GRAVITY

Chris Vuille, Ph.D.

January 6, 2003

Contents

Chapter 1

Introduction

Gravity is one of the oldest and most mysterious of forces. Ancients must have wondered why everything fell towards the Earth, and probably didn't even suspect that lighter objects attracted each other. According to record, only in the time of Galileo were experiments undertaken to describe the basic properties.

Galileo made a number of important, fundamental contributions. He measured the magnitude of the acceleration of gravity using balls on ramps, and showed that two objects of differing mass and composition fell with the same acceleration. This was a very important experiment, and is just as important today, where it has been repeated to considerably higher accuracy. The gravitational acceleration of a massive object, therefore, doesn't depend on the mass of the object nor on its composition. This is the **principle of equivalence**, which turns out to be one of the primary motivations for the development of general relativity.

1.1 Newtonian Gravity

In seventeenth century England, Isaac Newton, one of the all-time great scientific geniuses, developed calculus and with it a theory of gravitation that today is still the most useful for practical purposes. He posited the existence of a force between any two bodies given by

$$\mathbf{F} = -\frac{mMG}{r^2}\hat{\mathbf{r}} \tag{1.1}$$

This equation is fairly easy to understand. G, of course, is Newton's constant, which binds together the physical measurements into a force. It has a value of $6.672x10^{-11}kg - m^3/s^2$, a value that depends on the units chosen, and which is known with certainty only to about four digits. This makes it the least accurately known of all the fundamental constants of nature, primarily because the force is so weak. Stronger forces are easier to measure with our relatively large, clumsy apparati. The masses of the individual bodies attracting each other, m and M, are multiplied together because, apparently, the gravitational force doesn't saturate. Saturation means that the force law is weakened with the addition of more and more particles. This makes sense from our everyday experience: a bully on a playground may be able to exert a force F on one student, and force F on another student, but add a third student and it becomes difficult for the bully, who has only two arms, to simultaneously exert a force on that student, also. In the case of gravity, it appears that no matter how many separate bodies are involved, a given mass can exert an undiminished force on every single one of the rest of the bodies.

To see that m and M should be multiplied, imagine that there is a fundamentally smallest particle with mass m_0, and that m is composed of n such particles, while M is composed of N particles. Since the gravity field doesn't saturate, each of the n particles will attract every one of the N particles with the same force. Let f be the force between one fundamental particle in the glob m and one fundamental particle in the glob M. Then the total force between m and M would be given by

$$F_{tot} = nNf = nm_0 Nm_0 \frac{f}{m_0^2} = mM \frac{f}{m_0^2} \tag{1.2}$$

Each of the fundamental particles has the same mass, m_o, so it can be immediately seen that the total force between two objects must be proportional to mM.

The dependence on the distance between two particles can be understood in a similar way. Evidently, the force between two bodies is transmitted by some unseen particle, often called a graviton. The graviton must impart momentum to the object it encounters, telling it where and how to move. Imagine a continuous stream of gravitons radiating out from one body and striking another body. The force ought to be proportional to the number of gravitons actually intercepted, which in turn would be proportional to the intensity of gravitons. If, in a given instant, N gravitons are sent out in all directions, then after a suitable period of time N gravitons will cross a sphere at radius r. The graviton intensity at this distance would then be $N/4\pi r^2$. So the dependence on r can be understood as a consequence of elementary geometry.

Finally, the quantity $\hat{r}$, which points from the source towards the body of interest, is a consequence of the vector nature of the force. Whatever mediates this force travels directly towards the affected body, causing a straight-forward transfer of momentum along a geometrically shortest line between the two bodies.

In Newtonian theory, a scalar potential field Φpermeates all space, creating hills and valleys in an additional, fictitious dimension. For every point x^a in space, $\Phi(x^a)$ has a value, and the negative gradient gives a vector perpendicular to the level curves of the potential, and pointing in the direction of greatest decrease of Φ. This vector also gives the acceleration of other massive objects at that point. Here, x^a represents a vector, with $x^1 = x - component$, $x^2 = y - component$, and $x^3 = z - component$. The potential satisfies the Poisson equation,

$$\nabla^2 \Phi = 4\pi G \rho \tag{1.3}$$

For our later work, it's important to briefly review the method of calculating planetary orbits using Newton's theory. And, to make the connection with General Relativity, we use the Lagrangian method. In this method, $L = T - V$, where T is the kinetic en ergy and V is the potential energy. The potential energy V is related to the potential Φ by a factor of m: $V = m\Phi$ Minimizing the difference between kinetic energy and potential energy gives the differential equation of the orbit. The Lagrangian for gravity is therefore

$$L = T - V = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\dot{\theta}^2 + \frac{1}{2}mr^2 \sin^2\theta\dot{\phi}^2 + \frac{mMG}{r} \tag{1.4}$$

where a dot signifies a derivative with respect to time. Lagrange's equation is

$$\frac{d}{dt}\frac{\partial L}{\partial v^a} = \frac{\partial L}{\partial x^a} \tag{1.5}$$

where x^a (a=1,2,3) represents the coordinate variables and v^a represents the velocity associated with x^a. Carrying out this procedure results in the familiar $m\vec{a} = \vec{F}$ formulae, one equation for each coordinate.

Newton's law of gravity was powerful, and enabled the calculation of planetary orbits to incredible accuracy. Still, there were several small discrepancies that astronomers began to notice in the late nineteenth century. Notably, the perihelion shift of Mercury, even after the perturbing influences of Jupiter and other planets were taken into account, was still 43 seconds of arc per century greater than Newton's law predicted.

It was this tiny discrepancy, in part, that motivated Einstein to seek another theory of gravitation. In addition, Newton's law of gravity did not satisfy the laws of special relativity, while Maxwell's theory was in fact invariant in that theory. Finally, Newton's law required a body to react immediately to another body, even when separated by a vast distance. This action-at-a-distance didn't seem to make any sense, since it implied the instant transfer of information.

1.2 Einsteinian Gravity

The story behind general relativity starts with the Theorem of Pythagoras:

$$\Delta s^2 = \Delta x^2 + \Delta y^2 \tag{1.6}$$

This equation provides a means of calculating distances between two points in a plane given a set of cartesian coordinates. It can be easily generalized to a third spatial dimension. Later, Einstein, Minkowski, and others combined time and space in the Minkowski metric,

$$\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 \tag{1.7}$$

The minus signs in the metric are a matter of personal taste. About half the time, this metric appears with the signs -,+,+,+ associated with the four terms, as given above, and the other half of the time with

$$\Delta s^2 = c^2 \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta^2 z$$

Since particle physicists usually prefer +,-,-,-, this second definition will be used in this book. With the introduction of time into the measurement of the interval between points, and especially with sign opposite that of spatial distance, it might be expected that bizarre new predictions would arise—time dilation, length contraction, and so forth. This is indeed the case, and the effects are real and have been verified millions of times in particle accelerators.

Einstein, building on the genius of Gauss and Riemann, went a step further, postulating that spacetime was curved, and that the curvature of spacetime determined the paths of material particles, which travel along geodesics of the spacetime. Geodesics are extremal curves, curves of either minimal or maximal length. In Cartesian space, they are straight lines of minimal length. In spacetime, the geodesics actually have maximal length due to the indefinite metric. 'Indefinite' merely means the metric has both plus and minus signs on the diagonal. In general relativity, matter changes the shape of spacetime, which reflects in the measurement of lengths or intervals. For a general curved spacetime, the Theorem of Pythagoras takes the form

$$ds^2 = f(t, \vec{r})dt^2 - h(t, \vec{r})dx^2 - k(t, \vec{r})dy^2 - l(t, \vec{r})dz^2 \tag{1.8}$$

So in curved spacetime, functions appear in front of the various terms. Cross terms are, of course, also possible, such as $dxdy$. Using this infinitesimal form of the Theorem of Pythagoras, called the **metric**, all the results of general relativity can be calculated.

Curved spaces are in fact very common and are encountered even in elementary calculus. For example, everyone has learned to calculate the length of a curve in a plane. If $f(x, y) = \frac{1}{2}x^2$, the length of the segment from x=0 to x=1 can be calculated with a line integral. First the curve must be parametrized, using

$$\vec{s} = \left(x, \frac{1}{2}x^2\right)$$

The first slot is the x-component, of course, and the second the y-component, with the basis vectors $\hat{x}$ and $\hat{y}$ suppressed. The differential of $\vec{s}$ is

$$d\vec{s} = (1, x)\, dx$$

with a magnitude of

$$ds = \left(1 + x^2\right)^{1/2} dx$$

This is essentially a metric for a one-dimensional curved space. The integral from x=0 to x=1 can now be calculated with

$$Length = \int ds = \int_0^1 (1 + x^2)^{1/2} dx = 1.15 > 1 \tag{1.9}$$

In the last equation, routine integration by trig substitution and by parts leads to the answer. So in curved space, the simple coordinate displacements no longer give the correct distance: the function in front of the corresponding displacement must be figured in, as well. The above metric could be generalized to a metric on a paraboloid given by

$$z = \frac{1}{2}\left(x^2 + y^2\right)$$

The parametrization is

$$\vec{s} = \left(x, y, \frac{1}{2}\left(x^2 + y^2\right)\right)$$

where the unit vectors in the x,y, and z directions are understood. The differential of this parametrization is

$$d\vec{s} = (dx, dy, xdx + ydy)$$

The metric is then given by the dot product of the differential with itself:

$$ds^2 = d\vec{s} \cdot d\vec{s} = dx^2 + dy^2 + (xdx + ydy)^2$$

This can be arranged into a symmetric matrix, while suppressing the dx's and dy's:

$$g_{ab} = \begin{pmatrix} 1 + x^2 & xy \\ xy & 1 + y^2 \end{pmatrix} \tag{1.10}$$

g_{ab} is the matrix of the **components** of the **metric tensor**. It is important to realize that the differential forms, dx and dy, are still there–just suppressed for convenience, because everyone is supposed to understand they're still kicking around. They're important, from a conceptual viewpoint, when it comes to taking derivatives.

It turns out that metrics such as the one given above can always be diagonalized, at least in the neighborhood of a point. The process is to find the eigenvalues and eigenvectors, and use the eigenvectors as a new basis for all vectors and vector operators in the space. The eigenvectors must be chosen carefully, however, since they are only determined up to an overall factor (which in general is not constant). Ideally, the coordinate transformations can be integrated exactly, but again, this is not always possible. In General Relativity, a coordinate transformation is immaterial–the physics, in principle, is always the same regardless of the coordinates, and this is as it should be from an ideal point of view. From a practical standpoint, however, perturbation theory must typically be used to squeeze results out of the equations, and it has not been shown that perturbations in one set of coordinates always give the same answers as perturbations in other coordinates. For the moment and foreseeable future, this has to be taken on faith, and is probably true where spacetime is only slightly curved.

To find the metric, the analog of Poisson's equation is needed. The Einstein equivalent is

$$R_{ab} - \frac{1}{2}Rg_{ab} = \kappa T_{ab} \tag{1.11}$$

This equation won't make complete sense until the end of the second chapter, which covers Tensor Analysis. For now, each term may be thought of as a 4x4 matrix containing 16 functions. R_{ab} may be thought of as the average curvature of spacetime, while R is the average of the average curvature.. It may seem strange that an average involves sixteen different quantities, but this is because there are so many different directions to go in four dimensions. T_{ab} is a set of sixteen functions that represents the energy and matter present in the spacetime. Even if the spacetime is empty of matter, there can still be a curvature, since just the average curvature is zero, not the curvature itself. In words, Einstein's Equation reads

$$AverageCurvature = Stuff$$

The curvatures are all functions of the metric and its derivatives, so with this field equation it's possible to find the metric for a given distribution of matter.

To calculate orbital trajectories in General Relativity, there is no need to invoke potentials or kinetic energies, as these are always built into the metric. Unlike Newton's theory, however, it isn't very straightforward or easy to write down the equations for the gravity field of two or three or more sources. In Newton's theory, because of the linearity of the Poisson equation, fields of different particles can be found individually and then added after the fact. Because Einstein's equations are nonlinear in the metric, it isn't possible to find the metric for each body separately and then add them together. So in Newton's theory, the three-body problem cannot be exactly integrated, but in Einstein's theory, even the two body problem is highly complex. Orbital mechanics in GR involves one relatively large body, like a star, and one negligible body, like a planet. Instead of minimizing a Lagrangian, L, the orbits are found by finding the extremal length curves in the spacetime, i.e. the geodesics. The metric outside a typical spherical body (which ideally would be a black hole) is given by the Schwarzschild solution:

$$ds^2 = \left(1 - \frac{2MG}{r}\right) c^2 dt^2 - \left(1 - \frac{2MG}{r}\right)^{-1} dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right) \quad (1.12)$$

Note that dividing through by $d\lambda^2$, where λ is a parameter with dimensions of time, gives an expression similar to the Lagrangian in Newtonian Gravity.

$$\frac{d}{d\lambda}\frac{\partial ds/d\lambda}{\partial u^a} = \frac{\partial ds/d\lambda}{\partial x^a} \quad (1.13)$$

This is similar to the Lagrangian method, except that a maximal length curve is found instead of a minimum in $L = T - V$. The variable λ is a parameter which, after the variation is taken, may be associated with proper time of the particle, while $u^a = dx^a/d\lambda$, and may be taken as a four-velocity. This expression is unwieldy, and it can be proved that

$$\frac{d}{d\lambda}\frac{\partial (ds/d\lambda)^2}{\partial u^a} = \frac{\partial (ds/d\lambda)^2}{\partial x^a} \quad (1.14)$$

yields the same geodesics, while making the computations easier.

