of Newtonian cosmology. Having chosen a radius vector $\mathbf{r}$, we may say that the background outflow is opposed by the gravitational pull of the mass M interior to the sphere of radius $r = |\mathbf{r}|$. We express that as

$$\ddot{\mathbf{r}} = -\frac{GM\mathbf{r}}{r^3} \tag{10}$$

where $\mathbf{r}/r$ is the unit vector in the direction of $\mathbf{r}$. When we start off the motion described in this way with no peculiar velocity, the solution tells how the background expansion is proceeding. If the medium is uniform with density ρ, we have

$$M = \frac{4\pi}{3}\rho r^3\ . \tag{11}$$

This interior mass is expanding since it is part of the background medium. So as, r grows, ρ dimishes so as to keep M constant. Since r grows like R, ρ decreases like R^{-3}. Then

$$\ddot{\mathbf{r}} = -\frac{4\pi}{3}G\rho\mathbf{r} = -\frac{4\pi}{3}G\rho R\mathbf{x} \tag{12}$$

When we introduce this into (9), and choose the background dynamics according to

$$\ddot{R} = -\frac{4\pi}{3} G\rho R \tag{13}$$

we get

$$\left(R^2\dot{\mathbf{x}}\right)^{\cdot} = 0 \ . \tag{14}$$

Therefore

$$R^2\dot{\mathbf{x}} = \mathbf{constant}. \tag{15}$$

The magnitude of the peculiar velocity of a galaxy decreases as R^{-1}. This is a result of the fictitious drag force in an expanding medium. When you let the air out of a tire, the air expands and the peculiar velocities of the constituent molecules diminish. Since the temperature is a measure of the mean peculiar velocity, you expect the gas to cool. Of course, it does, and we shall later discuss how to estimate the cooling rate. We merely state for now that the universe must have been hotter in the past.

When we study the relativistic case, the expansion will be seen as an expansion of space itself. Waves traveling through space will have their wavelengths stretched like $R(t)$. Quantum mechanics tells us that, as the wavelength associated with a particle grows with R, its momentum diminishes like R^{-1}, in agreement with our present result. Photons too have their wavelengths stretched and this gives rise to the famous cosmological redshift. However you look at it, the acceleration produces an effective force but, in our discussion so far, we have thrown in gravity as if it were a real force. This is what we want to change.

2 Equivalence of Acceleration and Gravity

Suppose that you are in a box, or elevator, and cannot look out. Figure 1 shows two possible situations. In (a) the elevator is on the ground and subject to the downward acceleration of the earth's gravity, $\mathbf{g}$. In (b), the elevator is in space, far from any sources of gravity, and being towed from its ceiling with an acceleration $\mathbf{a} = \mathbf{g}$ in a direction that we shall call upward. Can you tell which situation you are in without looking outside? You may ask for any equipment you need for this determination, except perhaps for a mobile phone.

You could ask for a laser and send a beam across the elevator from one wall to the one opposite. In case (b), the beam would hit the opposite wall at a lower height than the one at which it started. Einstein's theory of gravity started from the premise that the two situations are equivalent and that a downward deflection would be observed in case (a) also. He had already demonstrated the equivalence of matter and energy, so it was natural to suppose that the light beam would be bent downward by gravity acting on its weight. This prediction was in fact later confirmed by the study of the deflection of light by the sun during an eclipse. However, there is a way that the distinction between cases (a) and (b) could be made. No real gravitational field is constant in space and the difference in gravitational force from place to place produces tidal distortions. This is what we shall look into next.

2.1 Force and Potential

** The material in this section can be skipped and in any case need not be worried over. **

In equation (I-11) we referred to the notion of a force derivable from a potential on the assumption that you knew a little about such things. I assume that you know that a body like the earth has a gravitational potential $\phi(\mathbf{x})$ where $\mathbf{x}$ is the position vector with respect to the earth's center. In principle, ϕ could depend on time, but let us leave out such refinements for now. An equation like $\phi(\mathbf{x}) = constant$ defines surfaces in space and these have dimension one less than the dimension of the space itself. The surfaces are then said to have codimension one.

Consider two points $\mathbf{x}$ and $\mathbf{x} + \delta\mathbf{x}$ close to each other. The difference in potential between the two points is

$$\delta\phi = \phi(\mathbf{x} + \delta\mathbf{x}) - \phi(\mathbf{x}) = \frac{\partial\phi}{\partial x^i}\delta x^i + ... \tag{16}$$

where x^i are the components of $\mathbf{x}$ (and likewise for the δx^i). Let us introduce a quantity

$$d\phi = \frac{\partial\phi}{\partial x^i}dx^i \tag{17}$$

that is not to be thought of as infinitesimal, though its form may have been inspired by that of $\delta\phi$, which is infinitesimal. Many people have a bit of trouble with this concept, so don't worry if you are worried. Now the differential forms make up an abstract vector space in which the dx^i are treated as basis vectors. Then the $\partial\phi/\partial x^i$ are considered components of what is called the gradient of ϕ. If you already know what a gradient is from your vector analysis (and I hope you do), then you need not worry. I include this to give some idea of why the components of gradients have their indices down. Also, I wanted you to have a sense of *deja vu* when we see some similar things later on. There is a book by Bill Burke called 'Spacetime, Geometry and Cosmology' (not to be confused with his geometry book) that tries to tell these things like they are in terms aimed at freshpersons. It is pleasant reading but it is pretty incomprehensible, but have a look at it.

2.2 Dynamics in a Falling Elevator

Suppose you are in an elevator well above the earth's surface and falling freely. Imagine that it is not tumbling — it it sufficiently symmetric laterally that the force of gravity from the earth does not exert a torque. We want to look at the dynamics of this falling.

So you are in a box falling in a potential *per unit mass* $\phi(\mathbf{x})$. Your position vector wherever you are in the box has components x^i. Your equation of motion in component form is

$$\ddot{x}^i = \delta^{ij}\frac{\partial\phi}{\partial x^j} \tag{18}$$

where δ^{ij} follows the usual rules of the Kronecker delta: it is one if i and j are the same and zero if they are not.

You may wonder why I have written Newton's second law in this seemingly odd way. The real reason is that I am being unnecessarily pedantic The fact is that it is more natural in some sense to express gradients in terms of the reciprocal basis because, when you take a derivative of a function with respect to x^i, the index winds up downstairs. We emphasize this by using a notation that people use in relativity (and some other subjects), namely

$$\frac{\partial \phi}{\partial x^i} = \phi_{,i} \tag{19}$$

In both forms of this partial differential equation, the index is a lower one. Then the gradient of ϕ is written as

$$\nabla\phi = \phi_{,i}\mathbf{e}^i \tag{20}$$

where $\mathbf{e}^i$ are the reciprocal basis vectors, of which we shall speak farther down the line. On the other hand, we have been writing the coordinate components as x^i, so I used the Kronecker delta in (18) to make everything look grammatical. I apologize for this unnecessary fussiness. If you like, just replace (18) by

$$\ddot{\mathbf{x}} = -\nabla\phi \tag{21}$$

Let $\mathbf{x}_{(0)}$ be the position vector of the center of the box with components $x^i_{(0)}$. The Taylor series for ϕ around the box center is

$$\phi(\mathbf{x}) = \phi(\mathbf{x}_{(0)}) + \phi_{,i}(\mathbf{x}_{(0)})\xi^i + \frac{1}{2}\phi_{,i,j}(\mathbf{x}_{(0)})\xi^i\xi^j + ... \tag{22}$$

where

$$\xi^i = x^i - x^i_{(0)} \tag{23}$$

When apply (21) at $\mathbf{x}_{(0)}$ and subtract it from (21) itself, we obtain

$$\ddot{\xi}^i = -\xi^k\delta^{ij}\phi_{,j,k}(\mathbf{x}_{(0)}) \tag{24}$$

This equation tells how the coordinates with respect to the center of the box are changing.

If we take as the potential

$$\phi(\mathbf{x}) = -\frac{GM}{r}\ , \tag{25}$$

we need to do some differentiating of this to make (24) explicit. With

$$r = \sqrt{\delta_{ij}x^ix^j} \tag{26}$$

we find

$$\left(\frac{1}{r}\right)_{,k} = -\frac{1}{2r^3}\left(\delta_{ij}x^ix^j\right)_{,k} \tag{27}$$

Since $x^i_{\cdot,j} = \delta^i_{\cdot j}$, we find that

$$\phi_{,i} = \frac{GM}{r^3} x_i \tag{28}$$

Taking the second derivative proceeds in a similar way and it is left for homework. We get

$$\phi_{,i,j} = \frac{GM}{r^3} \left(\delta_{ij} - \frac{3x_i x_j}{r^2} \right) . \tag{29}$$

We then conclude that

$$\ddot{\xi}^i = -\frac{GM}{r^3_{(0)}} \left[\delta^i_{\cdot j} - \frac{3x_j x^i}{r^2} \right]_{(0)} \xi^j \ . \tag{30}$$

Suppose we work in spherical coordinates, which is the natural choice for this spherical problem. The box falls down a radius vector and the coordinates of its center are $x^1 = r_{(0)}$, $x^2 = 0$, $x^3 = 0$. The equation of motion of a displacement from the center of the box becomes

$$\ddot{\xi}^i = -E_{ij} \xi^j \tag{31}$$

where

$$E_{ij} = -\frac{GM}{r_0^3} \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ . \tag{32}$$

This equation describes how the differential gravitational force deforms things. That effect is called a tide and, as you see, the tidal interaction varies as the inverse cube of the distance to the object.

General Relativity

Installment V.

September 22, 2003

1 Moving on the Earth's Surface

1.1 Relative Motion

What we saw in the last installment is that, in the box falling toward the earth, the horizontal motion of a particle not at the center of the box is toward the center of the box. If the horizontal displacement of the particle from box center is ξ, the equation for the horizontal motion with respect to the box is

$$\ddot{\xi} = -\frac{GM}{r^3}\xi \ . \tag{1}$$

The quantity GM/r^3 is a frequency squared and is a measure of the strength of the tidal force. This gathering in the horizontal is reminscent of motion on the earth.

Consider two travelers on earth starting out at the same time from the equator, but at slightly different longitudes. Each heads due north in heir own frame. If their northward velocities are the same, they will also be moving toward each other unintentionally till they meet at the north pole. This is a similar sort of motion to what we saw in the tidal case. Let us see how to treat this.

Suppose that the earth's surface is a perfect sphere with radius R. This geometrical object is called $\mathcal{S}^2$ and it will prepare us for thinking about Einstein's model of the universe, $\mathcal{S}^3$. As in the tidal problem, let us consider that the two travelers are near to each other and that they are moving along their respective meridians. Let their longitudes be ϕ and $\phi + d\phi$ and their (common) latitude be θ. The distance between them is $R\sin\theta d\phi$. This is measured on the earth's surface and perpendicularly to the meridians. This distance is small; let us express it as $\xi = R\sin\theta d\phi$. How does ξ vary as the travelers move north?

Along a meridian, the distance moved in a displacement $d\theta$ in latitude is $ds = Rd\theta$. So we calculate that

$$\frac{d^2\xi}{ds^2} = \frac{1}{R^2}\frac{d^2\xi}{d\theta^2} = \frac{1}{R^2}\frac{d^2(R\sin\theta d\phi)}{d\theta^2} = -\frac{1}{R}d\phi\sin\theta\ . \tag{2}$$

Hence, we have (essentially) what is called Jacobi's formula

$$\frac{d^2\xi}{ds^2} = -\frac{1}{R^2}\xi\ . \tag{3}$$

which is of the same form as (**??**). However, in this case, the coefficient which determines the way the two travelers approach one another is the radius of the earth, a (manifestly) geometric property, related to its curvature. This is a special case of a formula for what is called geodesic deviation that we shall derive in a higher dimensional example later on.

1.2 Geometry of $\mathcal{S}^2$

We can study the geometry of a space either by acting as someone living in the space or by being someone in a space of higher dimension looking at the space in question. Let us look at distances on the surface of the two-dimensional sphere, a.k.a. $\mathcal{S}^2$, from our standpoint of the three-dimensional Euclidean space, $\mathbb{E}^3$. Suppose two points on $\mathcal{S}^2$, one at x, y, z and the other at $x+dx, y+dy, z+dz$, are separated by a very small distance ds. We have

$$ds^2 = dx^2 + dy^2 + dz^2\ . \tag{4}$$

But the points are both on the sphere. So we require that

$$x^2 + y^2 + z^2 = R^2 \tag{5}$$

where R is the radius of the sphere. From this, we find that $xdx + ydy + zdz = 0$ so that

$$dz^2 = \frac{(xdx + ydy)^2}{z^2} = \frac{x^2dx^2 + y^2dy^2 + 2xydxdy}{R^2 - (x^2+y^2)} \tag{6}$$

The formula for ds^2 becomes

$$ds^2 = \frac{(R^2 - y^2)dx^2 + (R^2 - x^2)dy^2 + 2xydxdy}{R^2 - (x^2+y^2)} \tag{7}$$

And so we have a distance formula for two neighboring points on $\mathcal{S}^2$. But a resident of $\mathcal{S}^2$ might not like using x and y and would prefer some more convenient coordinates.

Nice coordinates on $\mathcal{S}^2$ are latitude θ and longitude ϕ defined by

$$x = R\sin\theta\cos\phi \qquad\qquad y = R\sin\theta\sin\phi \tag{8}$$

From this we find that

$$dx = R\cos\theta\cos\phi d\theta - R\sin\theta\sin\phi d\phi \tag{9}$$

$$dy = R\cos\theta\sin\phi d\theta - R\sin\theta\cos\phi d\phi \tag{10}$$

Now we may calculate $(dx)^2, (dy)^2$ and $dxdy$ and substitute and rearrange to obtain

$$ds^2 = R^2 d\theta^2 + sin^2\theta d\phi^2, \tag{11}$$

as we might have found directly. However, when we come to look at $\mathcal{S}^3$, we shall find the present approach more relaxing.

We have found the distance on $\mathcal{S}^2$ between two points separated in three dimensional Euclidean space by the infinitesimal vector $d\mathbf{x}$. In the latter space, we may write $d\mathbf{x} = dx^i\mathbf{e}_i$ where the i here is $1, 2, 3$. In the Euclidean case, we have that the distance is the square root of $d\mathbf{x}\cdot d\mathbf{x} = dx^i dx^j \mathbf{e}_i \cdot \mathbf{e}_j = dx^i dx^j g_{ij}$ where $g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$. Though we do not have the same situation in $\mathcal{S}^2$ — the vectors at a point are in the space tangent to $\mathcal{S}^2$ at that point, not in $\mathcal{S}^2$ itself — we see that the formula for distance in $\mathcal{S}^2$ looks like the one in Eucldean space. So, instead of calculating g_{ij} from an inner product as before, we specify the nature of the geometry by simply specifying g_{ij}. That is, we give what is called a metric on our space.

For $\mathcal{S}^2$, the result about the distance between two points that we found may be summarized as

$$ds^2 = g_{ij} dx^i dx^j \tag{12}$$

where $x^1 = \theta$, $x^2 = \phi$ and

$$g_{ij} = \begin{pmatrix} R^2 & 0 \\ 0 & R^2\sin^2\theta \end{pmatrix} . \tag{13}$$

This g_{ij} does what our old one did. For instance, it can be used to lower an index as in $dx_i = g_{ij}dx^j$ so that $ds^2 = dx_i dx^i$. And we define the inverse metric by

$$g^{ik}g_{kj} = \delta^i_{.j} \tag{14}$$

as before. (What is g^{ij} for $\mathcal{S}^2$?)