Now, one of the problems that Einstein's theory addressed was that of action-at-a-distance. Unlike Newton's theory, Einstein's general relativity is a local theory. The localcurvature of spacetime causes material particles to change course. There is no longer any question about how a force is transmitted through vast distances instantaneously.

Unfortunately, it is almost universally unrecognized that there is another problem, which is: how does matter over there create curvature over here? In a sense, then, the problem has merely been shifted back one step. And a final question: if gravity is transmitted by gravitons, and in particular an exchange of gravitons, how do the gravitons create curvature?

Despite these and many other unanswered questions, Einstein's Theory of General Relativity is the most beautiful and accurate of all physical theories. Something must be right about it.

1.3 Exercises

1. Find the metric on the upper nappe of a cone given by $z = \sqrt{x^2 + y^2}$.

2. Find the eigenvalues and eigenfunctions of the metric of the paraboloid. Placing the eigenvalues on the diagonal would represent the same metric in new coordinates defined by the eigenfunctions.

3. Set up the geodesic equations for polar coordinates, where the metric is

$$ds^2 = dr^2 + r^2 d\theta^2$$

. Show that rays through the origin are geodesics.

Chapter 2

Tensor Analysis

Einstein's theory of gravitation is based on the fundamentals of tensor analysis and differential geometry. The emphasis on geometry is what sets GR apart from quantum mechanics.

2.1 Scalars, Vectors, and Covectors

A **scalar**, of course, is a number, an element of the real line. A **scalar field**, also known as a **real-valued function**, is a mapping that takes each element of a given domain, for example the coordinates of a point in spacetime, and returns a real number. The potential field Φ of Newtonian gravity is an example of a scalar field. A scalar field is the simplest example of a tensor field, and is said to have rank (0,0). The meaning of rank will be made clear shortly. A **vector**, in elementary physics and math, is usually defined as a directed line segment. It consists of an ordered triple of scalars, called the components of the basis vectors, which in Cartesian space are usually taken to be $\hat{x}, \hat{y}$, and $\hat{z}$. A vector field is a mapping that takes each element of a domain, such as the coordinates of a point in spacetime, and returns a vector.

A vector in another example of a tensor, said to have rank (1,0). Vector fields are characterized by three scalar fields, called the components. The three scalar fields are coefficients of a basis of vectors that span all vectors in a space.

The usual notation is too clumsy for the purposes of general relativity. Instead, a system of indices is used. A typical vector might be represented by

$$\mathbf{V} = V^1\hat{\mathbf{x}} + V^2\hat{\mathbf{y}} + V^3\hat{\mathbf{z}} = \sum_1^3 V^a\mathbf{e_a}$$

Here the V^a are the components of the vector, with a running from 1 to 3, whereas the e_a are the basis vectors. In space-time, a fourth component defined, the time component, and will be assigned an index of 0. In Cartesian space you have $e_1 = \hat{x}$, etc., while V^1= x-component, etc. Notice that the same index–a– appears upstairs and downstairs. It's called a dummy index because it doesn't matter if it's an 'a' or a 'b' or any other letter, it's simply a placeholder for purposes of summation. This kind of summation occurs so often that Einstein invented a convention, now called the Einstein Summation Convention, whereby if a letter repeats itself upstairs and downstairs in a single expression, then summation over that index is understood. For example,

$$\sum_{a=1}^{3} V^a e_a \rightarrow V^a e_a$$

In elementary physics where almost everything is done in a cartesian coordinate system, the e_a don't have a very large role, and often are completely ignored. In a general curvilinear coordinate system, of course, this cannot be the case, because there the basis vectors are functions of the coordinates. This is especially important when taking derivatives, as will be seen. That said, the e_a, while not ignored, are usually dropped, with the vector V being represented by simply the components, V^a, with the e_a understood. Algebraically, there are several useful operations.

1. Addition and subtraction: this is performed component wise, in the obvious fashion. Example:

$$V^a e_a + W^a e_a = (V^0 + W^0)e_0 + (V^1 + W^1)e_1 + (V^2 + W^2)e_2 + (V^3 + W^3)e_3$$

2. Scalar multiplication. To multiply a vector by a real number, simply multiply each of the components by that number:

$$\alpha V^a e_a = \alpha V^0 e_0 + \alpha V^1 e_1 + \alpha V^2 e_2 + \alpha V^3 e_3$$

Obviously, vectors can be multiplied by scalar-valued functions, also. In modern differential geometry, the basis vectors are no longer thought of as passive, unit-sized chunks pointing in a certain direction. Instead, they are considered differential operators– directional derivative operators, to be exact. These derivative operators have special properties under coordinate transformations that come from what is actually just an application of the chain rule. Given coordinates t, x, y, z a vector $\mathbf{V}$ can be written as:

$$\mathbf{V} = V^a \frac{\partial}{\partial x^a} = V^0 \frac{\partial}{\partial t} + V^1 \frac{\partial}{\partial x} + V^2 \frac{\partial}{\partial y} + V^3 \frac{\partial}{\partial z}$$

where the e_a are now given by $\partial/\partial x^a$. To see the connection with the concept of direction derivative operators, recall that the directional derivative of f in the direction $\vec{V}$, $D_V f$, is given by

$$D_V f = \vec{V} \cdot \nabla f = V^x \frac{\partial f}{\partial x} + V^y \frac{\partial f}{\partial y} + V^z \frac{\partial f}{\partial z}$$

Again, for most purposes the basis vectors are suppressed, so the vector can be presented simply by V^a, the matrix of components.

Now consider a vector V in the coordinates $x^a = x^0, x^1, x^2, x^3$. How do we obtain the corresponding vector in the coordinates u^a? In this section, u^a represent coordinates, not a four-velocity. The answer is given by the basis theorem.

Basis Theorem. Let **V** be a vector in x_a coordinates. Then in u^a coordinates, V will be given by $\mathbf{V} = (\mathbf{V}u^a)\,\frac{\partial}{\partial u^a}$

Example. (A)Transform the partial derivative operators $\partial/\partial x$ and $\partial/\partial y$ from Cartesian coordinates to polar coordinates using the basis theorem. (B) Find the Laplacian in polar coordinates using the results of part (A).
Solution: In this case our u^a coordinates are given by $u^1 = r$ and $u^2 = \phi$, while $x^1 = x$ and $x^2 = y$. We apply the basis theorem for $\partial/\partial x$ in the x^a coordinates to the two coordinate transformation equations:

$$r = \sqrt{x^2 + y^2}$$

$$\phi = \arctan(y/x)$$

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \phi}{\partial x}\frac{\partial}{\partial \phi} = \frac{x}{r}\frac{\partial}{\partial r} + \frac{-y}{r^2}\frac{\partial}{\partial \phi} = \cos\phi\frac{\partial}{\partial r} - \frac{sin\phi}{r}\frac{\partial}{\partial \phi}$$

Similarly,

$$\frac{\partial}{\partial y} = \sin\phi\frac{\partial}{\partial r} + \frac{\cos\phi}{r}\frac{\partial}{\partial \phi}$$

With these, the 2-dimensional Laplacian in polar coordinates can be found.

$$\begin{gathered}\tfrac{\partial^2}{\partial x^2} + \tfrac{\partial^2}{\partial y^2} = \left(\cos\phi\tfrac{\partial}{\partial r} - \tfrac{sin\phi}{r}\tfrac{\partial}{\partial \phi}\right)\left(\cos\phi\tfrac{\partial}{\partial r} - \tfrac{sin\phi}{r}\tfrac{\partial}{\partial \phi}\right) + \\ + \left(\sin\phi\tfrac{\partial}{\partial r} + \tfrac{\cos\phi}{r}\tfrac{\partial}{\partial \phi}\right)\left(\sin\phi\tfrac{\partial}{\partial r} + \tfrac{\cos\phi}{r}\tfrac{\partial}{\partial \phi}\right) = \tfrac{\partial^2}{\partial r^2} + \tfrac{1}{r}\tfrac{\partial}{\partial r} + \tfrac{1}{r^2}\tfrac{\partial^2}{\partial \phi^2}\end{gathered}$$

Let's write down the basis theorem result in terms of vector components only, suppressing the differential basis operators. Let $\bar{V}^a$ represent the components expressed in the coordinate system $\{u^a\}$

$$\bar{V}^a = \frac{\partial u^a}{\partial x^b} V^b \tag{2.1}$$

All vectors, called contravariant vectors, transform in this way.

Example. An observer measures the energy-momentum four-vector of a particle to have Cartesian coordinates $V^a = (E/c, pc, 0, 0)$. What components are measured by an observer traveling at v in the positive x-direction with respect to the first observer?

Solution: This is simply a matter of applying the Lorentz transformations:

$$\bar{x} = \gamma\,(x - vt)$$

$$\bar{x^o} = \bar{t} = \gamma\left(t - vx/c^2\right)$$
$$\bar{y} = y$$
$$\bar{z} = z$$

Plugging into the formula, get

$$\bar{V}^0 = \bar{E} = \frac{\partial \bar{x}^0}{\partial t} V^t + \frac{\partial \bar{x}^0}{\partial x} V^x = \gamma \frac{E}{c} - \frac{v}{c}\gamma p$$

$$\bar{V}^1 = \bar{p}c = \frac{\partial \bar{x}^1}{\partial t} V^0 + \frac{\partial \bar{x}^1}{\partial x} V^x = -\gamma E \frac{v}{c} - \gamma pc$$

It might be thought that vectors will be sufficient for any physics we might have to do, but in fact there is a very similar object called a **covector**, also called a **covariant vector**. There are a variety of definitions, depending on taste, and all of them are equivalent. To motivate the existence of these objects, consider the differential of a real-valued function **f**. In a given coordinate basis, this is given by

$$\mathbf{df} = \frac{\partial f}{\partial x^a} dx^a$$

Now suppose we want to change to the coordinates u^a. We can do so by using the chain rule:

$$df = \frac{\partial f}{\partial x^a} dx^a = \frac{\partial f}{\partial x^a} \frac{\partial x^a}{\partial u^b} du^b$$

Evidently, we also have the direct calculation

$$df = \frac{\partial f}{\partial u^b} du^b$$

These two expansions must be equal. Comparing each side, we see that

$$\frac{\partial f}{\partial u^b} = \frac{\partial f}{\partial x^a} \frac{\partial x^a}{\partial u^b}$$

It's easy to see that the tranformation rule is similar but distinct from the transformation rule for vectors. Notice that the transformation factor appears to be upside down compared to the factor in equation for transformation of vectors. This is the hallmark of a covector. Furthermore, the index representing the components of these objects is downstairs, rather than upstairs. The components of a covector, $\bar{W}_a$, in the basis $\bar{x}^a$ are related to the components W_a in the basis x^a via

$$\bar{W}_a = W_b \frac{\partial x^b}{\partial \bar{x}^a} \qquad (2.2)$$

All **covectors**, also called **covariant vectors**, satisfy a relationship of this kind under coordinate transformations. So while vectors are associated with derivatives, covectors, also called 1-forms, are associated with differentials. A given covector W can be written

$$\mathbf{W} = W_a e^a = W_0 e^0 + W_1 e^1 + W_2 e^2 + W_3 e^3 = W_a dx^a$$

As will be seen in the next section, covariant vectors can be thought of as linear operators that map vectors to real numbers. Vectors are said to be tensors of rank (1,0), while covectors are tensors having rank (0,1). It turns out that higher rank tensors can be defined, as discussed in the next section.

2.2 Tensors

By taking what are called exterior products of vectors and covectors, it is possible to build up a whole collection of more complicated objects. These objects, together with the scalars, vectors, and covectors, are called Tensors.

2.2.1 Exterior Products

Definition. Let **A** and **B** be either covectors or vectors, each having n components. The exterior product of **A** and **B**, denoted $\mathbf{A} \bigotimes \mathbf{B}$consists of multiplying all elements of **A**with elements of **B** in the usual polynomial fashion, maintaining order of the basis vectors. In terms of components, we have

$$\mathbf{A} \bigotimes \mathbf{B} = A^a B^b e_a e_b \tag{2.3}$$

Here a double sum is, of course, implied by the Einstein summation convention.