1.3 A Trip

Let us think about taking a trip over the surface of the earth from point P to point Q. The distance covered is

$$s(P-Q) = \int_P^Q ds \tag{15}$$

which evidently depends on the path taken. But

$$ds = \sqrt{g_{ij}dx^i\, dx^j} \tag{16}$$

so the path with a stationary length is the one for which

$$\delta s(P-Q) = \delta \int_P^Q \sqrt{g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds}} ds = 0\ . \tag{17}$$

Since

$$\sqrt{g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds}} = 1, \tag{18}$$

we may replace this quantity in the integrand by any power of it. All these choices in fact lead to same result. (If you want to pursue this, look in 'General Relativity' by H. Stephani, a Cambridge Press publication suitable for a slightly more advanced course than ours. The equivalence of the different approaches is shown there.) It should not be surprising that, if the integral of a quantity equal to one is stationary, the integral of its square is also stationary. So let us look instead for the path such that

$$\delta\int_P^Q \left(g_{ij}\frac{dx^i}{ds}\frac{dx^j}{ds}\right) ds = 0 \ . \tag{19}$$

In fact, there is a reason in relativity to consider this case rather than the square root, as we shall see when we get there. Then

$$\int_P^Q \left[\frac{\partial g_{ij}}{\partial x^r}\delta x^r \dot{x}^i \dot{x}^j + g_{ij}\dot{x}^i \delta\dot{x}^j + g_{ij}\dot{x}^j\delta\dot{x}^i\right] = 0 \ . \tag{20}$$

where $dx^i/ds = \dot{x}^i$. This may also be written as

$$\int_P^Q \left[\frac{\partial g_{ij}}{\partial x^r}\delta x^r \dot{x}^i \dot{x}^j + 2g_{ij}\dot{x}^i \delta\dot{x}^j\right] = 0 \ . \tag{21}$$

We integrate by parts to get rid of the s-derivatives on the δx^ℓ and we have

$$\int_P^Q \left[\frac{\partial g_{ij}}{\partial x^r}\delta x^r \dot{x}^i \dot{x}^j - 2\frac{d}{ds}\left(g_{ij}\dot{x}^i\right)\delta x^j\right] = 0 \ . \tag{22}$$

Now we may rename some dummy indices to obtain

$$\int_P^Q \left[\frac{\partial g_{ij}}{\partial x^r}\dot{x}^i \dot{x}^j - 2\frac{d}{ds}\left(g_{ir}\dot{x}^i\right)\right]\delta x^r = 0 \ . \tag{23}$$

The perturbation δx^r represents the difference in position between the route being tested and a neighboring one that we are examining for its promise as a shorter alternative. Therefore, δx^r is arbitrary. And, since the perturbed integral must vanish for any (small) δx^r if we are to obtain a true extremum, the integrand must vanish. Hence

$$-2\frac{d}{ds}\left(g_{kr}\dot{x}^k\right) + \frac{\partial g_{k\ell}}{\partial x^r}\dot{x}^k\dot{x}^\ell = 0. \tag{24}$$

On carrying out the derivative of the product, we get

$$-2\dot{x}^i\frac{dg_{ij}}{ds} - 2g_{ij}\ddot{x}^i + \frac{\partial g_{ik}}{\partial x^j}\dot{x}^i\dot{x}^k = 0 \tag{25}$$

Then, since $dg_{ij}/ds = (\partial g_{ij}/\partial x\ell)\dot{x}^\ell$, we may write the equation for the optimal path as

$$2g_{ij}\ddot{x}^i = -2g_{ji,\ell}\dot{x}^\ell\dot{x}^i + g_{ik,j}\dot{x}^i\dot{x}^k \tag{26}$$

where we recall the notation from a previous installment, $[\cdot]_{,i} = \partial[\cdot]/\partial x^i$, where $[\cdot]$ is whatever you like.

Now we multiply by g^{jr}. Since $g^{jr}g_{ij} = \delta^r_{\cdot i}$, we see that

$$\ddot{x}^r = \frac{1}{2}g^{jr}\left[g_{ik,j} - g_{ji,k} - g_{jk,i}\right]\dot{x}^i\dot{x}^k \tag{27}$$

Finally, we may introduce the three-index quantity

$$\Gamma^k_{ij} = \frac{1}{2}g^{rk}\left[g_{ir,j} + g_{jr,i} - g_{ij,r}\right] \tag{28}$$

to give us

$$\ddot{x}^k + \Gamma^k_{ij}\dot{x}^i\dot{x}^j = 0\ . \tag{29}$$

The solution to this equation gives us stationary paths in the sense that a slight deviation from such a path makes a very small change in the distance traveled.

If you followed the section on Lagrange's equations, you may recognize the similarity of this derivation to the one there. Even if you skipped that part, you may see a resmblance of this equation to Newton's second law. Of course, the dots here refer to spatial derivatives, but that distinction will no longer be made once we get into spacetime.

General Relativity

Installment VI.

September 24, 2003

1 Spherical Coordinates

In the last Installment we saw how to describe motion along (the generalization of) "straight lines" in curved spaces. But we should be aware that, even in Euclidean space, this description could look complicated in other than Cartesian coordinates. Consider the case of spherical coordinates $x^1 = r, x^2 = \theta, x^3 = \phi$. It is not hard to see that the distance between two points separated by coordinate differentials $dr, d\theta, d\phi$ is given by

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2 \ . \tag{1}$$

This may be written as

$$ds^2 = g_{ij} dx^i \, dx^j \tag{2}$$

where

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix} \ . \tag{3}$$

If we repeat the calculation that we did for $\mathcal{S}^2$ to find the equation for paths with stationary lengths, we obtain the same equation as before, namely

$$\ddot{x}^i + \Gamma^i_{jk} \dot{x}^j \dot{x}^k = 0 \tag{4}$$

where

$$\Gamma^i_{jk} = \frac{1}{2} g^{i\ell} (g_{j\ell,k} + g_{k\ell,j} - g_{jk,\ell}) \tag{5}$$

You may (perhaps should) calculate the Christoffel symbols. They are

$$\Gamma^1_{22} = -r; \qquad \Gamma^1_{33} = -r\sin^2\theta; \qquad \Gamma^2_{12} = \Gamma^2_{12} = \frac{1}{r}; \qquad \Gamma^2_{33} = -\sin\theta\cos\theta \tag{6}$$

$$\Gamma^3_{13} = \Gamma^3_{31} = \frac{1}{r}; \qquad \Gamma^3_{23} = \Gamma^3_{32} = \cot\theta \ . \tag{7}$$

1.1 Lagrangian Mechanics

Another way to think about this result is to go back to mechanics. Along the trajectory followed by (say) a particle, we can measure distance, s. But we may also put down the time, t, along the path, if we could but find it out. In any case, we may presume a relation between distance traveled and time, $s(t)$. The speed of the particle is then ds/dt and the components of the velocity vector are dx^i/dt, where it is understood that the basis vectors are along the directions of the spherical coordinates. (We are allowed to presume this as we are in Euclidean space in this example.) Then (2) tells us how to compute the square of the speed and thus the kinetic energy. Since there is no potential energy here, this latter is the same as the Lagrangian,

$$L = \frac{m}{2} g_{ij} \frac{dx^i}{dt}\frac{dx^j}{dt} \tag{8}$$

where m is the mass. The action for a trip from P to Q is

$$\mathcal{A} = \int_{t_P}^{t_Q} L dt \tag{9}$$

and it should be stationary on the physically chosen path. That is,

$$\delta\mathcal{A} = \delta \int_{t_P}^{t_Q} L dt = 0 \ . \tag{10}$$

This gives us the equation (see the end of Installment I, which may now seem more readable to you),

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} - \frac{\partial L}{\partial x^i} = 0 \tag{11}$$

where $\dot{x}^i = dx^i/dt$

Now,

$$\frac{\partial L}{\partial \dot{x}^i} = m g_{ij} \dot{x}^j \tag{12}$$

This is basically the mass times the velocity and it is called the momentum (or canonical momentum) and indeed it does look like a momentum. But there is this g_{ij} in there. What it does is to lower the index on the mv^i. The momentum comes in with its index downstairs and so it is perhaps natural that in mechanics (both classical and quantum mechanical) it is said to be conjugate to the coordinate.

In explicit spherical coordinates,

$$L = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta \dot{\phi}^2\right) \tag{13}$$

Hence

$$\frac{\partial L}{\partial \dot{r}} = m\dot{r}; \qquad \frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta}; \qquad \frac{\partial L}{\partial \dot{\phi}} = m\dot{r}^2 \sin^2\theta\phi \tag{14}$$

Indeed, the first of these is a momentum and the other two are angular momenta. We find the equations of motion

$$\ddot{r} - r\,\dot{\theta}^2 - r\sin^2\theta\,\dot{\phi} \tag{15}$$

$$\ddot{\theta} + \frac{2}{r}\,\dot{r}\,\dot{\theta} - \sin\theta\ \cos\theta\,\dot{\phi}^2 \tag{16}$$

$$\ddot{\phi} + \frac{2}{r}\,\dot{r}\,\dot{\phi} + 2\cot\theta\,\dot{\phi}\,\dot{\theta} = 0\ . \tag{17}$$

What a way to describe a straight line! But, as we see, this set of equations is none other than the geodesic equation (4). We can then read off the Christoffel symbols by comparing with the general form and we get (6)–(7). (This is the secret easy way to do things.)

If we choose our coordinate system well, we can have the particle moving with constant ϕ. Then (15)–(16) become

$$\ddot{r} - r\,\dot{\theta}^2 = 0 \tag{18}$$

$$\ddot{\theta} + \frac{2}{r}\,\dot{r}\,\dot{\theta} = 0 \tag{19}$$

The second equation integrates to

$$r^2\dot{\theta} = \text{constant}. \tag{20}$$

This is an expression of the conservation of angular momentum in a system with no torques. (If you do not know how to find this result, please see me.) Call the constant ℓ and (18) becomes

$$\ddot{r} - \frac{\ell^2}{r^3} = 0\ . \tag{21}$$

This equation is of second order, so its solutions will have two arbitrary constants. A first integral is gotten by multiplying by $\dot{r}$ and integrating. We get

$$\frac{1}{2}\dot{r}^2 + \frac{\ell^2}{2\,r^2} = \text{constant} = E\ . \tag{22}$$

Since the origin of time is arbitrary we may write the solution as

$$r = \sqrt{\ell^2/(2E) + 2Et^2} \tag{23}$$

We can then solve (20) to obtain

$$t = \frac{\ell}{2E}\tan(\theta + \theta_0) \tag{24}$$

where θ_0 is an integration constant. And so we find that

$$r\ \cos(\theta + \theta_0) = \frac{\ell}{\sqrt{2E}} \tag{25}$$

So we have found that a free particle moving in Euclidean space with no forces acting on it moves in a straight line. This reminds me of what Voltaire wrote to Maupertuis when he went to Lapland to see whether the earth was oblate as Newton had said. Voltaire wrote (more or less):

Vous avez mesuré dans des lieux pleine d'ennui
ce que Newton savait sans sortir de chez lui.

Here a significant issue is raised. Someone coming in at the middle of this discussion could well imagine that we were talking about complicated motion in a curved space when in fact all we are doing is looking at the motion of a free particle. The compications arise only because we have injudiciously chosen curvilinear coordinates, which are not optimal for talking about straight lines. So how can we tell when this has happened and that we are really in a flat space? That involves what is called a test for curvature that we shall learn later.

1.2 Arbitrary Coordinates

As we see, the price of doing things in a general way is that we may overlook how simple things really are in our attempt to do it all at once. But once we are aware of that, we can throw caution to the winds and go into any coordinate system we feel like, knowing that there is a simplest version somewhere.

So suppose we have a coordinate system x^i, in whatever the space is and we go into another system. This time, let us see how we like using the notation that some people use for the other system of coordinates, that is $x^{i'}$. We then look for a transformation

$$x^i = x^i(x^{i'}). \tag{26}$$

If we have a particle moving on a particular trajectory, the tangent vector to the trajectory is

$$\frac{dx^i}{ds} = \frac{\partial x^i}{\partial x^{k'}}\frac{dx^{k'}}{ds} \ . \tag{27}$$

This is the transformation of the tangent vector from one coordinate system to the other and it is effected by the transformation matrix $\partial x^i/\partial x^{k'}$. Let us suppose that the transformation has an inverse. That is, that it is possible to solve for $x^{k'}(x^i)$. Then the transformation matrix taking vectors back the other way is $\partial x^{k'}/\partial x^i$ with

$$\partial x^i/\partial x^{k'}\partial x^{k'}/\partial x^j = \delta^i_{\cdot j} \tag{28}$$

The second derivative is

$$\frac{d^2x^i}{ds^2} = \frac{\partial^2 x^i}{\partial x^{k'}\partial x^{\ell'}}\frac{dx^{k'}}{ds}\frac{dx^{\ell'}}{ds} + \frac{\partial x^i}{\partial x^{k'}}\frac{d^2x^{k'}}{ds^2} \tag{29}$$

For a motion in a straight line in rectangular coordinates, we have $d^2x^i/ds^2 = 0$. If we impose this condition, and multiply by $\partial x^{i'}/\partial x^i$ we get

$$\delta^{i'}_{\cdot k'}\frac{d^2x^{k'}}{ds^2} + \frac{\partial x^{i'}}{\partial x^i}\frac{\partial^2 x^i}{\partial x^{k'}\partial x^{\ell'}}\frac{dx^{k'}}{ds}\frac{dx^{\ell'}}{ds} = 0 \ . \tag{30}$$

This may be written as

$$\frac{d^2 x^{i'}}{ds^2} + \Gamma^{i'}_{k'\ell'} \frac{dx^{k'}}{ds} \frac{dx^{\ell'}}{ds} = 0 \tag{31}$$

where

$$\Gamma^{i'}_{k'\ell'} = \frac{\partial x^{i'}}{\partial x^i} \frac{\partial^2 x^i}{\partial x^{k'} \partial x^{\ell'}} \ . \tag{32}$$

Notice that for a linear transformation, the nonacceleration property is invariant. More generally, the condition of no acceleration taken to arbitrary coordinates seems to tell us that the motion is geodesic; at least it will have once we shall have established the presumption that identification of the new expression for Γ is equivalent to the old one.

General Relativity

Installment VII.

September 28, 2003

1 Transformation of Vectors

The displacement is the vector *par excellence.* We express it in terms of the change of coordinates (dx^i) caused by the displacement ($d\mathbf{x}$). When we transform coordinates, we transform the components of $d\mathbf{x}$. To see how this works, recall that in a transformation of coordinates the new coordinates can be expressed as functions of the old coordinates, and vice versa. So we have

$$x^{i'} = x^{i'}(x^i). \tag{1}$$

Here the notation is such that the new coordinates get the primes on their indices so that $x^{i'}$ are the new coordinates and x^i are the old ones. In the new system, the components of the displacement vector are

$$dx^{i'} = \frac{\partial x^{i'}}{\partial x^i} dx^i \ . \tag{2}$$

So the transformation matrix $\partial x^{i'}/\partial x^i$ tell us what the components of $d\mathbf{x}$ are in the new coordinates (with the implied new basis vectors).

The infinitesimal displacement vector is such an exemplary vector that we shall demand that in our space of n dimensions any set of n quantities posing as the components of a vector has to establish its credentials by transforming like the components of $d\mathbf{x}$ as seen in (2). Suppose that we have a set of n numbers a^i given at every point in space. If these are supposed to be the components of a vector field $\mathbf{a}$ in the old system, then its components in the new coordinate system had better be

$$a^{i'} = \frac{\partial x^{i'}}{\partial x^i} a^i \ . \tag{3}$$

This then is our *d*efinition of what a vector field is — a set of n functions of position that transforms as in (3) when the coordinates transform as in (1).

We shall deal only with coordinate transformations that are invertible, so we can expect that

$$dx^i = \frac{\partial x^i}{\partial x^{i'}} dx^{i'} \tag{4}$$

where $\partial x^i/\partial x^{i'}$ is the inverse of the transformation matrix. That is, as in Installment VI,

$$\frac{\partial x^j}{\partial x^{i'}}\frac{\partial x^{i'}}{\partial x^k} = \delta^j_{\cdot k} \tag{5}$$

and

$$\frac{\partial x^{j'}}{\partial x^i}\frac{\partial x^i}{\partial x^{k'}} = \delta^{j'}_{\cdot k'} \tag{6}$$

If a point P has coordinates x^i in the old system and $x^{i'}$ in the new, a function $\phi(x^i)$ is called a *scalar* (or scalar field) if, under the transformation $x^i \to x^{i'}$, ϕ becomes

$$\phi'(x^{i'}) = \phi(x^i) \tag{7}$$

Now consider the derivative of a scalar with respect to the coordinates:

$$\frac{\partial \phi'}{\partial x^{i'}} = \frac{\partial \phi}{\partial x^j}\frac{\partial x^j}{\partial x^{i'}} = \frac{\partial x^j}{\partial x^{i'}}\frac{\partial \phi}{\partial x^j}\ . \tag{8}$$

If we let $b_i = \partial\phi/\partial x^i$, we see that this quantity transforms like

$$b_{i'} = \frac{\partial x^i}{\partial x^{i'}} b_i\ . \tag{9}$$

Components that transform in this way are called covariant while those (with upper indices) that transform as in (3) are called contravariant. The names have to do with the comparison to the transformation of coordinates with the transformation matrices

$$\Lambda^{i'}_{\cdot i} = \frac{\partial x^{i'}}{\partial x^i}; \qquad\qquad \Lambda^{i}_{\cdot i'} = \frac{\partial x^{i}}{\partial x^{i'}}\ . \tag{10}$$

The two sets of components actually represent geometrically different objects but, as they both form vector spaces of dimension n, we can match them up one by one using g_{ij}. The distinction is more important in situations where there is no g_{ij} provided, but that is a hardship we shall not face. (Mathematicians do seem to enjoy roughing it in that way, though.) It is nevertheless worth keeping the distinction in mind. It is exemplified by the tangent vector to a curve — a one dimensional object — being identified with vectors and the derivative of a function — with codimension one — being identified with differential forms. The former is often represented as an arrow and the latter by a pair of parallel line segments. (Burke's 'Spacetime, Geometry and Cosmology' (Univ. Science Books, 1980) has a long, chatty and bewildering discussion of all that at what is intended to be a very elementary level. Have a look at it if you can find it.)