Example. A classic example is the formation of **dyadics** by multiplying two Cartesian vectors. With $\vec{A} = \hat{\imath} + 2\hat{\jmath} + 3\hat{k}$ and $\vec{B} = 2\hat{\imath} - \hat{\jmath} + \hat{k}$ we have

$$\mathbf{C} = \vec{A} \bigotimes \vec{B} = (\hat{\imath} + 2\hat{\jmath} + 3\hat{k})(2\hat{\imath} - \hat{\jmath} + \hat{k}) =$$

$$= 2\hat{\imath}\hat{\imath} - \hat{\imath}\hat{\jmath} + \hat{\imath}\hat{k} + 4\hat{\jmath}\hat{\imath} - 2\hat{\jmath}\hat{\jmath} + 2\hat{\jmath}\hat{k} + 6\hat{k}\hat{\imath} - 3\hat{k}\hat{\jmath} + 3\hat{k}\hat{k}$$

The first number in the pair indicates the row number, while the second denotes the column number. Notice that C can be represented in terms of its components as a matrix:

$$C^{ab} = \begin{pmatrix} 2 & 1 & 1 \\ 4 & -2 & 2 \\ 6 & -3 & 3 \end{pmatrix} \tag{2.4}$$

It should be clear, however, that C^{ab} is not a matrix: the presence of the basis vectors, ii, ij, ik, etc. introduces geometric content that is absent from a simple matrix, which is purely a collection of scalars.

In modern differential geometry we dispense with the $\hat{\imath}, \hat{\jmath}, \hat{k}$ in favor of the partial derivative and differential operators. Nonetheless, most of the time in the course of calculations we deal with the components only, almost as if the tensors were matrices. Special rules apply in certain situations, for example when taking derivatives.

2.2.2 Tensorial Rank

The object **C** in the above example is called a tensor of rank (2,0). Scalar fields are tensors of rank (0,0), vector fields have rank (1,0), while covector fields have rank (0,1). In general, the rank can be found by simply counting upstairs and downstairs indices in the components. Indices that are summed over in the components are not counted. A geometric object with n upstairs, or contravariant, indices and m downstairs, or covariant, indices in the matrix of its components is said to have rank (n,m)—n times contravariant, and m times covariant.
Example. Find the rank of each of the following tensors: (A) ${F^a}_b$ (B) ${F^{ab}}_b$ (C)${Q^{abc}}_{de}$
Solution: (A) one upstairs and one downstairs index, so the rank is (1,1) (B) One upstairs index, and then an up and down which are summed over each other, which removes their tensorial character. Hence the rank is (1,0) (C) Three upstairs and two downstairs indices is a tensor of rank(3,2).

2.2.3 Contraction

Once we have formed a number of basic tensors, we can multiply them together with the exterior (tensor) product and create numerous other tensors of higher rank. There is another process, called a **contraction** or **inner product**, that produces tensors of lower rank. It consists of summing over a given pair of indices, one of them a covariant index, the other a contravariant index. The operation is similar to that of the trace of a matrix.

Example. Find the inner product of V^a and W_b, given that $V^a = (1, 2, 3, 4)$ and $W_b = (-1, 3, 0, -5)$.
Solution: $V^aW_a = V^0W_0+V^1W_1+V^2W_2+V^3W_3 = 1\cdot-1+2\cdot3+3\cdot0+4\cdot-5 = -15$

The **metric**, discussed previously and in more detail below, also plays a role in finding the inner product of two vectors. It can't be interpreted in the same way as in Cartesian coordinates, because the metric is indefinite–the time-time component is negative, while the others are positive, as in a normal Cartesian space.
Example. Find the inner product of the following two vectors in Minkowski space (let c=1). $V^a = (1, 2, 3, 4)$ and $S^b = (4, 3, 2, 1)$
Solution: $g_{ab}V^aS^b = -1 \cdot 1 \cdot 4 + 1 \cdot 2 \cdot 3 + 1 \cdot 3 \cdot 2 + 1 \cdot 4 \cdot 1 = 12$ Notice that in the double sum, since all the off-diagonal terms in the metric are zero, only the terms with $a = b$ are picked up. It's important, naturally, to specify the indices contracted over.

2.2.4 Tensor Transformation Laws

Tensors all transform according to a rule similar to that for vectors. Essentially, there must be one covariant transformation factor for each covariant index, and one contravariant transformation factor for each contravariant index. For example, a tensor of rank (1,2), once contravariant, twice covariant, would transform from the x^a coordinates to the y^a coordinates according to the following rule:

$$\bar{B}^a_{bc} = \frac{\partial y^a}{\partial x^d}\frac{\partial x^e}{\partial y^b}\frac{\partial x^f}{\partial y^c}B^d{}_{ef}$$

The above equation involves only the components, however. As will be customary, the basis vectors e_a and basis covectors e^a are suppressed.

2.2.5 Symmetric and Antisymmetric Tensors

A second-rank tensor Q_{ab} is called **symmetric** if

$$Q_{ab} = Q_{ba} \tag{2.5}$$

This is a slight abuse of notation, since a and b are simply place-keepers, however the idea is that the off-diagonal elements are swapped with their opposite numbers, like the transpose of a matrix. A given tensor can be symmetric in any pair of indices, regardless of the order. In addition, it's always possible to create symmetric tensors from an arbitrary tensor:

$$S_{ab} = \frac{1}{2}\left(C_{ab} + C_{ba}\right) \tag{2.6}$$

It can be easily verified that S_{ab} is symmetric. A higher-rank symmetric tensors can also be constructed:

$$U_{abc} = \frac{1}{3!}\left(P_{abc} + P_{abc} + P_{bac} + P_{bca} + P_{cab} + P_{cba}\right) \tag{2.7}$$

Again, it can be verified that U_{abc} is symmetric under interchange of any two of its indices. Further generalizations are obvious.

Another important type of symmetry is **antisymmetry**. An antisymmetric tensor of second rank has the property that

$$F_{ab} = -F_{ba} \tag{2.8}$$

It's evident that the diagonal elements are all zero. Given any tensor C_{ab}, it is possible to construct and antisymmetric tensor K_{ab}:

$$K_{ab} = \frac{1}{2}\left(K_{ab} - K_{ba}\right) \tag{2.9}$$

Just as in the case of symmetric tensors, it is always possible to create a completely antisymmetric tensor from a given arbitary tensor. Constructing such a tensor for a rank (0,3) tensor is left as an exercise.

2.3 The Metric

A metric is a rule that aids in the measurement of the distance between one point and another. It also allows the measurement of the lengths of vectors, which as we've seen are related to infinitesimal displacements and derivative operators. In Minkowski space, the metric is naturally given by

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 \tag{2.10}$$

The Minkowski metric is often written in its matrix form as

$$\eta_{ab} = \begin{pmatrix} c^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{2.11}$$

The metric provides an inner product, and divides vectors into three classes. If $g_{ab}V^aV^b > 0$, the vector V^a is called timelike. If $g_{ab}V^aV^b < 0$, the vector V^a is called spacelike. If $g_{ab}V^aV^b = 0$, the vector V^a is called null. In a timelike vector the time component is predominate, hence the norm is positive. Material particles, in general, are all timelike. A space-like vector has larger space components, hence the norm is negative. Tachyons are an example of particles that would travel on curves with spacelike tangent vectors. Finally, the inner product of zero corresponds to light, gravitons, and other massless particles, possibly neutrinos.

Example. Find the components of the Minkowski metric in spherical coordinates.

Solution: The easiest but somewhat lengthy way of doing this is to write down the coordinate transformations and then take differentials, substituting into the usual Cartesian metric. This is left as an exercise.

The speed of light, c, is often taken to be unity. This corresponds to a coordinate transformation whereby time is measured in meters, with c as the conversion factor. The metric g_{ab} has an inverse, which can be found by inverting the matrix of components. This inverse is represented by g^{ab}.

In curved spacetime, the metric takes a more general form, where the components may be functions of space and time. Given a metric, there is a natural way to turn vectors into covectors and vice-versa. This is called raising and lower indices. For example,

$$V^a = g^{ab}V_b$$

This gives the so-called the canonical correspondence between vector and covector components. Evidently, it works the other direction, also:

$$W_a = g_{ab}W^b$$

Note that any second rank tensor could be used for this definition, however using the metric tensor is standard and useful.

2.4 The Covariant Derivative

In order to compare vectors at nearby points, it's necessary to define a derivative operator, which is often refered to as a **connection**. A connection is something that tells you how a vector changes when you slide it to a nearby point. Once a connection is defined, then nearby vectors can be compared, and limits and derivatives can be taken. Conversely, we can define a derivative operator, and from there come up with a connection.

Fundamental to defining a derivative is the concept of parallel transport of a vector. If a vector at one point is carried to a nearby point, being always careful to keep it pointing–as much as possible– parallel to the direction it was pointing an infinitesimal distance before, how is it changed? In flat space, nothing happens, but in curved space, the curvature can cause the vector to point in a different direction compared to the original vector.

Let ∇_a represent a derivative operator. There are several desirable properties of this operator:

(1) Linearity: If A and B are tensors, and α and β are numbers (elements of the real line), then $\nabla_a(\alpha A + \beta B) = \alpha\nabla_a A + \beta\nabla_B$
(2) Satisfies the Leibnitz rule: $\nabla_a AB = (\nabla_a A)B + A(\nabla_a B)$
(3) Acts like a normal gradient operator on scalar fields
(4) Commutes with contraction operations
(5) Torsion free: $\nabla_a\nabla_b f = \nabla_a\nabla_b f$, f a scalar field.

Torsion refers to a symmetry in the connection, as will be seen. These properties are natural properties, though other conditions could conceivably be chosen. In principle, there are an infinite number of different derivative operators, but we'll pick out one that seems to be special, or right. One possible operator could be defined by

$$\nabla_a T^{b....h}_{ij...p} = \partial_a T^{b....h}_{ij...p} \tag{2.12}$$

This definition, whereby ordinary partial derivatives of the components are taken, might do in flat space, but is definitely inadequate in curved spacetime. To see this, consider what happens, via the Leibnitz rule, when this operator is applied to a vector in curved spacetime:

$$\nabla_b\left(V^a\mathbf{e_a}\right) = \left(\nabla_b V^a\right)\mathbf{e_a} + V^a\left(\nabla_b\mathbf{e_a}\right)$$

Since in general $\nabla_b e_a$ is not zero, it can be immediately seen that by the Leibnitz rule there will always be an additional term. This term involves what is called the **connection**, and connection coefficients, which technically, in an affine space, could be chosen to be anything whatsoever. The connection coefficients, Γ^c_{ab}, are defined by

$$\nabla_a e_b = \Gamma^c_{ab} e_c \tag{2.13}$$

so the derivative of the basis vectors returns a linear combination of the basis vectors. Note that the rank has also changed.The covariant derivative of a vector V^a is given by

$$\nabla_b V^a = \frac{\partial V^a}{\partial x^b} + \Gamma^a_{bc} V^c \tag{2.14}$$

The covariant derivative of a covector W_a is given by

$$\nabla_b W_a = \frac{\partial V_a}{\partial x^b} - \Gamma^c_{ab} W_c \tag{2.15}$$

This can be derived by using the idea that the covariant derivative of a scalar field is just the usual ordinary partial derivative, and then looking at a covector contracted with a vector (which is a tensor of rank zero—a scalar field).

Covariant derivative of a tensor of arbitrary rank is similar to the above, with a Γ term for each index, for example:

$$\nabla_a Q^{bc}_d = \frac{\partial Q^{bc}_d}{\partial x^a} + \Gamma^b_{ae} Q^{ec}_d + \Gamma^c_{ae} Q^{be}_d - \Gamma^e_{ad} Q^{bc}_e \tag{2.16}$$

Often, in the literature, partial derivatives of tensors are represented by commas followed by an index, for example $W_{b,a}$, while covariant derivatives are written with a semi-colon as $W_{b;a}$.

While the connection coefficients, Γ^c_{ab} can be chosen to be anything in principle, arbitrary choices lead to strange and arbitrary answers. It is more natural, then, to arrive at a definition that uses the metric somehow, since it is already known that curved spacetimes will affect vectors through the metric tensor.