2 Tensors

What happens to two-indexed quantities like g_{ij} when we transform coordinates? To think about this, consider the formula for the distance between two points that are only slightly separated so that their coordinates differ by only dx^i. The distance between them, ds, is given by the formula

$$ds^2 = g_{ij} dx^i\, dx^j \tag{11}$$

where g_{ij} is specified for the space we are in.

The value of the separation between the two points, ds, should not change when we change the coordinates, so we may also write that

$$ds^2 = g_{i'j'} dx^{i'}\, dx^{j'} \tag{12}$$

where $x^{i'}$ is the new coordinate system. On equating the two expressions for the distance interval, we find

$$g_{i'j'} dx^{i'}\, dx^{j'} = g_{ij} dx^i\, dx^j = g_{ij} \left(\frac{\partial x^i}{\partial x^{i'}} dx^{i'} \right) \left(\frac{\partial x^j}{\partial x^{j'}} dx^{j'} \right) . \tag{13}$$

That is,

$$\left(g_{i'j'} - g_{ij} \frac{\partial x^i}{\partial x^{i'}} \frac{\partial x^j}{\partial x^{j'}} \right) dx^{i'} dx^{j'} = 0 \ . \tag{14}$$

Since the coordinate differentials are arbitrary, we may conclude that

$$g_{i'j'} = g_{ij} \frac{\partial x^i}{\partial x^{i'}} \frac{\partial x^j}{\partial x^{j'}} \ . \tag{15}$$

Or, in the new notation,

$$g_{i'j'} = \Lambda^{i}_{\cdot i'}\, \Lambda^{j}_{\cdot j'}\, g_{ij} \tag{16}$$

So g_{ij}, with two indices downstairs, has a transformation matrix operating on each of its indices in the way that the components of vectors have transformations that work on their single indices. If you go through the same exercise with g^{ij}, you will find a similar result but with the upper indices transforming like the index of contravariant vector components.

More generally, one says that any set of n^2 functions labeled by two indices, as T_{ij}, where n is the range of i and j, makes up the components of a covariant tensor of second rank if, under transformation of coordinates, these components transform like

$$T_{i'j'} = \Lambda^{i}_{\cdot i'}\, \Lambda^{j}_{\cdot j'}\, T_{ij} \tag{17}$$

Likewise, a set of quantities T^{ij} that transforms like

$$T^{i'j'} = \Lambda^{i'}_{\cdot i}\, \Lambda^{j'}_{\cdot j}\, T^{ij} \tag{18}$$

is a contravariant tensor of rank two. Again, we may similarly define a mixed tensor $T^i_{\,j}$ of rank two. What would be the appropriate transformation law for this object?

If a^i and b^j are the contravariant components of two vectors, and T_{ij} are the components of a covariant tensor of rank two, what are the natures of $T_{ij}a^i$ and of $T_{ij}a^i\,b^i$ as determined by their transformation properties?

In this language, the vector components a^i are said to constitute the components of a contravariant tensor of rank one. And, naturally, a scalar is a tensor of rank zero. Tensors of higher rank than two, contravariant, covariant and mixed may also be defined, as we shall see later. (Try writing down now what would be the sensible definition for such objects.) But just because some quantities have lots of indices, you cannot assume that they are the components of a tensor. For example, the Christoffel symbols $\Gamma^{\,i}_{jk}$ are not the components of a third rank tensor.

When Columbia physicist I.I. Rabi learned of the discovery of a new and unexpected particle, he famously said "Who ordered that?" Perhaps you are entertaining similar thoughts about tensors. However, you can certainly see the usefulness of g_{ij}. But you may still wonder whether there is a real need for a whole class of such objects. You would not wonder at this if you had studied elasticity in any depth. In fact, that is where the word tensor (as in tension) comes from. If you have a continuous medium and you cut it into two parts (in your mind) by putting a plane through it, you could ask what is the force at a point in the plane exerted by the medium on one side on the material on the other side. There you have two vectors, one representing the surface (the plane could have any orientation) and the other representing the force itself. If you wanted to codify the information contained in that result, you would do well to introduce a two-index quantity. Of course, the two-indexed quantity need not be a tensor; you could use something called a dyad, for example. But the dyad does not have the nice properties of a tensor. For example, the sum of two dyads is not necessarily a dyad, and this is no good at all. So here the subject begins. Before we are through, you will stop worrying and learn to love tensors.

3 Differentiation of Tensors

3.1 Vectors First

As with vectors, you can add tensors by simply adding the corresponding components: for example $T^{ij} + S^{ij}$ is a second rank contravariant tensor with components U^{ij} say. But you may not add components of tensors of differing types: that is, $T^{ij} + S_{ij}$, for instance, has no meaning.

You may also multiply components pretty freely. So if a^i and b_j are components of vectors, then $a^i\, b_j$ is a mixed tensor of rank two as you may verify by checking the transformation properties. Also, you may multiply components and contract indices. For example, you should verify that if T^{ij} and S_i^k are tensors of rank two, then is $T^{ij} S_i^k$ is a tensor of rank three.

All those manipulations are worth practicing and they make good sense. But life becomes a little more complicated when we ask what tensor calculus may be like. You may want to look at some tensor books to really get into this subject — the one by Synge and Schild is particularly good, I find. Here, we need only some basics of the calculus for our purposes. Let us begin with differentiation of vectors.

Suppose we have the covariant components of a vector, a_i. They transform like

$$a_i = \frac{\partial x^{j'}}{\partial x^i} a_{j'} \tag{19}$$

If we differentiate with respect to x^k, using the notation $a_{i,k} = \partial a_i / \partial x^k$, we find from (19) on using the rule for differentiation of a product that

$$a_{i,k} = \frac{\partial x^{j'}}{\partial x^i} a_{j',\ell'} \frac{\partial x^{\ell'}}{\partial x^k} + \frac{\partial^2 x^{j'}}{\partial x^i\, \partial x^k} a_{j'} \tag{20}$$

In terms of the earlier notation for transformation matrices,

$$\Lambda^{j'}_{.i} = \frac{\partial x^{j'}}{\partial x^i} \tag{21}$$

we may write this as

$$a_{i,k} = \Lambda^{j'}_{.i} \Lambda^{\ell'}_{.k} a_{j',\ell'} + \frac{\partial^2 x^{j'}}{\partial x^i\, \partial x^k} a_{j'} \tag{22}$$

Because of the second term on the right, $a_{i.j}$ does not follow the rule for the transformation of tensor components: partial differentiation of a vector field destroys the tensor character of the field. This seems to be an undesirable feature and we may ask, what went wrong? (The problem does not arise when the transformation matrices are constant in space. For then $\partial^2 x^{j'} / (\partial x^i\, \partial x^k) = (\Lambda^{j'}_i)_{,k} = 0$.)

3.2 Another Difficulty

To see further into the problems with differentiation, let us consider the components of a vector, a^i, at point P with coordinates x^j. At a nearby point Q, the coordinates are $x^j + d\, x^j$. You might be tempted to suggest that the components at Q should be $a^i + d\, a^i$, where $d\, a^i = (\partial a^i / \partial x^j) dx^j$. But this would be wrong. Though dx^j is a vector, $a^i_{,j}$ is not, so the product is not a vector. We need to ask how to move $\mathbf{a}$ from P to Q in order to see what the difference in the components between the two places really is.

In a small enough region around P, we can go into Cartesian coordinates, designated as $x^{\hat{i}}$. This is possible because we are considering spaces that are Euclidean in small enough regions and it is standard to take advantage of this feature of Riemannian geometry. (In 'General Relativity' by H. Stephani (Cambridge, 1980) there is a nice discussion of locally flat coordinates.) The components of $\mathbf{a}$ in the local Cartesian system are

$$a^{\hat{i}} = \Lambda^{\hat{i}}_{\cdot i} a^i \tag{23}$$

In this situation, we can locally slide $\mathbf{a}$ from P to Q and we may expect expect the change $\delta\mathbf{a}$ in $\mathbf{a}$ from that operation to vanish. That is, $\delta a^{\hat{i}} = 0$ since there is no change in the Cartesian components from a slight parallel displacement of $\mathbf{a}$. But of course, there will be a change in the components with respect to general coordinates. Hence, from (23), we obtain

$$\delta a^{\hat{i}} = 0 = x^{\hat{i}}_{,i} \delta a^i + x^{\hat{i}}_{\;,i,j} a^i d\,x^j \tag{24}$$

(The corresponding change in coordinates, δx^i, is the same as $d\,x^i$.) Thus, we find that

$$x^{\hat{i}}_{,i} \delta a^i = -x^{\hat{i}}_{\;,i,j} a^i d\,x^j \tag{25}$$

When we multiply this by $x^k_{\;,\hat{i}}$ and recall that

$$x^k_{\;,\hat{i}} x^{\hat{i}}_{\;,i} = \frac{\partial x^k}{\partial x^{\hat{i}}} \frac{\partial x^{\hat{i}}}{\partial x_i} = \delta^k_{\cdot i} \ , \tag{26}$$

we obtain

$$\delta a^k = -\, x^k_{\;,\hat{i}} x^{\hat{i}}_{\;,i,j}\, a^i\, d\,x^j \tag{27}$$

This confirms what we have already seen (for example, in VI–(32)): that (27) can be written as

$$\delta a^k + \Gamma^{\;k}_{ij} a^i d\,x^j = 0 \tag{28}$$

where

$$\Gamma^{\;k}_{ij} = x^k_{\;,\hat{i}} x^{\hat{i}}_{\;,i,j} \tag{29}$$

Let us now, as promised earlier, see how this expression for Γ can be related to the other expression we gave for it in terms of g_{ij} and its derivatives. This calculation starts with the realization that, in local Cartesian coordinates, the metric is given by

$$g_{\hat{i}\hat{j}} = \delta_{\hat{i}\hat{j}} \tag{30}$$

When we transform to other, more general, coordinates we find the metric

$$g_{ij} = x^{\hat{i}}_{\;,i} x^{\hat{j}}_{\;,j}\, \delta_{\hat{i}\hat{j}} \tag{31}$$

When we differentiate (31) with respect to x^k, we find that

$$g_{ij,k} = \left(x^{\hat{i}}_{\;,i,k} x^{\hat{j}}_{\;,j} + x^{\hat{i}}_{\;,i} x^{\hat{j}}_{\;,j,k} \right)\, \delta_{\hat{i}\hat{j}} \tag{32}$$

We may interchange indices in this equation to obtain the additional relations

$$g_{ik,j} = \left(x^{\hat{i}}_{\ ,i,j} x^{\hat{j}}_{\ ,k} + x^{\hat{i}}_{\ ,i} x^{\hat{j}}_{\ ,k,j} \right) \delta_{\hat{i}\hat{j}} \tag{33}$$

$$g_{kj,i} = \left(x^{\hat{i}}_{\ ,k,i} x^{\hat{j}}_{\ ,j} + x^{\hat{i}}_{\ ,k} x^{\hat{j}}_{\ ,j,i} \right) \delta_{\hat{i}\hat{j}} \tag{34}$$

When we add (33) and (34) and subtract (32), allowing for the fact that the order of partial derivatives may be interchanged, we find that

$$g_{kj,i} + g_{ik,j} - g_{ij,k} = \left(x^{\hat{i}}_{\ ,i,j} x^{\hat{j}}_{\ ,k} + x^{\hat{i}}_{\ ,k} x^{\hat{j}}_{\ ,j,i} \right) \delta_{\hat{i}\hat{j}} = 2\, x^{\hat{i}}_{\ ,i,j} x^{\hat{j}}_{\ ,k}\, \delta_{\hat{i}\hat{j}} \tag{35}$$

An older notation is

$$[i\, j, k] = \frac{1}{2} (g_{kj,i} + g_{ik,j} - g_{ij,k}) \ , \tag{36}$$

and, when look back at V-(28), we see that the Christoffel symbol can be written as

$$\Gamma^{\,r}_{ij} = g^{rk}\, [i\, j, k] = \frac{1}{2} g^{rk} \left(g_{kj,i} + g_{ik,j} - g_{ij,k} \right) \ . \tag{37}$$

In analogy to (31), we may write

$$g^{rk} = x^{r}_{\ ,\hat{\ell}}\, x^{k}_{\ ,\hat{m}} \delta^{\hat{\ell}\hat{m}} \ . \tag{38}$$

We now multiply (**??**) by g^{rk} to find that

$$\Gamma^{\,r}_{ij} = x^{r}_{\ ,\hat{\ell}}\, x^{k}_{\ ,\hat{m}} \delta^{\hat{\ell}\hat{m}} x^{\hat{i}}_{\ ,i,j} x^{\hat{j}}_{\ ,k}\, \delta_{\hat{i}\hat{j}} = x^{r}_{\ ,\hat{\ell}}\, \delta^{\hat{j}}_{\hat{m}} \delta^{\hat{\ell}\hat{m}} x^{\hat{i}}_{\ ,i,j}\, \delta_{\hat{i}\hat{j}} = x^{r}_{\ ,\hat{\ell}}\, x^{\hat{\ell}}_{\ ,i,j} \ . \tag{39}$$

So we learn that, indeed,

$$\frac{1}{2} g^{rk} \left(g_{kj,i} + g_{ik,j} - g_{ij,k} \right) = x^{r}_{\ ,\hat{\ell}}\, x^{\hat{\ell}}_{\ ,i,j} \ , \tag{40}$$

as advertised. And the argument goes through even when the hatted coordinates are not Cartesian.

So here is the dilemma: neither of the expressions (28), that is,

$$\delta a^k = -\Gamma^{\,k}_{ij} a^i d\, x^j, \tag{41}$$

nor

$$d\, a^i = a^{i}_{\ ,j}\, dx^j \tag{42}$$

(see (22)) is a vector.

3.3 The Covariant Derivative

Although the sum of tensor and a non-tensor is a non-tensor, the difference of two non-tensors might just turn out to be a tensor if the bad parts cancel. (There is no word that I know of for quantities that are not tensors, so I have improvised.) Thus we might try looking at

$$D\,a^i = d\,a^i - \delta\,a^i = \left(a^i_{\;,j} + \Gamma^{\,i}_{jk}\,a^k\right) d\,x^j \ . \tag{43}$$

This suggests defining what is called the covariant derivative,

$$a^i_{\;;j} = a^i_{\;,j} + \Gamma^{\,i}_{jk}\,a^k \tag{44}$$

This is in fact a mixed tensor of rank two. The argument for that is not given yet so that you will have a chance to try finding it for yourself.

A similar argument will lead you to the formula for the covariant derivative of covariant components, but there is a more pleasant way to find a suitable expression. Suppose you have the scalar

$$\phi = a^i b_i \tag{45}$$

where a^i and b_i are components of vectors. We have already seen that the derivative of a scalar is a covariant derivative (or differential form) and we would like the outcome of the derivative of this scalar product to behave in a like manner. We would also like differentiation of this product to satisfy the rule for differentiating product (Leibniz' rule), so we look at

$$\phi_{,j} = a^i_{\;,j}\,b_i + a^i\,b_{i,j} \tag{46}$$

This can be written as

$$\phi_{,j} = \left(a^i_{\;,j} + \Gamma^{\,i}_{jk}\,a^k\right)\ b_i + a^i\ \left(b_{i,j} - \Gamma^{\,k}_{ji}\,b_k\right) \tag{47}$$

This suggests the definition for the covariant derivative of a covariant vector:

$$b_{i;j} = b_{i,j} - \Gamma^{\,k}_{ji}\,b_k \tag{48}$$

for which the standard mnemonic is "co below and minus." Now we have the agreeable form

$$\left(a^i\,b_i\right)_{,j} = \left(a^i\,b_i\right)_{;j} = +b_i a^i_{\;;j} + a^i b_{i\;;j}, \tag{49}$$

which is a covariant vector.

For tensors, a like procedure will work. For the example of a second rank covariant tensor, look at $\left(T^{ij}a_ib_j\right)_{,j}$ and find the appropriate expression for $T^{ij}_{\;\;;k}$.

Before we go on to other (related) things we might recall that the real reason for all this complication is that the basis vectors change from place to place in general spaces or in flat space with non-Cartesian coordinates. If we have a vector

$$\mathbf{a} = a^i\mathbf{e}_i \tag{50}$$

where $\mathbf{e}_i$ are basis vectors, the partial derivative of $\mathbf{a}$ may be written out as

$$\mathbf{a}_{,j} = a^i{}_{,j}\mathbf{e}_i + a^i\mathbf{e}_{i,j} \tag{51}$$

The issue is then that we need to say what $\mathbf{e}_{i,j}$ is. If we suppose that this quantity (for a given fixed i) is a linear combination of the $\mathbf{e}_i$, as in

$$\mathbf{e}_{i,j} = \Gamma^{\,k}_{ij}\,\mathbf{e}_k \tag{52}$$

The coefficients Γ have to be figured out for this purpose, but the form is right. We have

$$\mathbf{a}_{,j} = \left(a^k{}_{,j} + \Gamma^{\,k}_{ij}\,a^i\right)\mathbf{e}_i \tag{53}$$

and, if we can (later) show that this is the same Γ as before, we have that

$$\mathbf{a}_{,j} = a^i{}_{;k}\mathbf{e}_i \tag{54}$$

where the covariant differentiation is seen to allow for the real changes with position as well as the changes in components because of changes in the basis vectors.