To this end, we need a principle, and the principle is the following: if two vectors are parallel transported along a curve, then their inner product (under the metric) is unchanged. Briefly, parallel transport of a vector V^a along a curve C with tangent vector t^b is said to satisfy the equation

$$t^b \nabla_b V^a = 0 \tag{2.17}$$

In other words, the directional derivative along the curve is zero. If the inner product of two vectors, V^a and W^b, is to be preserved during parallel transport, then the following condition must hold:

$$t^c \nabla_c \left(g_{ab} V^a W^b\right) = t^c V^a W^b \nabla_c g_{ab} = 0 \tag{2.18}$$

where the Leibnitz rule has been applied. This must hold for any curve whatsoever, which implies that

$$\nabla_c g_{ab} = 0 \tag{2.19}$$

So the preservation of inner products under parallel transport means the covariant derivative of the metric tensor equals zero. This is taken to be a desirable property. In terms of the connection coefficients, this condition is written

$$\nabla_a g_{bc} = \frac{\partial g_{bc}}{\partial x^a} - \Gamma^d_{ab} g_{dc} - \Gamma^d_{ac} g_{bd} = 0$$

Rewriting this expression twice again while commuting the indices a, b, c results in:

$$\nabla_c g_{ab} = \frac{\partial g_{ab}}{\partial x^c} - \Gamma^d_{ca} g_{db} - \Gamma^d_{cb} g_{ad} = 0$$

$$\nabla_b g_{ca} = \frac{\partial g_{ca}}{\partial x^b} - \Gamma^d_{bc} g_{da} - \Gamma^d_{ba} g_{cd} = 0$$

Note that we have written the same equation down three times, merely rotating the indices. Now adding the first two and subtracting the last, and doing a little algebra, results in:

$$\Gamma^a_{bc} = \frac{1}{2} g^{ad} \left(g_{dc,b} + g_{bd,c} - g_{bc,d} \right) \tag{2.20}$$

These are called the **Christoffel symbols of the second kind**, and is one example of a **metric connection**. Christoffel symbols of the first kind can be obtained by simply lowering the upper index with the metric. Other connections exist, such as gauge connections, which are derived from the theory of fibre bundles and are extensively used in unified field theories.

Example. (A) Compute the Christoffel symbols for a 2-sphere. (B)Find the covariant derivative of the gradient of the function $f = \sin\theta$.

Solution: (A)The two variables are $x^2 = \theta$ and $x^3 = \phi$, and the metric is given by

$$ds^2 = R^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right)$$

so that $g_{22} = R^2$, $g_{23} = g_{32} = 0$, and $g_{33} = R^2 \sin^2\theta$. (The numbering of the coordinates,x^2 and x^3 is to be consistent with the four-dimensional case, where $x^0 = t$ and $x^1 = r$). The components of the inverse metric are given by

$$g^{ab} = \begin{pmatrix} R^{-2} & 0 \\ 0 & R^{-2}\sin^{-2}\theta \end{pmatrix}$$

There are only three non-zero Christoffel symbols. The first is calculated as follows:

$$\Gamma^2_{33} = \frac{1}{2} g^{2c} \left(g_{3c,3} + g_{c3,3} - g_{33,c} \right)$$

Remembering that upstairs and downstairs repeated indices are summed over, the index c can only take the value $c = 2$, since off-diagonal elements of

g^{ab} are identically zero. This is the case for most simple metrics of physical interest. Resuming:

$$\Gamma^2_{33} = -\frac{1}{2} g^{22} \frac{\partial g_{33}}{\partial \theta} = \frac{1}{2} \cdot \frac{1}{R^2} \cdot 2R^2 \sin\theta \cos\theta = -\sin\theta\cos\theta$$

The last non-zero Christoffel symbol can be calculated similarly:

$$\Gamma^3_{23} = \Gamma^3_{32} = \frac{1}{2} g^{33} \frac{\partial g_{33}}{\partial \theta} = \cot\theta$$

(B) Essentially, we're computing $\nabla_a \nabla_b f$. The first derivative gives the simple answer $\partial f / \partial x^b$, similar to the gradient in ordinary space. The second derivative gives

$$\nabla_a \nabla_b f = \frac{\partial^2 f}{\partial x^a \partial x^b} - \Gamma^c_{ab} \frac{\partial f}{\partial x^c}$$

The matrix of components is given by

$$\begin{pmatrix} -\sin\theta & 0 \\ 0 & \sin^2\theta\cos\theta \end{pmatrix}$$

2.4.1 Geodesics

A geodesic is a generalization of a straight line. Spacetime is curved, however, so there will be no absolutely straight lines. One property of lines in flat Cartesian space is the fact that the vector tangent to the line never changes. On a parabola, of course, it changes constantly, but on a straight line it never does. Hence, the derivative of a straight line's tangent vector must be zero. If we take this over to curved spacetime, we have that the covariant derivative of the tangent vector field along the curve will be zero. If that tangent vector field is called V^a, then the condition is:

$$V^b \nabla_b V^a = 0 \tag{2.21}$$

Using the fact that the tangent vector field is given by

$$V^a \nabla_a = \frac{dx^a}{d\tau} \nabla_a = \frac{d}{d\tau} \tag{2.22}$$

where τ is the proper time–that is, time as experienced by an observer traveling on the curve, the geodesic equation can be written

$$\frac{\partial^2 x^a}{d\tau^2} + \Gamma^a_{bc} \frac{dx^b}{d\tau} \frac{dx^c}{d\tau} = 0 \tag{2.23}$$

Note the use of the basis theorem in equation 2.22. Geodesic equations need not be written in terms of propertime, τ. The sole requirement is that it be written in terms of a linear function of the distance along the line, s. The parameter τ corresponds to a normalization, where s must be such that

$$\frac{dx^a}{ds} \frac{dx^b}{ds} = 1$$

Example. Show that the equator and great circles through the poles are geodesics on the 2-sphere of radius R.

Solution: This is a straightforward calculation using the Christoffel symbols calculated in the previous example. The geodesic equations are:

$$\frac{d^2\theta}{ds^2} - \sin\theta\cos\theta\left(\frac{d\phi}{ds}\right)^2 = 0$$

and

$$\frac{d^2\phi}{ds^2} + 2\cot\theta\frac{d\theta}{ds}\frac{d\phi}{ds} = 0$$

These can be easily first integrated, but that won't be necessary. It's easy to verify that $\phi = \mathit{constant}$,$\theta = s$ and $\phi = s$, $\theta = \pi/2$ satisfy both geodesic equations. All other geodesics, which might be found by full integration of the equations, correspond to great circles which do not fall on the ϕ and θ coordinate lines.

2.5 Curvature

Curved spacetime will cause a parallel-transported vector to change direction when it goes around a closed loop. This doesn't happen when the spacetime is flat. In curved spacetime, it always happens. This is most easily seen when parallel transporting a vector on a sphere.

Example. Take a vector on a sphere at the equator pointing north. Move it along a longitudinal line in such a way that it remains tangent to the curve and pointing north continuously. Upon reaching the North pole, again without changing the orientation of the vector, move it along a longitudinal line 90 degrees away from the original, back to the equator. Then, again parallel transporting the vector, move it back to the origin. You will find that the vector has been rotated through ninety degrees, even though at each step you were careful not to disturb the basic direction of the vector compared to infinitesimally nearby positions.

2.5.1 The Curvature Tensor, R_{abcd}

An expression measuring directly the change a vector goes through while being parallel transported around a curve can be derived. For convenience, define coordinates in a rectangle, with the points A, B, C, D at the corners. Under parallel transport of a vector V^a, we have the condition

$$t^b\nabla_b V^a = 0 \rightarrow t^b\frac{\partial V^a}{\partial x^b} = -t^b\Gamma^a_{bc}V^c$$

We will assume that the vector V^a is to be parallel-transported around a small rectangle, where the first side is parallel to e_j and the second side is parallel to e_k. The vector will then return, parallel to the e_j coordinate and then to the e_k coordinate. Since the directional derivative is in the direction of the coordinate basis vectors, the components t^b will all be zero except for the coordinate traveled on, which shall have value one. We will assume the x^j coordinate goes from a to $a + \delta a$, while the x^k coordinate runs from b to $b + \delta b$. In the following, summation over j and k is suspended: these are fixed quantities, taking on one value only. Using the above equation and computing from A to B, write

$$V^a(B) = V^a(A) + \int_A^B \frac{\partial V^a}{\partial x^j} dx^j = V^a(A) + \int_a^{a+\delta a} -\Gamma^a_{bj} V^b dx^j \tag{2.24}$$

Similarly, from B to C gives,

$$V^a(C) = V^a(B) + \int_B^C \frac{\partial V^a}{\partial x^k} dx^k = V^a(B) + \int_b^{b+\delta b} -\Gamma^a_{bk} V^b dx^k \tag{2.25}$$

From C to D:

$$V^a(D) = V^a(C) + \int_C^D \frac{\partial V^a}{\partial x^b} dx^b = V^a(C + \int_{a+\delta a}^a -\Gamma^a_{bj} V^b dx^j \tag{2.26}$$

And finally, from D back to A:

$$V^a(A) = V^a(D) + \int_D^A \frac{\partial V^a}{\partial x^k} dx^k = V^a(D) + \int_{b+\delta b}^b -\Gamma^a_{bk} V^b dx^k \tag{2.27}$$

The net change is given by the difference between the transported vector and the original vector. In flat space, this will be zero, but in curved spacetime the Γ^a_{bc} and V^b are functions of the coordinates, and so will have different values along different parts of the loop. To clarify this point, we define the following quantity, for example along the curve AB:

$$\Gamma^a_{bj} V^b = P^a_j(AB) \tag{2.28}$$

Thus the path along which the quantity takes its values is explicitly written. Using this notation, adding all four equations yields

$$\delta V^a = -\int_a^{a+\delta a} P^a_j(AB) dx^j - \int_b^{b+\delta b} P^a_k(BC) dx^k - \int_{a+\delta a}^a P^a_j(CD) dx^j - \int_{b+\delta b}^b P^a_k(DA) dx^k \tag{2.29}$$

A little rearrangement results in:

$$\delta V^a = \int_a^{a+\delta a} \left[P_j^a(CD)dx^j - P_j^a(AB)\right] dx^j - \int_b^{b+\delta b} \left[P_k^a(BC) - P_k^a(DA)\right] dx^k \tag{2.30}$$

Now, on CD, $x^k = b + \delta b$ while on AB, $x^k = b$. Thus the first integral is a difference in P_j^a at two neighboring points on the x^k coordinate line. Dividing and multiplying by δb gives approximately the derivative in the direction of x^k times δb. The other integral term is similar, except it's approximately a derivative in the direction of x^j times δa. Hence

$$\delta V^a = \int_a^{a+\delta a} \frac{\partial P_j^a}{\partial x^k} \delta b dx^j - \int_b^{b+\delta b} \frac{\partial P_k^a}{\partial x^j} \delta a dx^k \tag{2.31}$$

Integrating these expressions, considering them to be approximately constant over the small displacements involved, gives

$$\delta V^a = \delta a \delta b \left(\frac{\partial P_j^a}{\partial x^k} - \frac{\partial P_k^a}{\partial x^j} \right) \tag{2.32}$$

Reinserting the expression for the P's and using the product rule, followed by a substitution involving the definition of parallel transport, results in:

$$\delta V^a = \delta a \delta b V^b \left(\frac{\partial \Gamma^a_{bj}}{\partial x^k} - \frac{\partial \Gamma^a_{bk}}{\partial x^j} + \Gamma^a_{ck}\Gamma^c_{jb} - \Gamma^a_{cj}\Gamma^c_{kb} \right) \tag{2.33}$$

The expression in parentheses in the above equation defines the **Riemann curvature tensor**:

$$R^a{}_{bkj} = \frac{\partial \Gamma^a_{bj}}{\partial x^k} - \frac{\partial \Gamma^a_{bk}}{\partial x^j} + \Gamma^a_{ck}\Gamma^c_{jb} - \Gamma^a_{cj}\Gamma^c_{kb} \tag{2.34}$$

The Riemann tensor informs us how spacetime is curved throughout a manifold. While in principle it has 256 components, the symmetries of the Christoffel symbols eliminate all but twenty-one of these. Finally, the Bianchi identity, given by eliminates one more, leaving a maximum total of 20.

The expression obtained for the Riemann tensor can also be obtained by computing the commutator of two derivatives:

$$\left(\nabla_a \nabla_b - \nabla_b \nabla_a\right) V^c = -R^c{}_{abd} V^d \tag{2.35}$$

This is a special case of the Ricci identity, and can be generalized–with numerous additional terms– for more general tensors.

The Riemann tensor has several important symmetries, such as antisymmetry in the first pair and last pair of indices:

$$R_{abcd} = -R_{bacd} = -R_{abdc} \tag{2.36}$$

In addition, Riemann is symmetric when switching the first pair with the second pair of indices:

$$R_{abcd} = R_{cdab} \tag{2.37}$$

By permutation of any three indices, the following identity holds, for example:

$$R_{abcd} + R_{acdb} + R_{adbc} = 0 \tag{2.38}$$

Other curvature tensors can be defined. For example, performing one contraction results in the Ricci tensor:

$$R_{ab} = R^{c}{}_{acb} \tag{2.39}$$

Since this is essentially a trace, and from linear algebra we know that the trace is equal to the sum of eigenvalues of the matrix, the Ricci tensor may be thought of as giving an average, or mean curvature. Contracting over the final pair of indices results in the Ricci scalar:

$$R = g^{ab} R_{ab} = R^{a}{}_{a} \tag{2.40}$$

Another curvature tensor which is useful in the study of conformal spacetimes is the Weyl conformal tensor, defined by

$$R_{abcd} = C_{abcd} + \frac{1}{2}\left(g_{ac}R_{bd} - g_{ad}R_{bc} - g_{bc}R_{ad} + g_{bd}R_{ac}\right) + \frac{1}{6}\left(g_{ad}g_{bc} - g_{ac}g_{bd}\right) R \tag{2.41}$$

C_{abcd} shares all the symmetries of the Riemann tensor, and in addition has the property that a contraction over the first and third indices is equal to zero:

$$C^{a}{}_{bad} = 0 \tag{2.42}$$

It is preserved under conformal transformations of the metric: i.e. when

$$\bar{g}_{ab} = \Omega^2 g_{ab} \tag{2.43}$$

then

$$\bar{C}_{abcd} = C_{abcd} \tag{2.44}$$

Conformal spacetimes are an important subfield of general relativity.

One of the more famous and useful identities is the Bianchi identity, which is given by

$$\nabla_a R^{b}{}_{cde} + \nabla_e R^{b}{}_{cad} + \nabla_d R^{b}{}_{cea} = 0 \tag{2.45}$$

This identity is especially useful in obtaining a curvature tensor with zero divergence, known as the Einstein tensor. Contracting the Bianchi identity over b and e and c and d, and using the symmetries, a tensor can be found that has zero divergence. That tensor, G_{ab},called the Einstein Tensor, is

$$G_{ab} = R_{ab} - \frac{1}{2} R g_{ab} \tag{2.46}$$

Since the stress-energy tensor T_{ab} also has zero divergence, Einstein was led to believe these two tensors were proportional. This equation and its consequences shall be the subject of the next chapter.

Example: Compute the Riemann curvature, Ricci tensor, and Ricci scalar for the surface of a sphere.

Solution: Using the symmetries of the Riemann tensor, it can be shown that there is only one independent non-zero component, R_{2323}. There are three other non-zero components, which are $R_{3232} = -R_{2332} = R_{3223} = -R_{2323}$. Computing:

$$R_{2323} = g_{22}R^2{}_{323} = g_{22}\left(\frac{\partial\Gamma^2_{33}}{\partial x^2} - \frac{\partial\Gamma^2_{23}}{\partial x^3} + \Gamma^c_{33}\Gamma^2_{2c} - \Gamma^c_{32}\Gamma^2_{3c}\right)$$
$$= R^2\left(-\cos^2\theta + \sin^2\theta + \cot\theta\sin\theta\cos\theta\right) = R^2\sin^2\theta$$

The Ricci tensor has two components on the diagonal, given by

$$R_{ac} = g^{bd}R_{abcd} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

Finally, the Ricci scalar, which is exactly twice the Gauss curvature K, is given by

$$R = g^{ac}R_{ac} = g^{22}R_{22} + g^{33}R_{33} = \frac{2}{R^2}$$

2.6 Exercises

1. Let $B^a = (1, 2, 0, -1)$ and $C^a = (-2, 0, 3, 0)$. Find the following: (A) $2B^a + C^a$ (B) B^aC^b (C) $\eta_{ab}B^aB^b$ (D) $B^aC^a - B^bC^b$ (E)$\eta_{ab}B^aC^b$ (F) B_a

2. State the rank of the following tensors: (A) C^{ab} (B)D^b_c (C) $A^{bcd}B_{de}$

3. Show that in the special case of a tensor Ψ of rank zero, $\nabla_a\nabla_b\Psi = \nabla_b\nabla_a\Psi$ if and only if $\Gamma^c_{ab} = \Gamma^c_{ba}$.

4. (A) Find the metric corresponding to cylindrical coordinates. (B)Compute the Christoffel symbols (C) calculate $\nabla_a\nabla_b f$, where $f = re^\phi$.

5. Construct a totally antisymmetric tensor U_{abc} from an arbitrary tensor of rank (0,3), C_{abc}.

6. (A) Compute the Christoffel symbols for the metric $ds^2 = dr^2 + e^{2r}d\phi^2$. (B) Compute the components of the Riemann tensor (C) Ricci (D) Ricci scalar.

7. By explicit calculation, prove equation 2.35.

8. Prove equation 2.46.

9. Prove in flat space that $\nabla_a\nabla_b F^{ab} = 0$, where F^{ab} is an arbitrary anti-symmetric tensor.

10. Use the Ricci identity for second rank tensors (note: you need to put this in somewhere!) to prove that $\nabla_a \nabla_b F^{ab} = 0$ in an arbitary spacetime, where F^{ab} is an arbitrary antisymmetric tensor.

11. Consider a spacetime which is three-dimensional, with one time dimension and two space dimensions. Let the space dimensions be those of a cylinder of fixed radius, R. Now assume the components g_{tt} and g_{zz} are both functions of z given by $g_{tt} = 1 - 1/z$ and $g_{zz} = 1/(1 - 1/z)$. Compute the Christoffel symbols for this metric. (B) Compute the Ricci tensor for this metric.

Chapter 3

Special Relativity and Electromagnetism

3.1 Special Relativity

In order to warm up to the general theory, we'll take a look at the special theory from a more advanced viewpoint.

Accelerated Systems in Special Relativity

Contrary to popular belief, accelerated systems can be handled in special relativity. Essentially, the assumption is that there is a continuum of frames, each with a separate, constant velocity, with smooth transition from one to the next. Taken altogether, it is then possible to compute the dynamics of accelerated objects such as elementary particles and super-spacecraft. It should be emphasized, however, that this is still an open area of research.

We start with the elementary fact about four-velocities, u^b:

$$u^b u_b = -c^2$$

Taking the proper time derivative, which is the time experienced by an observer traveling with the spacecraft, gives

$$\frac{d}{d\tau}\left(u^b u_b\right) = 2\frac{du^b}{d\tau}u_b = a^b u_b = 0$$

In the initial rest frame, we have

$$u^b = \left(\frac{dt}{d\tau}, \frac{dx}{d\tau}\right) = (1, 0)$$

From the previous equation, therefore,

$$\eta_{bc}u^b a^c = -c^2 u^0 a^0 + u^1 a^1 = 0 \rightarrow a^0(0) = 0$$

Hence for an accelerated system we must solve simultaneously the following equations:

$$\eta_{bc}u^b u^c = -c^2 \tag{3.1}$$

$$\eta_{bc}u^b a^c = 0 \tag{3.2}$$

$$\eta_{bc}a^b a^c = g^2 \tag{3.3}$$

with initial conditions $u^0 = 1$ and $a^0 = 0$, where g is the acceleration of the system provided by some force, say a rocket thrust, or an electromagnetic field. Writing these out explicitly gives:

$$-c^2 {u^t}^2 + {u^x}^2 = -c^2$$

$$-c^2 u^t a^t + u^x a^x = 0$$

$$-c^2 {a^t}^2 + {a^x}^2 = g^2$$

Solve the second equation for a^x and plug into the third equation, yielding:

$$-c^2 {a^t}^2 + c^4 {a^t}^2 \frac{{u^t}^2}{{u^x}^2} = g^2$$

Some algebra and the use of the first equation results in:

$$-c^2 {a^t}^2 \left(-1 + c^2 \frac{{u^t}^2}{{u^x}^2}\right) = -c^2 {a^t}^2 \frac{\left(-{u^x}^2 + c^2 {u^t}^2\right)}{{u^x}^2} = c^2 {a^t}^2 \left(\frac{c^2}{{u^x}^2}\right) = g^2$$

Hence

$$a^t = \frac{\pm g u^x}{c^2}$$

Plug this back into the third equation,

$$-c^2 {a^t}^2 + {a^x}^2 = \frac{-c^2 g^2 {u^x}^2}{c^4} + {a^x}^2 = g^2$$

and solve for a^x

$${a^x}^2 = g^2 \left(1 + \frac{{u^x}^2}{c^2}\right) = g^2 \frac{\left(c^2 + {u^x}^2\right)}{c^2} = g^2 {u^t}^2$$

so finally

$$a^x = \pm g u^t$$

Newton's law of motion reads

$$\frac{d^2x}{dt^2} = F/m = g$$

where g is the acceleration created by the physical field, while Einstein's special relativity gives a pair of equations,

$$\frac{d^2t}{d\tau^2} = \frac{g dx/d\tau}{c^2}$$

$$\frac{d^2x}{d\tau^2} = g\frac{dt}{d\tau}$$

This pair of differential equations is much more difficult to solve, in general, than Newton's equation.

Example: The Relativistic Rocket. We assume it's a super rocket that can proceed forever at some constant acceleration without using any fuel, thus with unchanging total mass. First, we recast the equations in terms of the components of the four-velocity:

$$\frac{du^t}{d\tau} = \frac{gu^x}{c^2}$$

$$\frac{du^x}{d\tau} = gu^t$$

Take the derivative of the second equation, and substitute into it the first:

$$\frac{d^2u^x}{d\tau^2} = g\frac{du^t}{d\tau} = \frac{g^2u^x}{c^2}$$

This can be integrated easily:

$$u^x = A\cosh\left(\frac{g}{c}\tau\right) + B\sinh\left(\frac{g}{c}\tau\right)$$

This is the general solution for the x-component of the four-velocity. Initially, the rocket is at rest, so $u^x = v\gamma = 0 = A$, so that

$$u^x = B\sinh\left(\frac{g}{c}\tau\right)$$

This solution can be plugged back into the equation for u^t:

$$\frac{du^t}{d\tau} = \frac{g}{c^2}B\sinh\left(\frac{g}{c}\tau\right)$$

$$u^t = \frac{B}{c}\cosh\left(\frac{g}{c}\tau\right) + D$$

Initial conditions on u^t give

$$u^t(0) = 1 = \frac{B}{c} + D \rightarrow D = 1 - \frac{B}{c}$$

Hence

$$u^t = \frac{B}{c}\left(\cosh\left(\frac{g}{c}\tau\right) - 1\right) + 1$$

To determine the constant B, it is necessary to use the initial conditions on the acceleration. Note that

$$\eta_{bc}a^b a^c = \eta_{00}{a^0}^2 + \eta_{11}{a^1}^2 = -c^2\tfrac{g^2}{c^4}B^2\sinh^2\left(\tfrac{g}{c}\tau\right) + \tfrac{B^2g^2}{c^2}\cosh^2\left(\tfrac{g}{c}\tau\right) =$$

$$= \tfrac{B^2g^2}{c^2}\left(-\sinh^2\left(\tfrac{g}{c}\tau\right) + \cosh^2\left(\tfrac{g}{c}\tau\right)\right) = \tfrac{B^2g^2}{c^2} = g^2 \rightarrow B = \pm c$$

Without loss of generality, we can choose the positive root, which gives a positive velocity in the x-direction. This only means we've chosen the direction of motion of the rocket to coincide with the positive x-axis. So the four velocity components are, at last,

$$u^t = \left(\cosh\left(\frac{g}{c}\tau\right) - 1\right) + 1$$

$$u^x = c\sinh\left(\frac{g}{c}\tau\right)$$

With these two expressions, it is easy to find the coordinates of the super-rocket as a function of proper time:

$$t = \frac{c}{g}\sinh\left(\frac{g}{c}\tau\right) + t_0$$

$$x = \frac{c^2}{g}\cosh\left(\frac{g}{c}\tau\right) + x_0$$

The integration constants can be taken to be zero. We are now ready to answer the question concerning a trip to the center of the galaxy at one gee acceleration without mass loss. The distance is about 30,000 light years. Setting this equal to the x-coordinate and solving for τ gives an answer of about 10.5 years of subjective time. Plugging this into the time equation tells us how much time passes back on Earth during this voyage—a little over 30,000 years!

3.2 Relativistic Lagrangian Mechanics

3.3 Electromagnetism in Special Relativity

3.4 Exercises

1. Photon rocket. Assume you have a matter-antimatter rocket in which the fuel is turned completely into energy and shot out the back end at velocity c. Derive the relativistic equations for this rocket, and design a mission to Alpha Centauri, 4.3 light years away, based on your design parameters. (Ignore the problem of sweeping through the interstellar gas at relativistic velocities, which should in fact roast all those on board.)

2. Analyze the relativistic harmonic oscillator with equilibrium at x=0, assuming the four- acceleration has magnitude kx.

Chapter 4

Geometry and Matter

Einstein's work was ultimately built on that of Gauss and Riemann. Gauss invented many of the basic ideas of the curvature of a surface, and the differential geometry on that surface. Riemann developed the mathematical theory of general curved spaces, and even guessed that such spaces ought to have some fundamental importance for physical theory. Unfortunately, he died of consumption before the age of forty.

Einstein, after his great successes of 1905, embarked on a ten-year quest to develop a theory of gravity using the mathematics of Riemann. His answer was the general theory of relativity. The motion of particles in space under the influence of gravity was due to the local curvature, not from an action at a distance.