General Relativity

Installment VIII.

October 7, 2003

1 Homework

1.1 The Covariant Derivative

What we started to do in the last installment was to look at the introduction of the covariant derivative by way of the local flatness of the space. Suppose then that in an infinitesimal region around point P we may think in terms of the simple concepts of vectors and their bases. So a vector $\mathbf{a}$ can be written as in daily life as

$$\mathbf{a} = a^i\,\mathbf{e}_i \tag{1}$$

The coordinates at P are x^i and the coordinates at neighboring Q are $x^i + dx^i$. The change in $\mathbf{a}$ in going from P to Q is

$$d\mathbf{a} = a^i{}_{,j}\,dx^j\,\mathbf{e}_i + a^i d\mathbf{e}_i \tag{2}$$

Remember that

$$a_{i,j} = \frac{\partial a^i}{\partial x^j} \tag{3}$$

In (**??**) there are two terms. One represents the actual change in the vector as we go from P to Q. The other expresses the change in the component representation that results from possible changes in the basis vectors in going from P to Q. The latter term is not there when we are using Cartesian coordinates in a flat space. As we shall see, it is also not there in the flat space time of special relativity. (You may have read that cosmologists think that the universe is flat, but that is a rough statement and does not allow for the bumps and local warps in spacetime that make our lives so interesting.)

The issue is then that we need to say what $d\mathbf{e}_i$ is. Let us represent it as a linear combination of the basis vectors. The coefficients in this linear combination will depend on which $d\mathbf{e}_i$ we are dealing with; that is, the coefficient will carry the index i. We also expect that $d\mathbf{e}_i$ will be proportional to dx^j to good approximation. (More precisely, there will be higher order terms — quadratic and cubic and so on in dx^j — and for a more global study we need to go to the tangent space. But this local, linear treatment is adequate for our purposes. So we may write

$$d\mathbf{e}_i = \Gamma_{ij}^{\,k}\,\mathbf{e}_k\,dx^j \tag{4}$$

The coefficients Γ naturally have three indices, one for the sum over the dx^j, one for the combination of the $\mathbf{e}_k$ and one to tell us which $\mathbf{e}_i$ we are referring to. On using (**??**) we have

$$d\mathbf{a} = a^i_{\;,j}\,dx^j\,\mathbf{e}_i + a^i\Gamma_{ij}^{\,k}\,\mathbf{e}_k\,dx^j \tag{5}$$

With the renaming of a dummy index this is written as

$$d\mathbf{a} = \left(a^k_{\;,j} + \Gamma_{ij}^{\,k}\,a^i\right)\mathbf{e}_k\,dx^j \tag{6}$$

The covariant derivative is then defined as

$$a^k_{\;;j} = a^k_{\;,j} + \Gamma_{ij}^{\,k}\,a^i \tag{7}$$

For homework: Figure out what the $\Gamma_{ij}^{\,k}$ are within the context of what we have in this handout. Make use of the fact that

$$g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j \tag{8}$$

The manipulations that you need are exemplified in Installment VII, equations (32)-(37).

For homework in 4000: (a) Using results from the 3000 homework and the definition of the covariant derivative of a second rank covariant tensor given in class, show that $g_{ij:k} = 0$. (b) Try to show that the covariant derivative of a vector, say $a^i_{\;;j}$, is a mixed tensor of second rank. Do not worry if you don't get this one. Indeed look it up if you want to. But do learn how it is done.

General Relativity

Installment IX.

October 1, 2003

1 Geodesic Motion

We specify a parameterized curve by giving the coordinates of the points on the curve as functions of the parameter along the curve, say the distance on the curve from selected point. That is, we specify

$$x^i = x^i(s) \tag{1}$$

where s is the arclength along the curve. The components of the tangent vector to the curve at any point x^i are

$$v^i = \frac{d\,x^i}{ds} \tag{2}$$

and we may write the tangent vector as

$$\mathbf{v} = v^i \mathbf{e}_i \tag{3}$$

The derivative along the curve of the velocity is

$$\frac{d\,\mathbf{v}}{ds} = \frac{d\,v^i}{ds}\mathbf{e}_i + v^i \frac{d\,\mathbf{e}_i}{ds} \tag{4}$$

We saw (for example, in Installment VIII) that

$$d\mathbf{e}_i = \Gamma_{ij}^{\,k} \mathbf{e}_k d\,x^i \tag{5}$$

(However, the sign is different this time. Why is that?) We may then write that

$$\frac{d\,\mathbf{e}_i}{ds} = \Gamma_{ij}^{\,k} \mathbf{e}_k\, v^i \tag{6}$$

and so

$$\frac{d\,\mathbf{v}}{ds} = \left(\frac{d\,v^k}{ds} + v^i v^j \Gamma_{ij}^{\,k} \right) \mathbf{e}_k \tag{7}$$

If there is no external force, there is no acceleration, and we expect

$$\frac{d\,v^k}{ds} + v^i v^j \Gamma_{ij}^{\,k} = 0 \tag{8}$$

This is geodesic motion.

On the other hand, if we want to explain that the motion is not on straight lines, we might instead pretend that the deviation from Newton's first law is caused by an external force with a potential ϕ. We might guess that the force is

$$\nabla\phi = v^i v^j \Gamma_{ij}^{\,k} \mathbf{e}_k \tag{9}$$

The gradient operator is a vector operator and can be expressed as

$$\nabla = \mathbf{e}^i \frac{\partial}{\partial x^i} \tag{10}$$

Hence, if we dot $\mathbf{e}_\ell$ into (9), we get

$$\frac{\partial\phi}{\partial x^\ell} = v^i v^j \Gamma_{ij}^{\,k}\, g_{\ell k} \tag{11}$$

On the left we have first derivative of the potential. On the right we have a kind of first derivative of the metric tensor g_{ij}, for that what the Γ really is. So the (somewhat loose, but suggestive) conclusion is that we have motion that is like accelerated motion with the metric serving as the potential. The imagery however is totally different in the two cases. To complete the story, we need to replace the Newtonian outlook as a means of getting this 'potential.'

In Newtonian gravity, the potential is a scalar and is determined by the density of matter, but the metric is a tensor and we need to replace the rule for finding the metric as a result of the influence of mass and energy. But the rough idea for describing the motion is already seen here: if we can give a suitable metric, the geodesic equation can describe orbits in the way that Newton's second law does.

A complicating point is that in the term in (8) with Γ we have $\mathbf{v}$. But in fact, if s is arc length along the curve, $\mathbf{v}$ is a unit vector. That is, since

$$d\,s^2 = g_{ij} dx^i\, dx^j \tag{12}$$

we have

$$\mathbf{v}^2 = g_{ij} v^i\, v^j = 1. \tag{13}$$

This means that if the trajectory is specified, we know both Γ and $\mathbf{v}$ along it. But we don't in general know the trajectory and (8) is a sort of consistency condition to determine it. Things are really no simpler than in Newtonian mechanics and we don't even know how to pick the right g_{ij} yet.

However, if we multiply

$$\Gamma_{ij}^{\,k} = \frac{1}{2} g^{i\ell} \left(g_{\ell\, i,j} + g_{\ell\, j,i} - g_{i\, i,\ell} \right) \tag{14}$$

by $g_{k\, m}$ we find

$$g_{km} \Gamma_{ij}^{\,k} = \frac{1}{2} \left(g_{k\, i,j} + g_{k\, j,i} - g_{i\, i,k} \right) \tag{15}$$

If we interchange indices suitably and add the resulting expressions, we get a formula for the derivative of g_{ij} in terms of Γ (try this for yourself) and so can get a sort of equation for g_{ij} in terms of ϕ. It is not so easy to solve such equations, but you can see the emerging outline of the process from this.

General Relativity

Installment X.

October 6, 2003

1 Review of Coordinate Transformations

We require that the properties of spacetime and the contents of our statements of the physical laws should not depend on our choice of coordinate systems. We still do want to have the convenience of coordinates to formulate our statements about these things. So let us look again at the business of changing coordinates. (There is a nice discussion of this matter in Stephani's 'General Relativity.') Suppose that we have coordinates x^i, $i = 1, 2, ..., n$ and we wish to change to new coordinates, $x^{i'}$. The transformation can be indicated by

$$x^i = x^i(x^{i'}) \tag{1}$$

To simplify the discussion we may sometimes work with the coordinate differentials

$$dx^{i'} = \frac{\partial x^{i'}}{\partial x^i} dx^i \tag{2}$$

This is a set of simultaneous linear equations for the dx^i in terms of the $dx^{i'}$. The coefficients of the equations are in the transformation matrix

$$\Lambda^{i'}{}_i = \frac{\partial x^{i'}}{\partial x^i} \tag{3}$$

and we require that det $\left(\Lambda^{i'}{}_i\right) \neq 0$. For those not familiar with this condition in the solution of simulataneous linear equations a brief example may serve to explain or recall the situation.

Suppose that you have two simultaneous equations in the unknowns x and y,

$$Ax + By = a; \qquad\qquad Cx + DY = b. \tag{4}$$

This can be written as

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \tag{5}$$

Define the row vectors $\mathbf{A} = (A, B)$ and $\mathbf{C} = (C, D)$ and the column vector

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{6}$$

Then the simultaneous equations are

$$\mathbf{A} \cdot \mathbf{x} = a \qquad\qquad \mathbf{C} \cdot \mathbf{x} = b \tag{7}$$

Now observe that the determinant of the matrix can be written as $\mathbf{A} \times \mathbf{C}$ which has only one component (and not really in the plane either). So if the determinant did vanish, this would mean that $\mathbf{A}$ and $\mathbf{C}$ are colinear and that we could write that

$$\mathbf{A} = \lambda \mathbf{C} \tag{8}$$

where λ is a constant. Then, the equations (9) become

$$\lambda \mathbf{C} \cdot \mathbf{x} = a \qquad\qquad \mathbf{C} \cdot \mathbf{x} = b \tag{9}$$

These are compatible only if $a = \lambda b$, otherwise we cannot find a solution. That is why we require the condition $\det \left(\Lambda^{i'}{}_{i} \right) \neq 0$ for the differential transformations that we work with.

4000 homework for Oct. 13: Give a similar geometric argument for the three-dimensional case.

Going back to the inverse transformation, we have the transformation on differentials

$$dx^j = \frac{\partial x^j}{\partial x j'} dx^{j'} = \Lambda^j{}_{j'} dx j' \tag{10}$$

where

$$\Lambda^j{}_{j'} = \frac{\partial x^j}{\partial x j'} \tag{11}$$

Then

$$\Lambda^j{}_{j'} \Lambda^{j'}{}_k = \delta^j{}_k \ , \qquad\qquad \Lambda^{i'}{}_i \Lambda^i{}_{j'} = \delta^{i'}{}_{j'} \tag{12}$$

It may sometimes happen that you are given the differential transformation, but that there is no integrated transformation of the form of (1). Such a situation is called anholonomic. I doubt that you will ever see such thing here but I mention it because something like this arises in mechanics and in thermodynamics, and I wanted to get you a little prepared for that.

2 Spacetime

2.1 The Galilean Transformation

Figure 1 shows a section of Manhattan. The x-coordinate labels the Avenues and the y-coordinate labels the streets. The unit of length in x is the av and the unit of y is the st.

Our friends, points P and Q, are indicated on the figure. What is the distance between P and Q? To answer this question you measure Δx and Δy, the coordinate separations between the points. These are found to be 1.5 av and 1.5 st, respectively. But then how to compute the distance? The two coordinate differences are in different units and thus hard to combine into a single number such as distance. But, if I tell you that there are three sts to an av, you can do it. The answer is then $\sqrt{(3 \times 1.5)^2 + (1.5)^2} = 1.5\sqrt{10} \approx 4.7sts$. Evidentally, the two coordinates need to be in the same units if we are to use them effectively. The same is true in spacetime. These and related issues are nicely discussed in the Relativity book by Max Born and the book on Special Relativity by French and Wheeler.

In Figure 2, we show a picture of spacetime of an inertial observer. Only one spatial axis is shown and that is called x. The ordinate, t, is the time axis. This is the usual way of drawing spacetime diagrams. As with the streets and avenues, it is not practical to measure time and space in different units when working in spacetime, so we need a correction factor to convert one of them to the units of the other. Call that factor c and imagine that we have multiplied it into t. If c has the dimensions of speed, then t will have acquired the dimensions of length. In the new units, the value of c will have become unity. For now, c can be any conversion factor that is useful, though in relativity, the natural choice is the speed of light *in vacuo*. In that case, we are using light years instead of years to measure time (for example).

Suppose that there is another observer moving in the x-direction at constant velocity v. (If hey were moving in some other direction, we would have chosen that direction as the x-direction.) Heir time axis, t' is along heir worldline in spacetime, given by $x = vt$. This is arranged so that hey is stationary in heir own coordinate system. The Galileian transformation between heir system and ours is

$$t' = t \; ; \qquad\qquad x' = x - vt \tag{13}$$

as can be seen from the figure. This linear transformation of coordinates can be expressed in matrix form as

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -v & 1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \tag{14}$$

The inverse transformation is the same except that $-v$ is replaced by v so that

$$\begin{pmatrix} 1 & 0 \\ -v & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{15}$$

Those of you who have studied matrices will recognize that the matrix in (14) shears the medium and (14) may seem to you to be a rather strange looking transformation law. It may also be surprising that no one really thought about it much until Minkowski made explicit what was implied by the theory of relativity: space and time are one and ought to be thought of together in the manner of Figure 2.

In relativity, we know the metric of space time, but what is it in classical physics where we don't worry about exceeding the speed of light? In the Euclidean plane, the locus of points equidistant from the origin is a circle. In the Minkowskian x–t plane of relativistic spacetime, the locus of points equidistant from the origin is a hyperbola whose asymptotes are $t = \pm x$, as we shall see. But the classical case is not so nice. What we want from a metric is that it should produce an invariant object or scalar that we might associate with a line element or distance. Well, we do have $dt = dt'$, so that can serve as a line element. Then time is distance, and the import of this remark will take on more significance when we go to relativity. But clearly, the classical case is not so pretty as the relativistic one.

2.2 The Lorentz Transformation

It came as a surprise that Maxwell's equations for electricity and magnetism were not invariant under the Galilean transformation that we looked at in the previous subsection. Perhaps it was a bigger surprise that there was another transformation under which Maxwell's equations *a*re invariant. Theories were constructed to rationalize this finding but they did not really explain what was going on. Then Einstein cleared away the mystery by deriving the Lorentz transformation from a couple of simple postulates: that the laws of physics are the same in any inertial frame and that the speed of light is independent of the speed of its source. (Inertial frames are those that are not accelerated.) What this meant was that the Galilean transformation is not quite the right one between moving frames of reference. Rather it is only an approximation to the correct transformation and it is applicable only when the relative speeds of the moving frames have small magnitude.

In the case of two-dimensional space time, the Lorentz transformation is

$$t' = \frac{t - vx/c^2}{\sqrt{1-(v/c)^2}} \qquad\qquad x' = \frac{x - vt/c}{\sqrt{1-(v/c)^2}}) \tag{16}$$

where c is the speed of light (and when we say "speed of light," the qualification in a vacuum is generally understood). Now, we notice that when $|v|/c << 1$, the Lorentz transformation reduces to the Galilean transformation. Since speeds comparable to c were unknown before the twentieth century (except for light itself) it is not so surpising that the Galilean transformation was good enough.

In the Lorentz transformations, as expressed in (16), space and time have different units, so we want to fix that as we discussed in the previous paragraphs. For this we make the substitution $t \to ct$. In the new units, $c = 1$ and we may write the Lorentz transformation as

$$t' = \frac{t - vx}{\sqrt{1-v^2}} \qquad\qquad x' = \frac{x - vt}{\sqrt{1-v^2}} \tag{17}$$

In matrix form, the transformation is

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v \\ -\gamma v & \gamma \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \tag{18}$$

where $\gamma = \sqrt{1 - v^2}$.

A suggestive change of variables that is often made is

$$\cosh\chi = \gamma \qquad\qquad \sinh\chi = \gamma v. \tag{19}$$

The transformation then looks like this:

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh\chi & -\sinh\chi \\ -\sinh\chi & \cosh\chi \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \tag{20}$$

This version makes it very clear that the determinant of the transformation matrix is unity. A consequence of this (which you might like to verify) is that

$$t'^2 - x'^2 = t^2 - x^2 \tag{21}$$

That is, $t^2 - x^2$ is a scalar under transformation to a new inertial frame. This quantity looks a lot like the one in the theorem of pythagoras, except for the sign. In Minkowski spacetime, the distance from the origin of a point with spacetime coordinates t and x is $\sqrt{t^2 - x^2}$. The locus of points equidistant from the origin is a hyperbola. So spacetime is a world in which hyperbolic trignonmetry reigns.

One interesting property of spacetime that follows from this definition of distance is that the sum of the lengths of two sides of a triangle is *l*ess than the length of the third side. This feature of the geometry of spacetime gives rise to some results that may seem paradoxical at first glance. (The so-called twin paradox is a popular example of this.)