4.1 Einstein's Equation

Einstein guessed that gravitation was the result of the curvature of spacetime, rather than the effect of an action at a distance, as in Newtonian gravity theory. Newton's theory had been highly successful in predicting the motions of the planets and the tides, but failed to predict the correct perihelion shift of Mercury. It was this experimental result that Einstein wanted to calculate. His first attempt at a theory was to posit

$$R_{ab} = \kappa T_{ab}$$

and with this he calculated the correct advance in the perihelion of Mercury. He was dissatisfied with the equation, however, and a couple years later struck upon

$$R_{ab} - \frac{1}{2} R g_{ab} = \kappa T_{ab} \tag{4.1}$$

This is called Einstein's Equation. Hilbert derived it from a variational principle and published it five days before Einstein, but it is likely he got the idea from Einstein's lectures and prior published work. The left hand side is the Einstein tensor, introduced in the last chapter, and the right hand side is the stress-energy tensor, which describes the matter and its properties, such as energy density and pressure in various directions. The main reason for adding the additional term is to guarantee that the divergence of the left hand side, like the right hand side, vanishes. This is a consequence of the conservation of energy. Otherwise, the divergence of the Ricci tensor would have to be set equal to zero, forming an unaesthetic additional constraint. Einstein wasn't finished, however. His equation predicted a dynamic, evolving universe. At that time the only galaxy discovered was the Milky Way, and the others, though found by astronomers, were thought to be clouds of gas within the our galaxy. The universe appeared to be static and unchanging. Since his equations were inconsistent with a static universe, Einstein introduced the cosmological constant term, which allowed for static solutions. His equation became

$$R_{ab} - \frac{1}{2}Rg_{ab} + \Lambda g_{ab} = \kappa T_{ab} \tag{4.2}$$

which is called Einstein's Equation with cosmological constant. Einstein later called the cosmological constant the greatest blunder of his life, because by 1930 Hubble had discovered that the universe was indeed expanding. The blunder, however, wouldn't quite go away, and much later Guth, Linet, and others would use such a cosmological constant to drive an early period of exponential inflation. It may be that the cosmological constant will ultimately turn out to be Einstein's greatest success.

There was one more variant on Einstein's equation, which consisted of replacing the factor of in equation 4.1 by 1/4. This made the left hand side trace-free, which Einstein thought might be advantageous in describing the force that held the nucleus together–which he believed was the gravitational force. Now, of course, this idea has been discredited.

4.2 The Stress-Energy Tensor

The right hand side, called variously the stress-tensor, the stress-energy tensor, or the energy- momentum tensor, is critical in spacetimes where matter is found, particularly inside stars, which may be modeled as fluids, but also in regions of field energy of some kind, such as electrostatic fields or relativistic fields, such as Klein-Gordon or Dirac. Cosmology, also requires a fluid, the points of which represent galaxies.

There are various ways to construct stress-energy tensors, and of course it's critical that it make sense, since otherwise in non-empty space it will be impossible to get meaningful solutions. Generic matter it is usually assumed to be a perfect fluid, which means a fluid without viscosity or dissipative effects of any kind. This tensor is given by

$$T_{ab} = \rho u_a u_b + p(u_a u_b - g_a b) \tag{4.3}$$

This same tensor is used in fluid mechanics. u^a is the four-velocity of the fluid, while ρ is the energy density, and p is the pressure. Note that since the metric can always be written with a time coordinate parallel to u^a, the pressure term is effectively a spatial metric only.

In fields there is a formalism involving Lagrangians. The Lagrangian density is varied with respect to the metric, and this is taken to be the stress-energy tensor. This method is the method of choice for most of those in general relativity, though there are two other ways of getting that stress energy, one called the canonical stress-energy (which is not always symmetric), and then a symmetrization of that asymmetric stress-energy. This latter stress-energy is used extensively in quantum field theory, and it isn't clear which of the two approaches is truly correct, if either.

The stress energy for a field is given, in general, by:

$$T_{ab} = -\frac{\alpha_M}{8\pi}\frac{1}{\sqrt{-g}}\frac{\delta}{\delta g^{ab}}\sqrt{-g}\pounds \tag{4.4}$$

The Lagrangian for the field must be such that varying with respect to ψ yields the physical equation describing the field, such as Maxwell's equations. It will then follow that variation with respect to the metric will result in the stress-energy of the field. The Lagrangian can depend on the metric and ψ and its derivatives, usually only to first order, though some have investigated Lagrangians which are second order. Note that the constant, α_M, depends on the field in question. In principle, it would be necessary to do an experiment to determine its exact value, in all but possibly perfect fluid cases. In practice, it is often hard to be sure what number is appropriate, because it is not usually possible to measure the gravitational response of a given stress-energy.

4.3 Einstein's equation in the weak field limit

It's important to relate Einstein's equation to Newtonian physics, to make certain that its predictions will be consistent with Newton's gravity theory when fields are weak. In addition, this will allow the identification of the Einstein constant.

First, particles travel on geodesics:

$$\frac{\partial^2 x^a}{ds^2} + \Gamma^a_{bc}\frac{dx^b}{ds}\frac{dx^c}{ds} = 0 \tag{4.5}$$

In the limit of small velocities and weak fields, the temporal component of dx^b/ds will dominate over the spatial components. This is clear if units in which c=1 are used, as when $x^0 = ct$. In that case, the ordinary velocities are fractions of light speed. In addition, $dx^0/ds \approx 1$. Then

$$\frac{\partial^2 x^a}{ds^2} \approx -\Gamma^a_{00} \tag{4.6}$$

We will also neglect the variation in time of the source of the gravity field. This means $\Gamma^0_{00} = 0$, so that the time component of the above equation means that $s \to x^0$. Then the above equation becomes

$$\frac{\partial^2 x^k}{d{x^0}^2} \approx -\Gamma^k_{00} \tag{4.7}$$

where $k = 1, 2, 3$. This equation looks very much like Newton's second law. Inserting the definition of the Christoffel symbols yields

$$\Gamma^k_{00} = \frac{1}{2} g^k\lambda \left(2\frac{\partial g_{\lambda 0}}{\partial x^0} - \frac{\partial g_{00}}{\partial x^\lambda} \right) \approx -\frac{1}{2}\eta^{k\lambda}\frac{\partial g_{00}}{\partial x^\lambda} \tag{4.8}$$

The right hand side is actually the ordinary three-gradient g_{00}. Define, therefore,

$$g_{00} = 1 + \frac{2}{c^2}\phi \tag{4.9}$$

where g_{00} and ϕ are functions of the spatial coordinates only. This only corresponds to swapping out the function g_{00} in favor of ϕ. In addition, reinsert the value $x^0 = ct$. The results is

$$\frac{d^2 x^k}{dt^2} \approx -\nabla\phi \tag{4.10}$$

Thus in the limit of slow particles and weak fields, g_{00} acts like the classical potential. There remains to show that it satisfies a Poisson equation, at least in the weak field limit. In this limit, the '00' components always dominate. So Einstein's equation, for the 00 component, reads

$$R_{00} = \kappa\left(T_{00} - \frac{1}{2}g_{00}T\right) \approx \frac{1}{2}\kappa T_{00} = \frac{1}{2}\kappa c^2\rho \tag{4.11}$$

Neglecting nonlinear and time-derivative terms in the Ricci tensor, obtain

$$R_{00} = \frac{\partial\Gamma^a_{00}}{\partial x^a} - \frac{\partial\Gamma^a_{0a}}{\partial x^0} + \Gamma^a_{00}\Gamma^b_{ab} - \Gamma^a_{c0}\Gamma^c_{0a} \approx \frac{\partial\Gamma^k_{00}}{\partial x^k} \tag{4.12}$$

where $k = 1, 2, 3$. So

$$R_{00} \approx \frac{\partial\Gamma^k_{00}}{\partial x^k} = -\frac{1}{2}\eta^{kl}\frac{\partial^2 g_{00}}{\partial x^k \partial x^l} \approx \frac{1}{c^2}\nabla^2\phi \tag{4.13}$$

Putting this all together, the potential ϕ satisfies

$$\nabla^2\phi = \frac{1}{2}\kappa c^2\rho \tag{4.14}$$

which is Newton's gravity equation. The constant κ should therefore be identified as

$$\kappa = \frac{8\pi G}{c^4} \tag{4.15}$$

For weak fields, therefore, Einstein's theory reduces to Newton's, and the two theories are consistent. And as will be seen in chapter 6, Einstein's equations will return greater accuracy when the gravitational fields are somewhat stronger. It is still not known whether the theory is correct for extremely strong fields.

4.4 Lagrangians and Stress-Energies

The Einstein tensor, which gives something that could described as a divergence-free average of the curvature, is supposedly equal to the material in the spacetime, which is known as the stress-energy. The matter and energy in spacetime is represented by the stress-energy tensor, also called the energy-momentum tensor, which is a tensor of second rank just like the Einstein tensor. In most cases of interest, it gives the equivalent of the average of numerous interactions of fields at extremely short distances. Despite the fact GR is the best-tested theory in physics, there remain questions about how to represent matter. In addition, inserting the matter terms generally makes the equations much harder to solve. The primary applications are in the areas of stellar structure and cosmology.

4.4.1 Lagrangian Formulations

In some sense, anything placed opposite the Einstein Tensor can be regarded as the stress-energy of something. In fact, any metric can be written down, the Christoffel symbols and curvature terms calculated, and then inserted into the Einstein tensor. Whatever results–not likely to be zero, in general–can then be declared the stress-energy of the spacetime.

This is obviously not very satisfying from a physical or philosophical view. Still, some very interesting spacetimes have been discovered in this way. Most of the time, however, the result is impossible to interpret physically.

The stress-energy, therefore, is best derived from some general principles that make sense and connect with classical notions of pressure and density. And, when dealing with fields, it is best to obtain the stress-energy from a variational principle. Nature appears to have a preference for extremes–minimums and maximums–and a procedure that returns an extremum is considered more likely to be correct.

As it turns out, there are two different Lagrangian-based stress-energies, and they don't always agree with each other. The first is called the canonical stress energy. Using the Lagrangian for a field, an expression is constructed with explicit zero divergence. The Lagrangian is assumed, in most cases, to depend only on the field ϕ and its gradient, $\nabla_a \phi$, though in fact there is no reason it cannot depend on higher-order derivatives. Among relativists, the method of choice is to find the Lagrangian density that results in the field equation for ϕ when varied with respect to ϕ, and then gives the stress-energy when varied with respect to the metric, g_{ab}. "Taking variations" is very similar to taking derivatives and using the chain rule. The starting point is the variation of an integral over paths in function space for which we seek a stationary (maximal or minimal) solution:

$$\delta \int \pounds \left(\phi, \nabla_a \phi\right) \sqrt{-g} d^4 x = 0 \tag{4.16}$$

Moving the δ inside the integral and using the chain rule results in:

$$\int \left(\frac{\partial \pounds}{\partial \phi} \delta\phi + \frac{\partial \pounds}{\partial \nabla_a \phi} \delta \nabla_a \phi \right) \sqrt{-g} d^4 x = 0 \tag{4.17}$$

The variation operator δ is assumed to commute with the partial derivative operator, so that

$$\frac{\partial \pounds}{\partial \nabla_a \phi} \delta \nabla_a \phi = \frac{\partial \pounds}{\partial \nabla_a \phi} \nabla_a \delta\phi = \nabla_a \left(\frac{\partial \pounds}{\partial \nabla_a \phi} \delta\phi \right) - \left(\nabla_a \frac{\partial \pounds}{\partial \nabla_a \phi} \right) \delta\phi \tag{4.18}$$

The far right-hand term comes from an integration by parts, and results in a surface term of the form

$$\int \nabla_a \left(\frac{\partial \pounds}{\partial \nabla_a \phi} \delta\phi \right) \sqrt{-g} d^4 x \tag{4.19}$$

This perfect divergence allows the use of Gauss's general theorem which relates volume integrals to surface integrals. In this case, the volume integral is over all space, so the surface integral would be carried out over a sphere of infinite radius. At infinity, the field may be assumed to have gone to zero, and hence this term will not contribute to the integral. The remaining two terms will not in general be zero, and for the integral to be zero it follows that what's underneath the integral sign must be zero, namely:

$$\frac{\partial \pounds}{\partial \phi} - \bigtriangledown_a \frac{\partial \pounds}{\partial \bigtriangledown_a \phi} = 0 \tag{4.20}$$

This is the formula satisfied by extremal functions in the function space. The analogies with one- variable calculus and maxima and minima are clear. It should also be clear that not all equations have Lagrangians; the equations that do have them are special.

For fields having Lagrangians depending on the second derivative of the field ϕ, an additional term is required:

$$-\bigtriangledown_a \bigtriangledown_b \frac{\partial \pounds}{\partial \bigtriangledown_a \bigtriangledown_b \Psi} + \bigtriangledown_a \frac{\partial \pounds}{\partial \bigtriangledown_a \Psi} - \frac{\partial \pounds}{\partial \Psi} = 0 \tag{4.21}$$

Note the pattern in the terms, and the alternation of signs, which is a general result. Proving this relationship is left as an exercise.

After finding a Lagrangian for a given function, it is possible to find the stress energy. The stress-energy is found either by following the canonical procedure, or by taking the variation with respect to the metric. In principle, experiments or observations should decide which of these methods is best. Unfortunately, the gravitational effects are generally very small, so that it isn't possible to decide which formulation holds in reality, if either one does.