We should also be aware that, when we travel, in normal circumstances the spatial distance covered is measured by an odometer. But the spatio-temporal distance is measured by a chronometer, or clock. That is, in our own frames, we are simply moving forward in our own or *p*roper times designated as τ (though the less pretentious t' will do). As we travel through spacetime, the distances covered along our paths, or worldlnes, are recorded by our clocks. In your own frame, you have no spatial velocity and your speed in your time direction is $dt'/dt' = 1$ (remember though that $c = 1$). In the unprimed frame your velocity has components $dt/d\tau$ and $dx/d\tau$. These are the components of the vector tangent to your world line and its magnitude (your speed in spacetime) is always unity. You cannot stop moving through spacetime but if you move through space at the speed of light, you will stand still in time because your total speed must be unity. All this makes for an interesting dynamics.

3 Energy, Momentum and All That

Consider a particle moving through spacetime. As usual, we denote Its spatial coordinates as x^i. The time, t, we call x^0. In four-dimensional spacetime, we shall use Greek indices that run from 0 to 3. The spacetime coordinates of an event are thus x^μ with $\mu = 0, 1, 2, 3$. In special relativity (no gravity yet) we shall call the metric tensor in spacetime $\eta_{\mu]nu}$ and the line element is written as

$$d\tau^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{22}$$

From the considerations of the previous section, we have

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{23}$$

The spacetime displacement corresponding to the change in dx^μ is the change in the proper time of the displace object. The μ component of the corresponding velocity is

$$u^\mu = \frac{dx^\mu}{d\tau} \tag{24}$$

This is called the four velocity. According to (22),

$$\eta_{\mu\nu} u^\mu u^\nu = 1 \tag{25}$$

If the moving object has mass (or inertia) m, the momentum is

$$p^\mu = m u^\mu = m \frac{dx^\mu}{d\tau} \tag{26}$$

Then

$$\eta_{\mu\nu} p^\mu p^\nu = (p^0)^2 - \mathbf{p}^2 = m^2 \tag{27}$$

where $\mathbf{p}$ is the spatial momentum, or three-momentum, by contrast with p^μ, the four-momentum. The zeroth component of the four-momentum, p^0, is interpreted as the energy, E, of the object. Thus,

$$E^2 = \mathbf{p}^2 + m^2 \tag{28}$$

or

$$E = \pm\sqrt{m^2 + \mathbf{p}^2} \tag{29}$$

The two possible signs of the energy has great ramifications for quantum mechanics; it also has significance for classical (continuum) mechanics that has not been fully exploited as yet.

When the momentum is not large, as in ordinary life, we may approximate the energy as

$$E = m\left(1 + \frac{\mathbf{p}^2}{m^2}\right)^{\frac{1}{2}} = m\left(1 + \frac{\mathbf{p}^2}{2m^2} + ...\right) \tag{30}$$

(The minus sign has been suppressed for simplicity.) Since $\mathbf{p}^2/(2m^2)$ is the classical kinetic energy, we see that the energy of the object has two terms, the standard kinetic energy and m. With our scaling of t by c, we have a situation where energy and mass have the same units. But to give the mass–energy its normal look, we recall that a factor of c^2 $(= 1)$ is needed. Then, we find that the extra term in the energy (from the classical point of view) is the famous mc^2.

General Relativity

Installment XI.

October 13, 2003

1 The Differential

This is a subject that leads into some advanced topics that we shall not really get into, but as some people are not comfortable with differentials, it may be worth saying something about them here.

Consider a function f on the real line with coordinate x. The value of f at $x + dx$, where dx is a small increment in coordinate, is

$$f(x + dx) = f(x) + f'(x)\, dx + \frac{1}{2} f''(x)(dx)^2 + ... \tag{1}$$

where $f' = df/dx$ and so on. In the limit of small displacement, dx, we may approximate this as

$$df = f(x + dx) - f(x) = \frac{df}{dx}\, dx \tag{2}$$

In a space of dimension n with coordinates x^i, this generalizes to

$$df = \frac{\partial f}{\partial x^i}\, dx^i \tag{3}$$

One way to look at this is to think of the dx^i as a set of (reciprocal) basis vectors and of the $\partial f/\partial x^i$ as components of a (covariant) vector called the gradient of f. This image proved so appealing to mathematicians that they promoted df from its lowly status as an approximation with an infinitesimal character to being a finite object in its own right. It is a pity that a more compelling notation was not introduced with this elevation, though in some variants on the notation various doodads are put on the d, as in $\tilde{d}$. What matters most about this object though is that when you see something like $\tilde{d}f$, this thing is not necessarily integrable in the way that $(\partial f/\partial x)dx$ integrates up to f. That is, in the expression $a_i dx^i$, a_i

is not necessarily the derivative of anything. This feature of the generalization of differentials to forms is of fundamental importance in thermodynamics and it also plays an important role in some aspects of classical mechanics. In a less mathematical way, we might write the gradient operator in terms of more commonplace looking basis vectors as $\nabla = \mathbf{e}^i(\partial/\partial x^i)$. When applied to a function f, this is

$$\nabla f = \mathbf{e}^i \frac{\partial f}{\partial x^i} \tag{4}$$

Recall that we have already said that the derivative along a curve parameterized by s is

$$\frac{d}{ds} = \frac{dx^i}{ds}\frac{\partial}{\partial x^i} \tag{5}$$

This is also called the derivative along the tangent vector $v^i = dx^i/ds$. As we have seen, the operator d/ds is a vector with components v^i and basis vectors $\partial/\partial x^j$. If we think of the latter as $\mathbf{e}_i$ we may write

$$\mathbf{e}^i \cdot \mathbf{e}_j = dx^i[\frac{\partial}{\partial x^j}] = \delta^i_{\cdot j} \tag{6}$$

This is what underlies the distinction between covariant and contravariant, but we are not going into these more advanced topics. What we do need to appreciate though is that one may speak of the derivative along any given vector, $\mathbf{a}$. This is

$$\mathbf{a} \cdot \nabla = (a^i \mathbf{e}_i) \cdot (\mathbf{e}^j \frac{\partial}{\partial x^j}) = a^i \frac{\partial}{\partial x^i} \tag{7}$$

This will also work for the directional derivative of a vector.

The derivative of $\mathbf{b}$ along $\mathbf{a}$ is

$$\nabla_{\mathbf{a}}\mathbf{b} = \nabla_{(a^i \mathbf{e}_i)}(b^j \mathbf{e}^j) = a^i \nabla_{\mathbf{e}_i}(b^j \mathbf{e}_j) \tag{8}$$

As before, the derivative along $\mathbf{e}_i$ is the derivative in the x^i direction. Taking the derivative of the product, we obtain

$$\nabla_{\mathbf{a}}\mathbf{b} = a^i(\mathbf{e}_j b^j_{,i} + b^j \nabla_{\mathbf{e}_i}\mathbf{e}_j) \tag{9}$$

(The typesetting is not so successful here — $\mathbf{e}_i$ is a subscript on ∇ but the $\mathbf{e}_j$ is not.) From past experience, we surmise that with a little work we would find that

$$\nabla_{\mathbf{e}_i}\mathbf{e}_j = \Gamma^{\;k}_{ij}\mathbf{e}_k \tag{10}$$

where the Γ will turn out to be just what the notation suggests. Thus,

$$\nabla_{\mathbf{a}}\mathbf{b} = a^i b^k_{;i}\mathbf{e}_k \tag{11}$$

2 The Metric Tensor as Potential

2.1 Four-Velocity

Now let us go into an inertial frame with time coordinate $x^0 = t$ and the space coordinates x^i, $i = 1, 2, 3$, as usual. The distance between two neighboring points is $d\tau$ where $d\tau^2 = \eta_{\mu\nu} dx^\mu dx^\nu$. ($d\tau^2$ means $(d\tau)^2$.)

A particle moves through the spacetime at constant spatial velocity $\mathbf{v}$ and in its frame its proper (or personal) coordinates are $x^{0'} = t'$ and $x^{i'} = 0$. That is, $x^{\mu'} = (t', \mathbf{0})$ where $\mu = 0, 1, 2, 3$. The Lorentz transformation (with $c = 1$) is

$$t = \gamma(t' + vx') \qquad\qquad x = \gamma(x' + vt') \tag{12}$$

The particle is assumed to move in the x-direction with $x^1 = x$. The time-component of the velocity of the particle is

$$\frac{dt}{dt'} = \gamma \tag{13}$$

since t' is the proper time of the particle. Here, $\gamma = 1/\sqrt{1 - \mathbf{v}^2}$and the components of $\mathbf{v}$ are dx^i/dt so that

$$\frac{dx^i}{dt'} = \gamma v^i = u^i \tag{14}$$

We can also write this as

$$u^\mu = \left(\frac{1}{\sqrt{1 - \mathbf{v}^2}}, \frac{\mathbf{v}}{\sqrt{1 - \mathbf{v}^2}} \right) \tag{15}$$

As we can see,

$$\eta_{\mu\nu} u^\mu u^\nu = 1 \tag{16}$$

Another way to say this is provided by the introduction of

$$u_\mu = \eta_{\mu\nu} u^\nu = (u^0, -\mathbf{v}) \tag{17}$$

so that

$$u_\mu u^\mu = 1 \tag{18}$$

2.2 Geodesic Motion

We need not repeat the derivation of the equation of motion along geodesics since the derivation is as before except for some notational changes: we replace Latin indices by Greek ones, and now use τ, the spacetime distance along the path, instead of the former s. Then we have

$$\ddot{x}^\mu = -\Gamma^{\ \mu}_{\rho\sigma} \dot{x}^\rho \dot{x}^\sigma \tag{19}$$

where $\dot{x} = dx/d\tau$. (τ and t' are being used interchangeably in this example.) We may write (19) out as

$$\ddot{x}^\mu = -\Gamma^{\,\mu}_{00}\, u^0 u^0 - 2\Gamma^{\,\mu}_{0i}\, u^0 u^i - \Gamma^{\,\mu}_{ij}\, u^i u^j \tag{20}$$

But since $u^0 = \gamma$ and $u^i = \gamma v^i$,

$$\ddot{x}^\mu = -(u^0)^2 \left(\Gamma^{\,\mu}_{00} - 2\Gamma^{\,\mu}_{0i} v^i - \Gamma^{\,\mu}_{ij} v^i v^j\right) \tag{21}$$

Since c is the unit of speed, we may assume that $|\mathbf{v}| << 1$ in the case of the mild motions we are used to, so that

$$\ddot{x}^\mu \approx -(u^0)^2\, \Gamma^{\,\mu}_{00} \approx -\Gamma^{\,\mu}_{00} \tag{22}$$

We need to compute $\ddot{x}^i$ to compare with Newton's second law to get an idea of what replaces it.

We know that

$$\Gamma^{\,\mu}_{00} = \frac{1}{2} g^{\mu\beta} \left(g_{0\beta,0} + g_{0\beta,0} - g_{00,\beta}\right) \tag{23}$$

In the limit where the space is not very warped, we expect the Minkowski metric to be a reasonable approximation so we write

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{24}$$

where $h_{\mu\nu}$ is a small perturbation on the flatspace Minkowski metric. The inverse metric should be of the form

$$g^{\mu\nu} = \eta^{\mu\nu} + f^{\mu\nu} \tag{25}$$

since $\eta_{\mu\rho}\eta^{\rho\nu} = \delta_\mu^{\ \nu}$. We then require that

$$g_{\mu\rho} g^{\rho\nu} = (\eta_{\mu\rho} + h_{\mu\rho})(\eta^{\rho\nu} + f^{\rho\nu}) = \delta_\mu^{\ \nu} \tag{26}$$

When we write out the indicated product, we get

$$g_{\mu\rho} g^{\rho\nu} = \delta_\mu^{\ \nu} + f_\mu^{\ \nu} + h_\mu^{\ \nu} + ... = \delta_\mu^{\ \nu} \tag{27}$$

Then we conclude that

$$f_\mu^{\ \nu} = -h_\mu^{\ \nu} \tag{28}$$

We neglect terms quadratic in h and, since $\eta_{\mu\nu}$ is a constant,

$$\Gamma^{\,i}_{00} = \frac{1}{2}\eta^{i\beta}(2h_{0\beta,0} - h_{00,\beta}) = \frac{1}{2}\delta^{ij} h_{00,j} \tag{29}$$

And so,

$$\ddot{x}^i = -\frac{1}{2}\frac{\partial h_{00}}{\partial x^j}\delta^{ij} \tag{30}$$

On comparison of this with the Newton's second law,

$$\ddot{x}^i = -\frac{\partial \phi}{\partial x^j}\delta^{ij} \tag{31}$$

we conclude that in the classical limit

$$h_{00} = 2\phi \tag{32}$$

where ϕ is the potential of an imposed force. Then, since $g_{00} = \eta_{00} + h_{00}$, we have that

$$g_{00} = 1 + 2\phi \tag{33}$$

We had already seen that in Galilean spacetime, a suitable line element is dt. Now we find that the effect of an external potential is to produce a slight warping through g_{00}. More importantly, we have the clue that the metric tensor replaces the gravitational potential when we ascribe the influence of gravity to an effect of geometry rather than to a force.

General Relativity

Installment XII.

October 13, 2003

1 Topics in Tensor Calculus

1.1 Parallel Transport

There seem to remain questions among you about the meaning of covariant differentiation. So let us look another facet of this process. Remember though that if you want to differentiate a vector field, you have in effect to look at the difference in the field at two points that are close by, separated by coordinate differentials, dx^i. To get that difference you need to move one of the vectors over to the other to make the comparison. This is no problem in flat space but it is more subtle in curved space. So we need to define what we mean by moving a vector from place to place in a manner parallel to itself. To express this in formulae, we need to be aware that components of a vector will change as we move it, not only because the space is curved but also because the basis vectors typically change. And even if you have a good definition of parallel transport, when you take a vector from one point to another in a curved space, keeping it parallel to itself as you go, you may wind up with a different end product depending on the path you take.

Recall that the absolute differential of a vector **a** is, in terms of components,

$$Da^i = da^i + \Gamma^{\,i}_{jk} a^j dx^k = a^i{}_{,k} dx^k + \Gamma^{\,i}_{jk} a^j dx^k \tag{1}$$

where the comma denotes partial differentiation as usual. The second term takes care of the differences in the basis vectors at two points. The derivative along a curve $x^i = x^i(s)$ is then

$$\frac{Da^i}{Ds} = a^i{}_{,k} \frac{dx^k}{ds} + \Gamma^{\,i}_{jk} a^j \frac{dx^k}{ds} = \frac{da^i}{ds} + \Gamma^{\,i}_{jk} a^j \frac{dx^k}{ds} \tag{2}$$

where Ds and ds are really the same thing, the difference in s between the neighboring points. When you move **a** along the curve so that

$$\frac{da^i}{ds} + \Gamma^{\,i}_{jk} a^j \frac{dx^k}{ds} = 0 \tag{3}$$

then you are keeping it parallel to itself as you go. To find what the vector is at any point once it has been transported in this way, you need to solve the indicated differential equations. This is doable since the problem is linear, but it may take some work.

Now suppose that the vector field that you want to transport along the curve is its tangent vector, $\mathbf{v}$. Then the condition of parallel transport is

$$\frac{dv^i}{ds} + \Gamma^{\ i}_{jk} v^j v^k \ = 0 \tag{4}$$

where we need to recall that $v^i = dx^i/ds$. But this is the equation for a geodesic and so we conclude that, along a geodesic, the tangent vector is parallel to itself. Look back at equation (11) of Installment XI. You should convince yourself that the geodesic equation can be written as

$$\nabla_{\mathbf{v}} \mathbf{v} = \mathbf{0} \tag{5}$$

Along a geodesic, the tangent vector moves parallel to itself. This is a property of straight lines in flat space and that is why the geodesics are considered the natural generalization to curved space of straight lines. If time permits, we shall see how these things can be used in some problems. But now we must go on to the next step that is needed to set up the Einstein equations.

1.2 Second Covariant Derivative

We saw that the metric is like a potential. In the classical theory of gravity, the potential is derived from the Poisson equation

$$\nabla^2 \phi = 4\pi\, G\rho \tag{6}$$

where ρ is the density of matter and G is Newton's gravitational constant. To extend this theory into relativity, we shall have to deal with second derivatives. So let us bite the bullet.

We have a four-vector a — no boldface for four-vectors in this game — with covariant components a_λ (covariant for ease of typing). What is its second derivative? Staying in spacetime with Greek indices, we write

$$a_{\lambda;\mu;\nu} = (a_{\lambda;\mu})_{;\nu} \tag{7}$$

Since $a_{\lambda;\mu}$ is a second rank covariant tensor, we recall the rule 'co below and minus.' Then

$$a_{\lambda;\mu;\nu} = (a_{\lambda;\mu})_{,\nu} - \Gamma^{\ \rho}_{\nu\mu} a_{\lambda;\rho} - \Gamma^{\ \rho}_{\nu\lambda} a_{\rho;\mu} \tag{8}$$

As we recall, a Christoffel symbol is needed for each index.