The canonical stress-energy is developed by construction. We will assume that the Lagrangian is a function of ψ and its first derivatives only (the second derivative case can also be calculated and is straightforward, though it is considerably more involved). Let a coordinate transformation be given, ${x'}_a = x_a + s_a$, where s_a is an infinitesimal coordinate- independent displacement, and let $\pounds$ be a Lagrangian which depends also on the second derivatives of the field. Then

$$\delta\pounds = \pounds(x') - \pounds(x) = s^a\partial_a\pounds = \frac{\partial\pounds}{\partial\Psi}\delta\Psi + \frac{\partial\pounds}{\partial(\partial_a\Psi)}\delta\partial_a\Psi \tag{4.22}$$

In addition, evidently, the following relationships hold:

$$\delta\Psi = \Psi(x') - \Psi(x) = s^a\partial_a\Psi \tag{4.23}$$

$$\delta\partial_a\Psi = \partial_a\Psi(x') - \partial_a\Psi(x) = s^b\partial_b(\partial_a\Psi) \tag{4.24}$$

The expression for $\partial\pounds/\partial\Psi$can be eliminated from equation 4.22 by using equation 4.20.

Using equations 4.20, and equations 4.23 through **??**, and remembering that repeated upstairs and downstairs indices are dummy indices, equation 4.22 can be rearranged in such a way as to produce an expression with explicitly zero divergence. This, then, is the desired expression for the canonical stress-energy:

$$T^{ab} = \frac{\partial\pounds}{\partial(\partial_a\Psi)}\partial^b\Psi - g^{ab}\pounds \tag{4.25}$$

In general, this expression may not even be symmetric in its indices. There are procedures for making it so, by adding in divergence-free terms.

To find the stress-energy via variation of the metric is straightforward and always yields a symmetric tensor. From an aesthetic point of view, this is highly preferable, and until there is definitive experimental evidence to the contrary, is the method of choice. Recall the following variation:

$$\delta\sqrt{-g} = -g_{ab}\sqrt{-g}\delta g^{ab} \tag{4.26}$$

Then the stress energy can be defined as

$$T_{ab} = \frac{\alpha}{\sqrt{-g}}\frac{\delta}{\delta g^{ab}}\sqrt{-g}\pounds \tag{4.27}$$

The parameter α must be chosen for each different case. Generally, the choice is made so that the Hamiltonian takes on the intuitively correct form, consistent with energies in classical physics. Obviously, without experimental evidence, it isn't clear what choice should be made. No one knows, for example, how strongly electromagnetic fields gravitate.

4.5 The Klein-Gordon Equation

The first relativistic quantum field equation was derived by Schrodinger, and later also obtained by Kaluza and Klein:

$$\nabla^a \nabla_a \psi + \frac{m^2 c^2}{\hbar^2} \psi = 0 \tag{4.28}$$

The Lagrangian for this equation is given by

$$\pounds = \frac{1}{2} \nabla_a \psi \nabla^a \psi - \frac{1}{2} \frac{m^2 c^2}{\hbar^2} \psi^2 \tag{4.29}$$

In this simple case the two stress-energies are identical except for a numerical constant. Obtain:

$$T_{ab} = \nabla_a \psi \nabla_b \psi - g_{ab} \left(\frac{1}{2} \nabla_a \psi \nabla^a \psi + \frac{1}{2} \frac{m^2 c^2}{\hbar^2} \psi^2 \right) \tag{4.30}$$

Let's try this another way, using the Hamiltonian, which is an energy. The Hamiltonian for a field ψ is given by

$$H = \int \psi_{,t} \frac{\partial \pounds}{\partial \psi_{,t}} d^3x - L = \int \left(\psi_{,t} \frac{\pounds}{\partial \psi_{,t}} - \pounds \right) d^3x \tag{4.31}$$

The integrand is obviously an energy density, and hence could be taken as the definition of the energy density in the stress-energy tensor. By analogy with classical mechanics, this expression should return a kinetic energy term, which is one-half the momentum squared divided by the mass, along with effective potential energy terms. The field momentum, π is defined by

$$\pi = \frac{\partial \pounds}{\partial \psi_{,t}} \tag{4.32}$$

Now, the overall normalization of a given Lagrangian is arbitrary, as far as generating the correct field equation. It is generally chosen, however, so that the Hamiltonian will have a leading term equal to $\frac{1}{2}\pi^2$. This is still a rather arbitrary criterion, because ultimately, the correct choice will be that which gives the correct gravity field, which ultimately can only be determined by experiment. In the case of the Klein-Gordon equation,

$$\mathcal{H} = \pi \frac{\partial \psi}{\partial \sqcup} - \mathcal{L}_{\mathcal{KG}} = \left(\frac{\infty}{\in \alpha} - \frac{\infty}{\triangle \alpha} \right) \}'' \frac{\partial \psi}{\partial \sqcup}^{\in} + \alpha \nabla \psi \cdot \nabla \psi + \alpha \frac{\Updownarrow^{\in} \rfloor^{\in}}{\zeta^{\in}} \psi^{\in} \tag{4.33}$$

By inspection, $\alpha = \frac{1}{2}$ gives the correct energy density, hence the correct value of α by convention.

4.6 The Faraday Tensor

The stress-energy of electromagnetic fields is already known, in principle. One way or another, it should contain within it the squares of the electric and magnetic fields. There is in fact only one natural invariant that can be created out of the Faraday tensor, F_{ab}:

$$\pounds = F_{ab} F^{ab} \tag{4.34}$$

where

$$F_{ab} = \nabla_a A_b - \nabla_b A_a \tag{4.35}$$

First, let's verify that this Lagrangian gives the correct equations when varied with respect to A^c.

Calculating the stress energy is straightforward:

$$T_{ab} = F_{ac} F_b{}^c - \frac{1}{4} g_{ab} F_{cd} F^{cd} \tag{4.36}$$

Note that no experiment has determined how much gravity field a given electromagnetic energy density creates. (Note to self: the bending of light by the sun ought to tell us something!)A variation on this theme is the Proca equation, which corresponds to a massive spin-1 field (see the exercises).

4.7 Perfect Fluids

4.8 Exercises

1. Derive Lagrange's equation in the case where the $\pounds = \pounds(\psi, \nabla_a \psi, \nabla_a \nabla_b \psi)$.

2. Find a suitable Lagrangian for the Proca equation, and derive the corresponding stress-energy tensor.

Homework, Due Friday, October 27.

1. Derive the equations for matter under pressure inside a spherical star, but with a cosmological constant term. To do this, follow the notes, step by step, starting from Einstein's equations with cosmological constant. For extra credit, find a solution.

2. Let the Lagrangian for a physical system be given by

$$S_{ab} S^{ab} + A_a A^a$$

, where S is the same as the electrodynamic F except symmetric instead of anti-symmetric, and A is the field. Derive the equation for A, and the stress-energy for the system.

3. (Review) Let A=(1,2,3,4), B=(-2,-1,1,1). Find all quantities as in previous homework problem of the same type.

4. Let a two-dimensional spacetime be given by

$$ds^2 = f(x, y)\left(dx^2 + dy^2\right)$$

. Write down the Christoffel symbols and curvature terms, assuming f depends on both y and x. Try to find a solution that gives R=constant.

Chapter 5

The Schwarzschild Solutions

While Einstein's theory is highly aesthetic, it only has meaning if it can be related to observations and the results of experiments. Einstein believed that the equation could only be solved approximately. Schwarzschild, a German medic on the Russian front, managed to find two exact solutions, both for spherical symmetry, one of them for an exterior vacuum, the other an interior fluid with constant energy density. Einstein, receiving the manuscript, was enthusiastic, and replied with his congratulations by return mail. Shortly thereafter, Schwarzschild died tragically of a skin disease contracted while serving for the Russian Army during the Great War.

The prescription for solving the Einstein equation is straightforward:

(1) Choose an appropriate general form for the metric.

(2) Compute the Christoffel symbols, which, given the symmetries on the indices, amounts to less than 32 terms.

(3) Using the Christoffel symbols, obtain the Ricci tensor and Ricci scalar components; load them into Einstein's equation.

(4) Solve the resulting system by any means you are able.

Choosing the general form of the metric can usually be done intuitively, and often involves assumptions based on symmetry and on the exact form of the stress-energy tensor. More rigorous is to find the Killing vectors appropriate to a given symmetry and use them to write down a trial metric. From this trial metric, some simplifying changes may be made through coordinate transformations. This is one of the great principles Einstein introduced, the idea that the physics shouldn't depend on what system of coordinates is used, hence all are equivalent, and it therefore makes sense to choose the system that makes the problem easier.

5.1 Exterior Schwarzschild Solution

A static, spherically symmetric body should have a gravity field that is independent of time. For a trial metric, it is natural to choose

$$ds^2 = -e^\nu dt^2 + e^\lambda dr^2 + r^2(d\theta^2 + sin^2\theta d\phi^2) \tag{5.1}$$

where $\nu = \nu(r)$ and $\lambda = \lambda(r)$. Next, the Christoffel symbols must be calculated. Note that given the symmetry in the lower two indices, at most 32 must be calculated rather than 64. (There are ways of reducing this number further, but generally it's not worth the trouble). Note that off-diagonal metric terms are zero, so in fact many terms drop immediately out in the course of the calculations.

$$\begin{array}{ccc} \Gamma^0_{01} = \frac{\nu'}{2} & \Gamma^1_{00} = \frac{\nu'}{2}e^{\nu-\lambda} & \Gamma^1_{11} = \frac{\lambda'}{2} \\ \Gamma^1_{22} = -re^{-\lambda} & \Gamma^1_{33} = -r\sin^2\theta e^{-\lambda} & \Gamma^2_{12} = \frac{1}{r} \\ \Gamma^2_{33} = -\sin\theta\cos\theta & \Gamma^3_{13} = \frac{1}{r} & \Gamma^3_{23} = \cot\theta \end{array} \tag{5.2}$$

In this case the Einstein equation reduces to $R_{ab} = 0$, so there is no need to compute the Ricci scalar. It can be verified that there are only four equations,

$$R_{00} = e^{\nu-\lambda}\left(\frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} + \frac{\nu'}{r}\right) = 0 \tag{5.3}$$

$$R_{11} = \left(-\frac{\nu''}{2} + \frac{\nu'\lambda'}{4} - \frac{\nu'^2}{4} + \frac{\lambda'}{r}\right) = 0 \tag{5.4}$$

$$R_{22} = e^{-\lambda}\left(-1 - \frac{r\nu'}{2} + \frac{r\lambda'}{r}\right) + 1 = 0 \tag{5.5}$$

$$R_{33} = \sin^2\theta R_{33} \tag{5.6}$$

with all others equal to zero. Multiplying the first equation by $e^{-\nu+\lambda}$ and adding to the second gives

$$\frac{\lambda}{r} + \frac{\nu}{r} = 0 \rightarrow \lambda = -\nu \tag{5.7}$$

Substituting this result into the third equation gives:

$$e^\nu(-1 - r\nu') + 1 = 0 \rightarrow (re^\nu)' = 1 \tag{5.8}$$

So

$$e^\nu = 1 + \frac{\beta}{r} \tag{5.9}$$

From the linear equations and weak-field limit,

$$\beta = \frac{2GM}{c^2} \tag{5.10}$$

The Schwarzschild exterior metric reads:

$$ds^2 = -\left(1 - \frac{2MG}{c^2 r}\right) c^2 dt^2 + \frac{dr^2}{\left(1 - \frac{2MG}{c^2 r}\right)} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right) \tag{5.11}$$

Notice that as r gets large, we recover the Minkowski metric, as we should. This very famous metric was the first exact solution found, and gave a precise correction to the perihelion shift of Mercury and predicted the bending of light by the sun. Much later, after the death of Einstein, it also correctly predicted the delay in the reflection of radar, the redshift of light from the sun, and the change in period of binary pulsars.