Slogging on, we get the second covariant derivative

$$a_{\lambda;\mu;\nu} = (a_{\lambda,\mu} - \Gamma^{\ \rho}_{\lambda\mu}\, a_\rho)_{,\nu} - \Gamma^{\ \rho}_{\nu\mu}(a_{\lambda,\rho} - \Gamma^{\ \sigma}_{\lambda\rho} a_\sigma) - \Gamma^{\ \rho}_{\nu\lambda}(a_{\rho,\mu} - \Gamma^{\ \sigma}_{\rho\mu} a_\sigma) \tag{9}$$

$$= a_{\lambda,\mu,\nu} - \Gamma^{\ \rho}_{\lambda\mu} a_{\rho,\nu} - \Gamma^{\ \rho}_{\mu\nu} a_{\lambda,\rho} - \Gamma^{\ \rho}_{\nu\lambda} a_{\rho,\mu} - \left[a_\rho \Gamma^{\ \rho}_{\lambda\mu,\nu} - \Gamma^{\ \rho}_{\nu\mu}\, \Gamma^{\ \sigma}_{\lambda\rho} a_\sigma - \Gamma^{\ \rho}_{\nu\lambda} \Gamma^{\ \sigma}_{\rho\mu} a_\sigma\right] \tag{10}$$

1.3 The Commutator of Covariant Derivatives

When you look at the expression for the second covariant derivative, you may not be surprised to learn that, unlike partial derivatives, covariant derivatives do not commute: the order in which you take the derivatives matters. When you repeat the differentiations in the other order, you get

$$a_{\lambda;\nu;\mu} = a_{\lambda,\nu,\mu} - \Gamma^{\,\rho}_{\lambda\nu} a_{\rho,\mu} - \Gamma^{\,\rho}_{\nu\mu} a_{\lambda,\rho} - \Gamma^{\,\rho}_{\mu\lambda} a_{\rho,\nu} - \left[a_\rho \Gamma^{\,\rho}_{\lambda\nu,\mu} - \Gamma^{\,\rho}_{\mu\nu} \Gamma^{\,\sigma}_{\lambda\rho} a_\sigma - \Gamma^{\,\rho}_{\mu\lambda} \Gamma^{\,\sigma}_{\rho\nu} a_\sigma\right] \tag{11}$$

The effect of changing the order of differentiation is seen in the evaluation of $a_{\lambda;\mu;\nu} - a_{\lambda;\nu;\mu}$. This is the difference of two third rank covariant tensors and, amazingly enough, it is expressible as a linear combination of the components of a. The intermediate steps are easy enough to work out but too difficult to type, so here is the result:

$$a_{\lambda;\mu;\nu} - a_{\lambda;\nu;\mu} = R^{\rho}_{\cdot\lambda\mu\nu} a_\rho \tag{12}$$

where

$$R^{\rho}_{\cdot\lambda\mu\nu} = -\Gamma^{\,\rho}_{\lambda\mu,\nu} + \Gamma^{\,\rho}_{\lambda\nu,\mu} + \Gamma^{\,\sigma}_{\nu\lambda} \Gamma^{\,\rho}_{\sigma\mu} - \Gamma^{\,\sigma}_{\mu\lambda} \Gamma^{\,\rho}_{\nu\sigma} \tag{13}$$

is called the Riemann tensor or the curvature tensor.

1.4 The Transition from Newton

We saw that the orbits produced by an external force — at least one coming from a potential — could be duplicated by geodesics in a suitably curved space. In the Newtonian limit, the appropriate geometry was given a metric with $g_{00} = 1 + 2\phi$ where ϕ is the potential. When we studied the motion of a particle in a freely falling elevator, we also saw that the spatial variations in the force are significant. In the elevator, the displacement, ξ^i, of a particle from the center satisfied the equation

$$\ddot{\xi}^i = -E_{ij}\xi^j \tag{14}$$

where

$$E_{ij} = \frac{\partial\phi}{\partial x^i \partial x^j} \tag{15}$$

We also discussed the similarity of this equation to the equation for geodesic deviation on $\mathcal{S}^2$. To see how this idea really works, we need the formula for geodesic deviation in a curved space. The derivation of the appropriate formula is too long for this discussion, so we shall simply state the results. For a nice discussion of these matters, I would recommend the book 'General Relativity and Cosmology' by Jayant Narlikar. (I assume that you are reading in Kenyon as well and have looked at Fig. 3.8 on p. 33. You ought to aim to have read the first six chapters by the midterm since much of this is covered in the notes as well.)

Consider geodesics diverging from a point P. Along each of these geodesics, the distance from P is s. For s not too large, consider a pair of geodesics not too far from each other.

Suppose that Q_1 on one of the geodesics is at the same distance from P as is point Q_2 on the other. Let Q_1 be at x^i and Q_2 be at $x^i + \xi^i$. Then it is possible to establish that

$$\frac{d^2\xi^i}{ds^2} = R^i_{jk\ell} v^j v^k \xi^\ell \tag{16}$$

where $\mathbf{v}$ with components v^i is the tangent vector on the first geodesic. What Einstein did was to use the curvature tensor to construct a suitable second rank tensor to serve the same purpose as E_{ij}. That is, he replaced E^i_j by $R^i_{k\ell j} v^j v^k$. Thus he captured not only the basic effect of the equivalence between uniform accelerations and uniform gravity fields, but also could describe the basic differential gravitational forces. Then he needed to give a method for determining the curvature tensor.

Now

$$\mathrm{Trace}\,\mathbb{E} = E^i_{\ i} = \triangle\phi = 0, \tag{17}$$

where $\triangle = \nabla^2$ is the Laplacian. Hence the trace of the surrogate for E_{ij} is going to vanish and this statement replaces Laplace's equation for the gravitational potential. That is a good place to start with the new equation that is to replace Poisson's equation.

Queries

What is a null geodesic?

What is there that we have discussed so far that suggests that nothing can go faster than the speed of light?

When you see the astronauts in orbit on the tv, you observe that objects in the space ship seem to float in the air. Why is that?

Do covariant derivatives commute in flat space?

General Relativity

Installment XIII.

October 15, 2003

1 The Stress-Energy Tensor

1.1 Conservation Laws

Let $\mathcal{D}$ be the density of some stuff (scalar, vector or tensor). The total amount of that stuff in a given volume $\bullet$ is $\int_\bullet \mathcal{D} dV$ and the conservation law for the stuff is

$$\frac{\partial}{\partial t}\int_\bullet \mathcal{D} dV = -\int_\circ (\mathcal{D}\mathbf{v})\cdot \mathbf{n} dS + \int_\bullet \mathcal{S} dV \tag{1}$$

where $\mathbf{v}$ is the flow velocity of the stuff. The first integral on the left is a surface integral and the outward normal on the surface is $\mathbf{n}$. The quantity $\mathcal{S}$ is the net rate of creation and destruction of the stuff per unit volume.

The volume is fixed, so we can move the time derivative under the integral. We use the divergence theorem on surface integral and we obtain

$$\int_\bullet \left[\frac{\partial}{\partial t}\mathcal{D} + \nabla\cdot(\mathcal{D}\mathbf{v}) - \mathcal{S}\right] dV = 0 \tag{2}$$

Since the volume is arbitrary, this implies that

$$\frac{\partial}{\partial t}\mathcal{D} + \nabla\cdot(\mathcal{D}\mathbf{v}) = \mathcal{S} \tag{3}$$

1.2 Mass and Momentum

If the conserved substance is mass, whose density we denote as ρ, we write

$$\frac{\partial}{\partial t}\rho + \nabla\cdot(\rho\mathbf{v}) = 0 \tag{4}$$

since we are not allowing matter to be created or destroyed.

The momentum density is $\rho\mathbf{v}$ and it is generated by a force per unit volume as in

$$\frac{\partial}{\partial t}(\rho\mathbf{v}) + \nabla\cdot(\rho\mathbf{v}\mathbf{v}) = \mathbf{force}/\text{volume} \tag{5}$$

The force on a fluid or other continuous medium, besides external body forces, comes from the pressure. A pressure force on an element of fluid is the pressure difference across it. For example, in flat space with Cartesian coordinates, the pressure difference in the x-direction across a gap of dx is

$$p(x) - p(x+dx) \approx -\frac{\partial p}{\partial x} \tag{6}$$

Since pressure is force per unit area, the force on a small element of fluid with transverse area $dydz$ and thickness dx is $-(\partial p/\partial x)dx\,dy\,dz$ so the pressure force per unit volume in the x-direction is $-(\partial p/\partial x)$. Then the equation of momentum conservation is

$$\frac{\partial}{\partial t}(\rho\mathbf{v}) + \nabla\cdot(\rho\mathbf{v}\mathbf{v}) + \nabla p = \mathbf{0}. \tag{7}$$

These are the equations of motion of the perfect fluid. You need not attempt to follow this derivation; it is intended only that you should get an impressionistic feeling for it.

1.3 Preparing for the Jump to Relativity

Let $x^0 = t$ and let the spatial coordinates x^i be Cartesion for the first steps in the following rearrangements. Let us introduce a quantity v^0 whose numerical value is unity. Then the equation mass conservation may be written as

$$\frac{\partial(\rho v^0)}{\partial x^0} + \frac{\partial(\rho\, v^i)}{\partial x^i} = 0 \tag{8}$$

where v^i are the components of $\mathbf{v}$. Now let $\mu = 0, 1, 2, 3$. The conservation equation for mass becomes

$$(\rho v^\mu)_{,\mu} = 0 \tag{9}$$

For the momentum equation, we proceed along similar lines. We rewrite (7) as

$$\frac{\partial(\rho v^i v^0)}{\partial x^0} + \frac{\partial(\rho v^i v^j)}{\partial x^j} + \delta^{ij}\frac{\partial p}{\partial x^j} = 0. \tag{10}$$

This can be written as

$$\frac{\partial(\rho v^\mu v^i)}{\partial x^\mu} + \delta^{ij}\frac{\partial p}{\partial x^j} = 0 \tag{11}$$

or

$$\left(\rho v^\mu v^i\right)_{,\mu} + \delta^{ij}\frac{\partial p}{\partial x^j} = 0 \tag{12}$$

If we then rewrite (9) as

$$\left(\rho v^{\mu} v^{0}\right)_{,\mu} = 0 \tag{13}$$

we see that (13) and (12) together have a nice look. But what to do with the pressure term?

Introduce a quantity $h^{\mu\nu}$ such that $h^{ij} = \delta^{ij}$ with $h^{00} = 0$ and $h^{i0} = 0$. We'll see more about this in a minute, but first notice that (13) and (12) combine into

$$\left(\rho v^{\mu} v^{\nu}\right)_{,\mu} + h^{\mu\nu} \frac{\partial p}{\partial x^{\mu}} = 0 \tag{14}$$

The quantity $h^{\mu\nu}$, is called a projection operator. In this case, it projects onto the spatial components of a vector. In order to have the pressure term fit into the rest of the scheme, we really want to write it as something like $(h^{\mu\nu} p)_{,\mu}$. Indeed, it fits in fairly well if we write $h^{00} = v^0 v^0 - \eta^{00}$. This of course all seems off the wall. In fact it comes from long experience with these basic equations and some educated guesses about what the relativistic generalization must be. In going over this sort of thing carefully, you may get a strong impression about the transition between the transition between classical and relativistic theory. Though it must be admitted that it is easier to see through in the other direction, you need a good familiarity with the tensor calculus in that case.

Finally, we can write the equations for the conservation of mass and momentum in a classical perfect fluid as

$$T^{\mu\nu}_{\ \ ,\mu} = 0 \tag{15}$$

where

$$T^{\mu\nu} = \rho v^{\mu} v^{\nu} + h^{\mu\nu} p \tag{16}$$

This version of the equations has nothing to do with relativity except that it was constructed with relativity in mind.

1.4 Relativistic Fluid Dynamics

We need now to compare $v^i = (v^0, \mathbf{v})$, where $v^0 = 1$, to the four velocity. In fact, the v^i are just the conventional space velocity components, given by dx^i/dt. But in the frame of the moving observer (particle, or whatever) the four velocity is

$$u^{\mu} = \frac{dx^{\mu}}{d\tau} \tag{17}$$

where τ is the proper time. Thus, $u^{\mu} = \gamma v^{\mu}$ where $\gamma = \sqrt{1 - \mathbf{v}^2}$. For $\mathbf{v}^2 << 1$ — the classical limit — u^{μ} and v^{μ} are equal to good approximation and we have no hesitation in writing (9) as

$$\left(\rho u^{\mu}\right)_{,\mu} = 0 \tag{18}$$

and so on. For special relativity, we surmise then that with u^μ as the four-velocity field of the fluid, the full equations of motion are

$$T^{\mu\nu}_{\ \ ,\mu} = 0 \tag{19}$$

where now

$$T^{\mu\nu} = \rho u^\mu u^\nu + h^{\mu\nu}\, p \tag{20}$$

If you are moving with the fluid, you perceive a four-velocity $(1, 0, 0, 0)$; that is, your motion is entirely through time. We want this equation to be equally valid in all inertial frames, so $T^{\mu\nu}$ had better be a tensor. But a careful examination of the $h^{\mu\nu}$ as we have so far defined it, shows that it will not do. So having dared thus far, we generalize it to

$$h^{\mu\nu} = u^\mu u^\nu - \eta^{\mu\nu} \tag{21}$$

This expression boils down to the former h in the frame moving with the fluid, so it is a sensible generalization. It is the simplest tensor extension that has been thought of and there are other reasons for adopting it, but you could not be blamed for regarding it with suspicion.

By now, you have guessed the next step: we go into curved space and maintain our demand that our equation be expressed as a tensor. The conservation of mass and momentum should be expressed as

$$T^{\mu\nu}{}_{;\mu} = 0 \tag{22}$$

Of course, this being relativity, energy is also involved and this we see when we reexpress the stress (or stress-energy) tensor as

$$T^{\mu\nu} = (p + \rho) u^\mu u^\nu - \eta^{\mu\nu}\, p \tag{23}$$

Since p is a measure of the internal energy of the fluid (that is the kinetic energy of its constituent particles) we see that it contributes to the mass and so the kinetic energy term has become $(p+\rho)u^\mu u\nu$. But we must skip all that this entails as it would going too far afield to get into the thermodynamic issues. Suffice it to say that $T^{\mu\nu}$ is what Einstein decided as the surrogate for ρ in his generalization of the Poisson equation to the case of relativity.

General Relativity

Installment XIV.

October 20, 2003

1 The Einstein Equation

In his book, A First Course in General Relativity', Bernard Schutz gives these useful guidlines:

"If two tensors of the same type have equal components in a given basis, they have equal components in all bases and are said to be identical (or equal, or the same). In particular, if a tensor's components are all zero in one basis they are zero in all, and the tensor is said to be zero.

If an equation is formed using components of tensors combined only by the permissible tensor operations, and if the equation is true in one basis, then it is true in any other. This is a very useful result. It comes from the fact that the equation ... is simply an equality between components of two tensors of the same type, which [by the preceding remark] is then true in any system."

A fluid with four-velocity u^μ and stress-energy tensor $T^{\mu\nu}$ has a rest frame density given by the scalar $u_\mu u_\nu T^{\mu\nu}$. We could try to generalize Poisson's equation on the basis of this information, but that would be giving preference to the rest frame of the fluid. We want to make all frames equal (though it is sometimes true that some frames are more equal than others). So we shall seek a tensor formulation of the basic equation of gravity with $T^{\mu\nu}$ as the source of the gravitational field in the classical (nonrelativistic) limit. On the left side of Poisson's equation we can try to replace the Laplacian of the potential by second derivatives of the metric tensor. So our thoughts naturally turn to the curvature tensor, a beast with four indices. But $T^{\mu\nu}$ has only two, so a little trimming down needs to be done.

As we saw in Installment XII, the curvature tensor is

$$R^{\kappa}_{\cdot\lambda\mu\nu} = -\Gamma^{\ \kappa}_{\lambda\mu,\nu} + \Gamma^{\ \kappa}_{\lambda\nu,\mu} + \Gamma^{\ \sigma}_{\nu\lambda}\Gamma^{\ \kappa}_{\sigma\mu} - \Gamma^{\ \sigma}_{\mu\lambda}\Gamma^{\ \kappa}_{\nu\sigma} \tag{1}$$

Since the Christoffel symbols involve derivatives of $g_{\mu\nu}$, the Riemann tensor is made up of second derivatives. It is closely connected with the curvature of space. Indeed, it can be shown that a necssary and sufficient condition for the space to be flat is that all the components of the Riemann tensor should be zero. That may sound like a lot of components since there are $4^4 = 256$ components of the Riemann tensor. But because of the symmetries of this tensor, many of the components are not independent. In fact, as we see on inspection,

$$R_{\kappa\lambda\mu\nu} = -R_{\lambda\kappa\mu\nu} = -R_{\kappa\lambda\mu\nu} = R_{\mu\nu\kappa\lambda} \tag{2}$$

When we work through the combinatorics implied here, we find that, in four dimensions, $R_{\kappa\lambda\mu\nu}$ has a mere twenty independent components.

We need a second rank tensor that carries much of the geometrical information about the space we are working in. So a natural thing to try would be a contracted form of the Riemann tensor such as the Ricci tensor

$$R_{\mu\nu} = R^{\kappa}_{\cdot\mu\kappa\nu} \tag{3}$$

It is natural to wonder how the choice was made to contract that particular pair of indices. If you try contracting with the first index and any of the other three, you will find that you get the same result because of the symmetry. Another pleasant exercise is to show that

$$R_{\mu\nu} = R_{\nu\mu} \tag{4}$$

This is a good thing since the stress tensor is also symmetric in its two indices.

Now, one's first thought might be to try setting the Ricci proportional to the stress tensor. But by our construction of the stress tensor, it has a zero divergence:

$$T^{\mu\nu}{}_{;\nu} = 0 \tag{5}$$

The divergence of $R_{\mu\nu}$ does not vanish. But Einstein realized that the divergence of

$$G^{\mu\nu} = R^{\mu\nu} - Rg^{\mu\nu} \tag{6}$$

is zero, where

$$R = R^{\mu}_{\mu} \tag{7}$$

This contraction of the Ricci is called the curvature scalar; $G^{\mu\nu}$ is the famous Einstein tensor. Einstein's generalization of the Poisson equation is then

$$G^{\mu\nu} = \kappa T^{\mu\nu} \tag{8}$$

where κ is a constant whose value needs to be specified. It is basically the Newton constant G; we shall see in the weak field limit that there is also a factor or 4π that comes in.

2 The Schwarzschild Solution

Immediately after the Einstein equations were published, Karl Schwarzschild gave the solution for the metric around a spherical mass. This has been the starting point of much of the excitement in the subject to do with strong gravitational fields. In fact, the interest in this sort of thing actually goes back to the end of the 18th century when John Michel in England Pierre Laplace in France wondered about the possible existence of bodies whose escape velocities were greater than the speed of light. Such bodies would be dark, they reckoned, and would be detectable only through their obscuration of stars. Thus they were, in effect, thinking about Newtonian black holes over two hundred years ago.

The Einstein equations are not so easy to solve as Newton's. Instead of attacking them frontally, one normally looks for metrics with various simple properties and, by putting them into the equations, derives the full solution. In this way, Schwarzschild imagined a sphere of mass m whose center is at the origin of spherical coordinates, r, θ, ϕ such that

$$x = r\sin\theta\cos\phi \; ; \qquad y = r\sin\theta\sin\phi \; ; \qquad z = r\cos\theta \tag{9}$$

The coordinate differentials are

$$dx = dr\sin\theta\cos\phi + r\cos\theta\cos\phi\,d\theta - r\sin\theta\sin\phi\,d\phi \tag{10}$$

$$dy = dr\sin\theta\sin\phi + r\cos\theta\sin\phi\,d\theta - r\sin\theta\cos\phi\,d\phi \tag{11}$$

$$dz = dr\cos\theta - r\sin\theta\,d\theta \tag{12}$$

From this we find that

$$dx^2 + dy^2 + dz^2 = dr^2 + dr^2 + r^2\sin^2\theta\,d\phi^2 \tag{13}$$

We recognize the differential of solid angle

$$d\Omega = d\theta^2 + \sin^2\theta d\phi^2 \tag{14}$$

Schwarzshild sought a stationary, spherically symmetric solution with a metric of the form

$$d\tau^2 = F(r)\,dt^2 - G(r)\,dr^2 - r^2 D\Omega \tag{15}$$

The minus signs are there so that we get back to the right signs for Minkowski space in the limit of weak gravitational fields.

Now, it is possible to insert these forms into the expression for the Christoffel symbols, get them and go on to find the components of the Ricci tensor and the curvature scalar. There are arbitrary constants in the outcome and they are chosen to give back the weak field limit that we found earlier. This is a calculation that we shall not go through since it is not conceptually informative, though it does involve some tricks of the trade that you ought to learn someday if you want be involved in this subject. If the units are selected so that $G = 1$ and $c = 1$, the outcome is the famous Schwarzschild metric

$$d\tau^2 = (1 - \frac{2m}{r})dt^2 - \frac{1}{1 - \frac{2m}{r}}\,dr^2 - r^2\,d\Omega \tag{16}$$

The g_{00} is consistent with what we knew but the singularity in the g_{11} at $r = 2m$ alerts us to the fact that something interesting is going on. In fact, when we use ordinary units, this singularity occurs at

$$r = \frac{2Gm}{c^2}, \tag{17}$$

which is called the Schwarzschild radius. (This was in fact the radius found for the Newtonian black holes.)

3 Precession of the Perihelion of Mercury

3.1 Prerelativity

Kepler found that the orbit of each planet around the sun is an ellipse with the sun at one focus. The point in the ellipse of closest approach to the sun is called the perihelion. Observations revealed that the perihelia of the planets closest to the sun move slowly around the sun as if the ellipses those planets were trying to follow were themselves slowly turning in space. This curious behavior was for a time attributed to perturbations from the gravitational fields of other planets. However, long and extensive calculations failed to produce the right answer for the perihelion of Mercury whose calculated precession rate was off the observed value by some 43"/century. This was not so bad since the full precessional rate was more like 5558"/century. Still, the discrepancy was outside the estimated errors and so it seemed to be pointing to an interesting effect.

For a time, some thought that the solution lay in the influence of an unseen planet within the orbit of Mercury. They were pretty sure of this and even gave the mysterious object a name: Vulcan. Another possibility was that the sun was not perfectly round and had a quadrapole moment. This would produce a bit of inverse cube force and could cause the observed precession. But attempts to detect an oblateness were not successful. This problem was a fly in the Newtonian ointment and was becoming worrisome. Indeed, people were right to worry since the discrepancy pointed to a limitation of Newton's theory that Einstein's theory dealt with.

General Relativity

Installment XV.

October 22, 2003

1 The Exterior Schwarzschild Metric

On second thought, we had better look a little more into the derivation of the Schwarzschild metric, even though there is a good bit of typing involved. Some authors do this by putting the metric in the form

$$d\tau^2 = e^A dt^2 - e^B dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\phi^2 \tag{1}$$

When we go into index notation, we set $x^0 = t, x^1 = r, x^2 = \theta, x^3 = \phi$ so that, for example,

$$d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{2}$$

For a diagonal metric like this, one may denote it as $g_{\mu\nu} = \mathrm{diag}(e^A, -e^B, -r^2, -r^2\sin^2\theta)$. Similarly, we write $g^{\mu\nu} = \mathrm{diag}(e^{-A}, -e^{-B}, -r^{-2}, -r^{-2}\sin^{-2}\theta)$. Recall that the Christoffel symbol is

$$\Gamma^{\ \sigma}_{\mu\nu} = \frac{1}{2} g\sigma\rho \left(g_{\rho\mu,\nu} + g_{\rho\nu,\mu} - g_{\mu\nu,\rho}\right) \tag{3}$$

We also recall the notation

$$[\mu\nu, \lambda] = \frac{1}{2}\left(g_{\rho\mu,\nu} + g_{\rho\nu,\mu} - g_{\mu\nu,\rho}\right) \tag{4}$$

Let us now try a funny notation. The index (σ) is not σ but it has the same value as σ. We do this to make the next step easier to visualize (with any luck). We may then write

$$\Gamma^{\ \sigma}_{\mu\nu} = \frac{1}{2} g^{(\sigma)\sigma}[\mu\nu, (\sigma)] \tag{5}$$

In this expression, (σ) is summed over. The notation says that unless σ has the same value as (σ), there is no contribution since $g^{\mu\nu}$ is diagonal. What this is supposed to help you see is that if μ, ν and σ are all different, the Christoffel symbol is zero.

We then proceed to evaluate the nonzero terms. First we note that, since the solution is supposed to be stationary,

$$g_{\mu\nu,0} = 0 \tag{6}$$

This is just an assumption that would be proven in a more thorough discussion than this one.

Moreover,

$$g_{00,1} = g_{00}A', \qquad g_{11,1} = g_{11}B', \qquad g_{22,1} = -2r, \qquad g_{33,1} = -2r\sin^2\theta \tag{7}$$

and so on, where the prime means differentiation with respect to r. Now we see that

$$[11,1] = (g_{11}B') \tag{8}$$

and so on. We then go to

$$\Gamma = \frac{1}{2}g^{1\mu}[11,\mu] \tag{9}$$

Since the metric does not depend on t, θ or ϕ, we get

$$\Gamma = -\frac{1}{2}e^{-B}\left(-e^{B}B'\right) = \frac{1}{2}B' \tag{10}$$

In this manner, we find

$$\Gamma^{\,1}_{11} = \frac{1}{2}B'; \qquad \Gamma^{\,1}_{22} = -re^{-B}; \qquad \Gamma^{\,1}_{33} = -r\sin^2\theta\, e^{-B}; \qquad \Gamma^{\,1}_{00} = \frac{1}{2}e^{A-B}A' \tag{11}$$

$$\Gamma^{\,0}_{10} = \frac{1}{2}A'; \qquad \Gamma^{\,2}_{12} = \frac{1}{r}; \qquad \Gamma^{\,3}_{13} = \frac{1}{r}; \qquad \Gamma^{\,3}_{23} = \cot\theta; \qquad \Gamma^{\,2}_{33} = -\sin\theta\cos\theta \tag{12}$$

So much for the simple part.

The curvature tensor is

$$R^{\lambda}_{\cdot\mu\nu\sigma} = \Gamma^{\,\alpha}_{\mu\sigma}\,\Gamma^{\,\lambda}_{\alpha\nu} - \Gamma^{\,\alpha}_{\mu\nu}\,\Gamma^{\,\lambda}_{\alpha\sigma} + \Gamma^{\,\lambda}_{\mu\sigma,\nu} - \Gamma^{\,\lambda}_{\mu\nu,\sigma} \tag{13}$$

We contract it to

$$R^{\sigma}_{\cdot\mu\nu\sigma} = \Gamma^{\,\alpha}_{\mu\sigma}\,\Gamma^{\,\sigma}_{\alpha\nu} - \Gamma^{\,\alpha}_{\mu\nu}\,\Gamma^{\,\sigma}_{\alpha\sigma} + \Gamma^{\,\sigma}_{\mu\sigma,\nu} - \Gamma^{\,\sigma}_{\mu\nu,\sigma} = R_{\mu\nu} \tag{14}$$

To evaluate this it is useful to note that

$$\Gamma^{\,\sigma}_{\mu\sigma} = \frac{1}{2}g^{\sigma\rho}\left(g_{\mu\rho,\sigma} + g_{\sigma\rho,\mu} - g_{\mu\sigma,\rho}\right) = \frac{1}{2}g^{\sigma\rho}g_{\sigma\rho.\mu} \tag{15}$$

Then we simply start evaluating the components of $R_{\mu\nu}$. For instance

$$R_{11} = \Gamma^{\,\alpha}_{1\sigma}\,\Gamma^{\,\sigma}_{\alpha 1} - \Gamma^{\,\alpha}_{11}\,\Gamma^{\,\sigma}_{\alpha\sigma} + \Gamma^{\,\sigma}_{1\sigma,1} - \Gamma^{\,\sigma}_{11,\sigma} \tag{16}$$

But

$$\Gamma^{\,\sigma}_{1\sigma} = \frac{1}{2}g^{\sigma\rho}g_{\sigma\rho,1} \tag{17}$$

and

$$\Gamma^{\sigma}_{1\sigma,1} = \frac{1}{2}g^{\sigma\rho}g_{\sigma\rho,1,1} + \frac{1}{2}g^{\sigma\rho}{}_{,1}\, g_{\sigma\rho,1} \tag{18}$$

Plugging everything in, we get

$$R_{00} = -e^{A-B}\left(\frac{1}{2}A'' - \frac{1}{4}A'B' + \frac{1}{4}(A')^2 + \frac{A'}{r}\right) \tag{19}$$

$$R_{11} = \frac{1}{2}A'' - \frac{1}{4}A'B' + \frac{1}{4}(A')^2 + \frac{B'}{r} \tag{20}$$

$$R_{22} = e^{-B}[1 + \frac{1}{2}r(A' - B')] - 1 \tag{21}$$

$$R_{33} = R_{22}\sin^2\theta \tag{22}$$

A nice exercise that is left for you is to show that the curvature (or Ricci) scalar vanishes. Then, since we are in the external case where there is no matter, the Einstein equation (recall Laplace's equation) is $R_{\mu\nu} = 0$. If we combine $R_{00} = 0$ and $R_{11} = 0$, we get

$$e^{A-B}\left(\frac{A'}{r} + \frac{B'}{r}\right) = 0 \tag{23}$$

Then $A' = -B'$ and $A = -B+$constant. In the metric, since A and B are in the exponents, the integration constant merely introduces a change in the scale of time, which is anyway arbitrary so far. We can set the constant equal to zero.

Now, with $R_{22} = 0$, we find

$$e^A(rA' + 1) = 1 \tag{24}$$

or

$$(re^A)' = 1 \tag{25}$$

so that

$$g_{00} = e^A = 1 + \frac{const.}{r} \tag{26}$$

But, we found that $g_{00} = 1+2\phi$, where ϕ is the gravitational potential, so that $const. = 2Gm$ (really $2Gm/c^2$) and this is then Schwarzschild's famous exterior solution. We may also choose units so that $G = 1$ and then $const. = 2m$. In these units, time, distance and mass are all measured in units of length. For example, $m_\odot = 1.5\,km$.

The Schwarzschild exterior metric is then

$$d\tau^2 = (1 - \frac{2m}{r})dt^2 - (1 - \frac{2m}{r})^{-1}\,dr^2 - r^2\,d\Omega \tag{27}$$

with

$$d\Omega = d\theta^2 + \sin^2\theta\,d\phi^2 \tag{28}$$

2 Tests of General Relativity

2.1 Prerelativity

Kepler found that the orbit of each planet around the sun is an ellipse with the sun at one focus. The point in the ellipse of closest approach to the sun is called the perihelion. Observations revealed that the perihelia of the planets closest to the sun move slowly around the sun as if the ellipses those planets were trying to follow were themselves slowly turning in space. This curious behavior is mainly due to perturbations from the gravitational fields of other planets. However, long and extensive calculations failed to produce the right answer for the perihelion of Mercury whose calculated precession rate was off the observed value by some 43”/century. This was not so bad since the full precessional rate was more like 5558”/century. Still, the discrepancy was outside the estimated errors and so it seemed to be pointing to an interesting effect.

For a time, some thought that the solution lay in the influence of an unseen planet within the orbit of Mercury. They were pretty sure of this and even gave the mysterious object a name: Vulcan. Another possibility seriously considered was that the sun is not perfectly round and has a quadrapole moment. This would produce a bit of inverse cube force and could cause the observed precession. But attempts to detect a significant oblateness were not successful. This problem was a fly in the Newtonian ointment and was becoming worrisome. Indeed, people were right to worry since the discrepancy pointed to a limitation of Newton’s theory. Einstein’s theory correctly predicted the discrepancy.

2.2 Orbital Equations

A body moving in the Schwarzschild geometry will naturally follow the geodesics there if there are no (nongravitational) forces acting and if we may neglect the distortions of the geometry produced by the body itself. A body with such a negligible effect is called a test particle. The equation of motion is then

$$\ddot{x}^\alpha + \Gamma^{\,\alpha}_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 0 \tag{29}$$

For $\alpha = 2$ we have

$$\ddot{x}^2 + 2\Gamma^{\,2}_{12}\dot{x}^1\dot{x}^2 + \Gamma^{\,2}_{33}(\dot{x}^3)^2 = 0 \tag{30}$$

or

$$\ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} - \sin\theta\,\cos\theta\,\dot{\phi}^2 = 0 \tag{31}$$

That is,

$$(r^2\dot{\theta}\dot{\,)} - r^2\sin\theta\,\cos\theta\,\dot{\phi}^2 = 0 \tag{32}$$

To make life simple, we let the particle at $t = 0$ be at $\theta = \pi/2$ with $\dot{\theta} = 0$. Then, we see that $\ddot{\theta} = 0$ at $t = 0$, so that $\dot{\theta}$ remains zero for all time and θ stays at the value $\pi/2$. The orbit is in a plane.

Now, for the other components of the geodesic equation, we get

$$\ddot{r} + \frac{1}{2}B'\dot{r}^2 - re^{-B}\dot{\phi}^2 + \frac{1}{2}e^{A-B}A'\dot{t}^2 = 0 \tag{33}$$

$$\ddot{\phi} + \frac{2}{r}\dot{r}\dot{\phi} = 0 \tag{34}$$

$$\ddot{t} + A'\dot{r}\dot{t} = 0 \tag{35}$$

where $A = 1 - 2m/r$. The ϕ equation is just

$$\left(r^2\,\dot{\phi}\right)^{\cdot} = 0 \tag{36}$$

which gives us

$$r^2\dot{\phi} = h \tag{37}$$

where h is a constant of integration (whose meaning you should surmise). The t equation becomes

$$\left(\ln \dot{t}\right)^{\cdot} + \dot{A} = 0 \tag{38}$$

which integrates to

$$\dot{t} = Ce^{-A} \tag{39}$$

where C is another integration constant.

Finally, instead of using the (forbidding) r equation explicitly, we sneak around it by using the metric statement that

$$e^A\dot{t}^2 - e^{-A}\dot{r}^2 - r^2\dot{\phi}^2 = 1 \tag{40}$$

into which we substitute our expression for $\dot{t}$ to get

$$C^2e^{-A} - e^{-A}\dot{r}^2 - \frac{h^2}{r^2} = 1 \tag{41}$$

If we merely want the orbit itself but not (for the moment) the ephemeris, we can look for $r(\phi)$, that is, we examine

$$\dot{r} = \frac{dr}{d\phi}\dot{\phi} \tag{42}$$

Then, with

$$\dot{r}^2 = \left(\frac{dr}{d\phi}\dot{\phi}\right)^2 = \frac{h^2}{r^2}\left(\frac{dr}{d\phi}\right)^2 \tag{43}$$

we get

$$\frac{h^2}{r^4}e^{-A}\left(\frac{dr}{d\phi}\right)^2 + \frac{h^2}{r^2} - C^2e^{-A} = -1 \tag{44}$$

or

$$e^{-A}\left(\frac{h}{r^2}\frac{dr}{d\phi}\right)^2 + \frac{h^2}{r^2} - C^2 e^{-A} = -1 \tag{45}$$

Then

$$\left(\frac{h}{r^2}\frac{dr}{d\phi}\right)^2 + (1-\frac{2m}{r})\frac{h^2}{r^2} = C^2 - (1-\frac{2m}{r}) \tag{46}$$

Now (as is commonly done in the Newtonian case) we set $u = 1/r$ and we find the orbital equation

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{C^2-1}{h^2} + \frac{2m}{h^2}u + 2mu^3 \tag{47}$$

In the Newtonian case, the analogous equation turns out to be

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{C^2-1}{h^2} + \frac{2m}{h^2}u \tag{48}$$

So the extra term of relativity is like that from an inverse cube term in the classical case. You can see why people in the nineteenth century looked for an oblateness of the sun.

General Relativity

Installment XVI

October 27, 2003

1 Orbiting a Schwarzschild Object

1.1 The Newtonian Approach

When I say Newtonian, I really mean classical, though we even see that through Einsteinian lenses. This Newtonian subsection is for those of you who seem shaky on the provenance of Kepler's laws, so some of you can skip this.

Let us accept that the motion of a test body around a spherically symmetric mass is in the Euclidean plane. If there were no forces, the motion would be in straight lines. This is geodesic motion and it is described in Cartesian coordinates, x, y, by

$$\ddot{x} = 0; \qquad \ddot{y} = 0 \tag{1}$$

Since the Newtonian force of gravity is radial, it is advisable to use polar coordinates r, φ where

$$x = r\cos\varphi; \qquad y = r\sin\varphi \tag{2}$$

This decision is made on the basis of past experience in working with this problem. But you should note that it is always wise to select your coordinate system to suit the problem at hand. To accommodate our present choice, we note that

$$\dot{x} = \dot{r}\cos\varphi - r\dot{\varphi}\sin\varphi \tag{3}$$

$$\dot{y} = \dot{r}\sin\varphi + r\dot{\varphi}\cos\varphi \tag{4}$$

and

$$\ddot{x} = \ddot{r}\cos\varphi - 2\dot{r}\dot{\varphi}\sin\varphi - r\dot{\varphi}^2\cos\varphi - r\ddot{\varphi}\sin\varphi \tag{5}$$

$$\ddot{y} = \ddot{r}\sin\varphi + 2\dot{r}\dot{\varphi}\cos\varphi - r\dot{\varphi}^2\sin\varphi + r\ddot{\varphi}\cos\varphi \tag{6}$$

Then $\ddot{x}\cos\varphi + \ddot{y}\sin\varphi = 0$ gives us

$$\ddot{r} - r\dot{\varphi}^2 = 0 \tag{7}$$

and $\ddot{x}\sin\varphi + \ddot{y}\cos\varphi = 0$ is

$$r\ddot{\varphi} + 2\dot{r}\dot{\varphi} = 0 \tag{8}$$

These are the geodesic equations in polar coordinates and, from them, you can read off the Christoffel symbols

$$\Gamma^{\ r}_{rr} = 0; \qquad \Gamma^{\ r}_{r\varphi} = \Gamma^{\ r}_{\varphi r} = 0; \qquad \Gamma^{\ r}_{\varphi\varphi} = -r \tag{9}$$

$$\Gamma^{\ \varphi}_{rr} = 0; \qquad \Gamma^{\ \varphi}_{\varphi r} = \Gamma^{\ \varphi}_{r\varphi} = \frac{2}{r}; \qquad \Gamma^{\ \varphi}_{\varphi\varphi} = 0 \tag{10}$$

If there is a central mass, we need to modify (7) to include the central force

$$\ddot{r} - r\dot{\varphi}^2 = -\frac{Gm}{r^2} \tag{11}$$

while (8) remains the same. As usual, we may integrate (8) to obtain

$$r^2\dot{\varphi} = h \tag{12}$$

where h is the orbital angular momentum of the test particle. Then

$$\ddot{r} = \frac{h^2}{r^3} - \frac{Gm}{r^2} \tag{13}$$

When we multiply this equation by $\dot{r}$ we have something that is easily integrated and we find that

$$\frac{1}{2}\dot{r}^2 = -\frac{h^2}{2r^2} + \frac{Gm}{r} + E \tag{14}$$

where E is a constant of integration whose meaning you will readily perceive.

Our first interest is in finding the orbit, that is in $r(\varphi)$, and so we note that

$$\dot{r} = \frac{dr}{d\varphi}\dot{\varphi} = \frac{h}{r^2}\frac{dr}{d\varphi} \tag{15}$$

We then obtain

$$\frac{h^2}{2r^4}\left(\frac{dr}{d\varphi}\right)^2 = E - \frac{h^2}{2r^2} + \frac{Gm}{r} \tag{16}$$

Experience has shown that the variable $u = 1/r$ is a convenient one in this problem, so we introduce it and obtain (after some minor rearrangements) that

$$\left(\frac{du}{d\varphi}\right)^2 + u^2 = K + \frac{2Gm}{h^2}u \tag{17}$$

where $K = 2E/h^2$. Surprisingly, perhaps, it is easier to solve this equation by first differentiating it. On doing that, we get

$$\frac{d^2u}{d\varphi^2} + u = \frac{Gm}{h^2} \tag{18}$$

This is an inhomogeneous, linear differential equation with constant coefficients. A general solution will be of the form of a particular solution, such as $u = Gm/h^2$, plus the solution of the homogeneous equation, $u = A\cos(\varphi - \varphi_0)$ where A and φ_0 are constants of integration. We make things look a bit nicer by introducing the new constant $e = Ah^2/(Gm)$. Then the orbit is given by

$$r = \frac{h^2/(Gm)}{1 + e\cos(\varphi - \varphi_0)} \tag{19}$$

If you know your conic sections, you will observe that e is the eccentricity and when $e < 1$ the curve is an ellipse. What has the cone got to do with the motion of a planet? One way to think about this is to say that there is no force but that the sun (in the case of the solar system) has deformed the two-dimensional space we are talking about into a conical shape that we can look at from our three-dimensional space. But in the cone, the intrinsic geometry is flat. If you make a cone out of paper and draw a closed orbit around it, you get an ellipse. Unwrap the cone and lie it flat and see what that orbit really is. Of course, in spacetime, the worldline has a helical character as it winds around the time axis

Finally, the point of closest approach to the sun, or perihelion, occurs at $\varphi = \varphi_0$ and of greatest distance, or aphelion, at $\varphi = \varphi_0 + \pi$ so that we have

$$r_{min} = \frac{h^2/(Gm)}{1+e}; \qquad r_{max} = \frac{h^2/(Gm)}{1-e} \tag{20}$$

The major diameter of the ellipse is

$$2a = r_{min} + r_{max} \tag{21}$$

so that the semi-major axis is

$$a = \frac{h^2}{Gm(1-e^2)} \tag{22}$$

The perihelion distance then turns out to be $q = r_{min} = a(1-e)$ which I ask you to work out for yourselves. All these things have been well measured for all the planets.

1.2 Back to Schwarzschild

The conical model space is singular at the apex. A more satisfactory model is a curved two-space, which, for our purposes, is provided by Schwarzschild's exterior metric. As we saw, that is

$$d\tau^2 = (1 - \frac{2m}{r})dt^2 - (1 - \frac{2m}{r})^{-1}\, dr^2 - r^2\, d\Omega \tag{23}$$

with

$$d\Omega = d\theta^2 + \sin^2\theta\, d\phi^2 \tag{24}$$

We may try to draw the same kind of closed orbit on the stretched rubber sheet that this leads to when we limit ourselves to the planar orbit of the theory. Then we again find something very close to an ellipse. However, the ellipse does not quite close on itself. in terms of the ellipse we found in the Newtonian case — equation (19) — it is as if φ_0, is slowly drifting. The rate of drift, known as the precession of the perihelion, may be calculated from the orbital equation we found in the last installment:

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = K + \frac{2m}{h^2}u + 2mu^3 \tag{25}$$

where K is the renamed constant, $(C^2-1)/h^2$. Again, $u = 1/r$ and h is the angular momentum over the mass, but now we have units with $G = 1$ and $c = 1$.

When we differentiate (25), we obtain

$$\frac{d^2u}{d\varphi^2} + u = \frac{m}{h^2} + 3mu^2 \tag{26}$$

We improve the appearance of this equation with the substitution $v = h^2u/m$. Then we have

$$\frac{d^2v}{d\varphi^2} + v = 1 + \epsilon v^2 \tag{27}$$

where

$$\epsilon = \frac{3m^2}{h^2} \tag{28}$$

Roughly speaking, ϵ measures the square of the ratio of the Schwarzschild radius to the radius (or semi-major axis) of the orbit. This equation is a little more difficult to solve than the classical one because of the nonlinear term v^2 but it can be done. As in many such cases, an approximate solution is more directly informative, at least when the approximation is of the right kind. Here we use what is called (singular) pertubation theory.

When $\epsilon = 0$, the equation reduces to the classical one whose solution — call it v_0 — we found readily. In fact, ϵ is very small for the planets, so we look for a solution close to the classical one by assuming an expansion of the form

$$v = v_0 + \epsilon v_1 + \epsilon^2 v_2 + \ldots \tag{29}$$

When we introduce this expression into the equation for v, we get

$$\frac{d^2v_0}{d\varphi^2} + v_0 - 1 + \epsilon[\frac{d^2v_1}{d\varphi^2} + v_1 - v_0^2] + \epsilon\{\ldots\} + \ldots = 0 \tag{30}$$

When we let $\epsilon \to 0$, the only part of this equation that survives must be zero, that is,

$$\frac{d^2v_0}{d\varphi^2} + v_0 - 1 = 0. \tag{31}$$

But this is essentially the equation we solved in the classical problem so we know that the solution is

$$v_0 = 1 + e\cos(\varphi - \varphi_0) \tag{32}$$

Equation (31) does not involve ϵ and it remains true even in studying (30). Hence, when we introduce it into (30), we are left with

$$\epsilon[\frac{d^2 v_1}{d\varphi^2} + v_1 - v_0^2] + \epsilon\{...\} + ... = 0 \tag{33}$$

We can cancel a factor of ϵ from this equation and again let $\epsilon \to 0$. This forces the result

$$\frac{d^2 v_1}{d\varphi^2} + v_1 - v_0^2 = 0 \tag{34}$$

And now the plot thickens.

We find that

$$v_0^2 = 1 + 2e\cos(\varphi - \varphi_0) + e^2\cos^2(\varphi - \varphi_0) \tag{35}$$

which we may write as

$$v_0^2 = 1 + 2e\cos(\varphi - \varphi_0) + \frac{e^2}{2}\left(1 + \cos[2(\varphi - \varphi_0)]\right) \tag{36}$$

We then have the equation

$$\frac{d^2 v_1}{d\varphi^2} + v_1 = 1 + \frac{e^2}{2} + 2e\cos(\varphi - \varphi_0) + \frac{e^2}{2}\cos[2(\varphi - \varphi_0)] \tag{37}$$

This is another inhomogeneous differential equation, but a particular solution is a bit harder to find than for the last inhomogeneous equation we encountered. It is not so hard to see that if we take as our particular solution

$$v_1^{(p)} = 1 + \frac{e^2}{2} + V_1^{(p)} \tag{38}$$

we get

$$\frac{d^2 V_1^{(p)}}{d\varphi^2} + V_1^{(p)} = 2e\cos(\varphi - \varphi_0) + \frac{e^2}{2}\cos[2(\varphi - \varphi_0)] \tag{39}$$

And if we let

$$V_1^{(p)} = -\frac{e^2}{6}\cos[2(\varphi - \varphi_0)] + U_1^{(p)} \tag{40}$$

we obtain

$$\frac{d^2 U_1^{(p)}}{d\varphi^2} + U_1^{(p)} = 2e\cos(\varphi - \varphi_0) \tag{41}$$

The solution of this equation is the nub of the problem: the term on the right of this equation is resonant. What this means is that the solution to this problem grows without bound as a function of φ, thus as a function of time as well. One says that the perturbation theory is singular for this reason. This sort of situation arises in classical celestial mechanics and in a large number of problems of a practical kind. We shall not go into the methods for dealing with these difficulties but shall settle for the growing solution on the understanding that if we do not use it too extensively we can find out what we need to know. Thus the morality of physics (according to G.E. Uhlenbeck): "Do it! If it blows up in your face, then you worry."

In this instance it does no good to try a solution to (41) that is proportional to $2e\cos(\varphi-\varphi_0)$ because, when you insert that on the left of (41), you get zero and you cannot balance the right hand side. In such cases, you are advised to try next something like $\varphi\cos(\varphi-\varphi_0)$. In fact, as you may verify by substitution,

$$U_1^{(p)} = e\varphi\sin(\varphi-\varphi_0) \tag{42}$$

is a solution of (41). A particular solution of the equation for v_1 is then

$$v_1^{(p)} = 1 + \frac{e^2}{2} - \frac{e^2}{6}\cos[2(\varphi-\varphi_0)] + e\varphi\sin(\varphi-\varphi_0) \tag{43}$$

To this, we can (and should) add a solution to the homogeneous equation as before. Such a solution may be useful in fitting to initial conditions but we need not concern ourselves with it here. What want is to see whether the additional terms from Einstein's theory makes for any interesting effects. The one we need to focus on is the resonant term since such terms always tend to change the character of the results and books have been written on this kind of problem. We shall settle here for a down-to-earth simplified version but, if ever you learn the so-called two-time method, go back and look at this problem.

As we are thinking about planetary orbits for now, we have the simplification that, for them, $e^2 << 1$. Then we may write an approximate solution to first order in e as

$$v = 1 + e\cos(\varphi-\varphi_0) + \epsilon[1 + e\varphi\sin(\varphi-\varphi_0)] + ... + \text{homogenous solution} \tag{44}$$

The terms represented by dots do not play much of a role and we shall not worry about them.

We may make the approximations

$$\cos(\epsilon\varphi) = 1 + \mathcal{O}(\epsilon^2); \qquad \sin(\epsilon\varphi) = \epsilon\varphi + \mathcal{O}(\epsilon^3) \tag{45}$$

Then we note that we may write that

$$\begin{aligned}\cos(\varphi-\varphi_0) + \epsilon\varphi\sin(\varphi-\varphi_0) &\approx \cos(\varphi-\varphi_0)\cos(\epsilon\varphi) - \sin(\varphi-\varphi_0)\sin(\epsilon\varphi) \\ &= \cos\left[\varphi - (\varphi_0 + \epsilon\varphi)\right] \end{aligned} \tag{46}$$

Then we write that

$$v = 1 + e\cos(\varphi-\hat{\varphi}_0) + \text{nonsingular terms} \tag{47}$$

This is basically the classical solution except that, instead of φ_0, we have

$$\hat{\varphi}_0 = \varphi_0 + \epsilon\varphi \tag{48}$$

Every time one orbit is completed and φ increases by 2π, $\hat{\varphi}_0$ increases by $4\pi\epsilon/3$. Therefore, for a planet whose orbital period is P, the rate at which the perihelion advances is

$$\frac{6\pi m^2}{h^2 P} \tag{49}$$

If we introduce the semi-major axis, and retreat to conventional units, this is

$$\frac{6\pi G m}{a c^2 P} \tag{50}$$

where $e^2 << 1$ and so is neglected.

The application of this formula is not so simple since there are other factors that may influence the position of the perihelia of planets, most importantly the influence of the other planets. But all those effects had been pretty well worked out before the end of the nineteenth century and the discrepancy for Mercury was known quite accurately. So it was a triumph for relativity when the answer came out right and that had been the major empirical support of general relativity, at least until it became clear that black holes do exist. Still, we may expect that there is more to come in the theory of gravity.