5.2 Schwarzschild Interior Solution

The interior solution for constant density requires an energy-momentum tensor. In general, this is written as

$$T_{ab} = \rho u_a u_b + \frac{p}{c^2}\left(u_a u_b - g_{ab}\right) \tag{5.12}$$

where u^a is the four-velocity, ρ is the energy density, and p is the pressure. For static, spherical symmetry, $u^a = (u^0, 0, 0, 0)$. Further, $g_{00}{u^0}^2 = 1$ so $u_0 = sqrt g_{00}$. Using these facts, the stress tensor can be written as

$$T_{ab} = \begin{pmatrix} \rho e^{\nu} & 0 & 0 & 0 \\ 0 & \frac{p}{c^2} e^{\lambda} & 0 & 0 \\ 0 & 0 & \frac{p}{c^2} r^2 & 0 \\ 0 & 0 & 0 & \frac{p}{c^2} r^2 \sin^2\theta \end{pmatrix} \tag{5.13}$$

Einstein's equations can be written in the form

$$R_{ab} = 8\pi \frac{G}{c^2}\left(T_{ab} - \frac{1}{2} T g_{ab}\right) \tag{5.14}$$

which, with the same trial metric as in the previous section, becomes

$$R_{00} = e^{\nu-\lambda}\left(\frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} + \frac{\nu'}{r}\right) = C\left(\frac{\rho}{2} + \frac{3p}{2c^2}\right) \tag{5.15}$$

$$R_{11} = \left(-\frac{\nu''}{2} + \frac{\nu'\lambda'}{4} - \frac{\nu'^2}{4} + \frac{\lambda'}{r}\right) = C\left(\frac{\rho}{2} - \frac{p}{2c^2}\right) \tag{5.16}$$

$$R_{22} = e^{-\lambda}\left(-1 - \frac{r\nu'}{2} + \frac{r\lambda'}{r}\right) + 1 = C\left(\frac{\rho}{2} - \frac{p}{2c^2}\right) r^2 \tag{5.17}$$

$$R_{33} = \sin^2\theta R_{22} \tag{5.18}$$

with all others equal to zero. The first two equations can be easily combined, as in the exterior solution, resulting in

$$\left(\frac{\nu' + \lambda'}{r}\right) = C\left(\rho + \frac{p}{c^2}\right) e^{-\lambda} \tag{5.19}$$

Note that if the left hand side is equal to zero, then either both ρ and p are zero, or one of them is negative. Negative mass has not been found, though negative pressures, corresponding to tensions, do make some sense, as will be seen in the chapter on cosmological defects. Equations 5.17 and 5.19 can be solved for ρ and p, while ρ and p can be eliminated from from equations 5.16 and 5.17. This results in the following system of three equations:

$$C\rho = e^{-\lambda}\left(\frac{1}{r^2} - \frac{\lambda'}{r}\right) - \frac{1}{r^2} \tag{5.20}$$

$$C\frac{p}{c^2} = \frac{1}{r^2} - e^{-\lambda}\left(\frac{1}{r^2} + \frac{\nu'}{r}\right) \tag{5.21}$$

$$\frac{e^{\lambda}}{r^2} = \frac{1}{r^2} - \frac{\nu'^2}{4} + \frac{\nu'\lambda'}{4} + \frac{\nu' + \lambda'}{2r} - \frac{\nu''}{2} \tag{5.22}$$

These three equations can be solved just by specifying ν, which makes equation 5.22 into a simple equation for λ. Solving it and plugging into the other equations yields the density and pressure, respectively. Such a fishing expedition is constrained, however, since $e^{\nu} = 1 - 2m/r$ must hold at the boundary, and the pressure and energy density, at least most of the time, must be positive definite, unless the matter is exotic.

Next, define

$$e^{-\lambda} = 1 - \frac{2m(r)}{r} \tag{5.23}$$

This is just a renaming of the metric component in favor of a function that not only facilitates matching at the boundary, but also leads to an easy geometric interpretation of the function $m(r)$. From equation 5.20, substitute and get

$$-C\rho = \frac{\left(re^{-\lambda}\right)'}{r^2} - \frac{1}{r^2} = \frac{\left(r - 2m\right)'}{r^2} - \frac{1}{r^2} = -\frac{2m'}{r^2} \tag{5.24}$$

Rearranging this equation gives

$$\frac{dm}{dr} = \frac{4\pi G}{c^2} r^2 \tag{5.25}$$

This can be recast as an integral:

$$m(r) = \frac{G}{c^2}\int_0^r 4\pi r^2 \rho dr \tag{5.26}$$

This makes m look like the rest mass of the star, given radius r, but in fact it isn't. That's because the integral doesn't include the factor of $\sqrt{-h}$, where $h = deth_{ab}$, with h_{ab} being the spatial metric. So because the expression doesn't include the right volume element, it isn't the correct rest mass. There should be a factor of $e^{\lambda/2}$ next to the dr. However, m does indeed correspond to the total mass as viewed by an exterior observer, rest mass plus internal energy plus gravitational energy:

$$m(r) = m_0(r) + U(r) + \Omega(r) \tag{5.27}$$

Note that the rest mass energy is

$$m_0(r) = \frac{G}{c^2} \int_0^r 4\pi r^2 \mu_0 n \left(1 - 2m/r\right)^{-1/2} dr \tag{5.28}$$

where μ_0 is the mass of a typical particle and n is the number density of those particles, and whereas the total internal energy is

$$U(r) = \frac{G}{c^2} \int_0^r 4\pi r^2 \left(\rho - \mu_0 n\right) \left(1 - 2m/r\right)^{-1/2} dr \tag{5.29}$$

Subtracting these from the total mass energy should give the gravitational potential energy of the system, which is

$$\Omega = -\frac{\kappa}{c^2} \int_0^r \rho \left[\left(1 - 2m/r\right)^{-1/2} - 1\right] 4\pi r^2 dr \approx -\frac{\kappa}{c^2} \int_0^r \frac{\rho m}{r} 4\pi r^2 dr \tag{5.30}$$

This last equation agrees with what is usually thought of as the gravitational potential energy, which lends support to the idea that m is the total mass of the star, including gravitational and thermal contributions. Using m as defined, it is possible to get the following equation from equation 5.21:

$$\nu' = 2\frac{m + 4\pi G p r^2/c^4}{r(r - 2m)} \tag{5.31}$$

Differentiating 5.31 and eliminating ν'' from the result by using equation 5.22 gives:

$$-\frac{8\pi G}{c^4} p' = e^{-\lambda} \left(\nu' + \lambda'\right) \frac{\nu'}{2r} \tag{5.32}$$

Then by inspecting equation 5.19:

$$\frac{p'}{c^2} = -\frac{\nu'}{2} \left(\rho + \frac{p}{c^2}\right) \tag{5.33}$$

Combining this and equation 5.31 results in the celebrated Tolman-Oppenheimer-Volkoff (TOV) equation:

$$p' = -\frac{\left(\rho c^2 + p\right)\left(mc^2 + 4\pi G p r^3/c^2\right)}{r(r - 2m)} \tag{5.34}$$

These two equations together with equations 5.25, 5.23 and an equation of state

$$p = p(\rho) \tag{5.35}$$

form a nonlinear system of first order equations for the structure of a static, non-rotating star. This set of equations will be studied in greater detail in the chapter on stellar structure. Schwarzschild solved them for the special case where the equation of state is given by

$$\rho = \rho_0 \tag{5.36}$$

where ρ_0 is a contant. Obviously this is not the best model of a stellar interior, since the density would be expected to decline along with the pressure when moving from the core to the exterior boundary of the star. However, some valuable information might be obtained, nonetheless, from the resulting exact solution. After all that analysis, the solution is easy. First, the mass equation 5.26 can be immediately integrated:

$$m(r) = \frac{4\pi G\rho r^3}{3c^2} \tag{5.37}$$

With this result in hand, equation 5.23 can also be integrated:

$$e^{-\lambda} = 1 - \frac{8\pi G\rho r^2}{3c^2} \tag{5.38}$$

For convenience, let

$$R^2 = \frac{3c^2}{8\pi G\rho} \tag{5.39}$$

Then

$$e^{-\lambda} = 1 - \frac{r^2}{R^2} \tag{5.40}$$

In the case of constant energy density, equation 5.33 can be solved for ν' and easily integrated:

$$Le^{\nu/2} = \frac{8\pi G}{c^2}\left(\rho + \frac{p}{c^2}\right) \tag{5.41}$$

with L the constant of integration. Substituting this into equation 5.19 yields

$$Le^{-\nu/2} = \frac{e^{-\lambda}}{r}\left(\nu' + \lambda'\right) \tag{5.42}$$

Since λ is known this gives a differential equation for ν. Multiplying by $re^{\lambda}/2$, rearranging and substituting, we obtain:

$$\left(1 - \frac{r^2}{R^2}\right)\nu' e^{\nu/2} + \frac{r}{R^2}e^{\nu/2} = \frac{1}{2}Lr \tag{5.43}$$

This is clearly an inhomogeneous differential equation for $f = e^{\nu/2}$:

$$\left(1 - \frac{r^2}{R^2}\right)f' + \frac{r}{R^2}f = \frac{1}{2}Lr \tag{5.44}$$

The particular solution is

$$f_p = \frac{1}{2}LR^2 \tag{5.45}$$

The homogeneous equation can be easily solved by separation of variables. The solution is

$$e^{\nu/2} = \frac{1}{2}LR^2 - B\left(1 - \frac{r^2}{R^2}\right)^{1/2} \tag{5.46}$$

Setting

$$A = \frac{1}{2}LR^2$$

the metric may be written as

$$ds^2 = \left(A - B\left(1 - \frac{r^2}{R^2}\right)^{1/2}\right)^2 c^2dt^2 - \left(1 - \frac{r^2}{R^2}\right)^{-1} dr^2 - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) \tag{5.47}$$

Using the convenient substitution

$$L = \frac{2}{3}\frac{8\pi\rho G}{c^2}A$$

equation 5.41 can be integrated:

$$\rho + \frac{p}{c^2} = \frac{2\rho A/3}{A - B(1 - r^2/R^2)^{1/2}} \tag{5.48}$$

The pressure vanishes at the periphery of the star at $r = r_0$, which yields

$$A = 3B\left(1 - \frac{r_0^2}{R^2}\right)^{1/2} \tag{5.49}$$

To evaluate B, invoke the condition that the interior and exterior metrics ought to be continuous at the boundary. This results in the **interior Schwarzschild metric**, given by

$$ds^2 = \left[\frac{3}{2}\left(1 - \frac{r_0^2}{R^2}\right)^{1/2} - \frac{1}{2}\left(1 - \frac{r^2}{R^2}\right)^{1/2}\right]^2 c^2dt^2 - \left(1 - \frac{r^2}{R^2}\right)^{-1} dr^2 - r^2d\Omega^2 \tag{5.50}$$

Some bounds may be placed on the mass of such a constant density star. First, the radius ought to be such that the Schwarzschild coordinate singularity doesn't occur. This means that

$$r_0 > 2MG/c^2 \tag{5.51}$$

which means that the mass must satisfy

$$M < \frac{c^2 r_0}{2G} \tag{5.52}$$

A stronger condition can be put on M by requiring that the pressure is never infinite. This means that the denominator of equation 5.48 must never vanish. This leads to

$$\frac{3}{2}\left(1-\frac{r^2}{R^2}\right)^{1/2} > \frac{1}{2} \tag{5.53}$$

which, using the expression for ρ, leads to

$$M < \frac{4}{9}\frac{c^2 r_0}{G} \tag{5.54}$$

This can also be turned into a restriction on M depending on central density, ρ_0.

5.3 Exercises

1. Verify the Christoffel symbols calculated for the Schwarzschild solution, together with the Ricci tensor components.

2. Set up Einstein's equations for a constant density interior solution in the case where there is a non-zero cosmological constant term.

3. Using 5.20 through 5.22, solve Einstein's equations for the case $\nu = -\alpha \ln r$, where α is a constant, which should be judiciously chosen to facilitate solution. Is the solution a reasonable stellar model?

Chapter 6

Tests of General Relativity

6.1 Red-shift of Light from the Sun

A light wave with origin at the surface of the sun will go through n oscillations in a given proper time, $\Delta\tau_s$. If ν_s is the frequency as measured at the surface of the sun, then

$$n = \nu_s \Delta\tau_s$$

When the light wave reaches Earth, it will have a new frequency. In a given proper time, $\Delta\tau_e$, n complete waves will be received. Thus

$$\nu_s \Delta\tau_s = \nu_e \Delta\tau_e$$

So

$$\nu_e = \nu_s \frac{\Delta\tau_s}{\Delta\tau_e}$$

The passage of coordinate time can be assumed to be the same on Earth as on the sun. Hence

$$\Delta t = \frac{\Delta\tau_e}{\sqrt{g_{00}(x_e^a)}} = \frac{\Delta\tau_s}{\sqrt{g_{00}(x_s^a)}}$$

This gives the ratio of proper times needed. Together with the previous equation, this results in:

$$\nu_e = \nu_s \frac{\Delta\tau_s}{\Delta\tau_e} = \nu_s \left(\frac{g_{00}(x_s^a)}{g_{00}(x_e^a)}\right)^{\frac{1}{2}} = \nu_s \left(\frac{1 - 2MG/c^2 r_s}{1 - 2MG/c^2 r_e}\right)^{\frac{1}{2}}$$

Note that

$$F(x, y) = \left(\frac{1 + x}{1 + y}\right)^{1/2} \approx 1 + \frac{1}{2}x + \frac{1}{2}y$$

Setting $x = 2MG/c^2r_s$ and $y = 2MG/c^2r_e$, substituting into the equation and doing some algebra gives:

$$\frac{\nu_e - \nu_s}{\nu_s} = \frac{MG}{c^2}\left(\frac{1}{r_e} - \frac{1}{r_s}\right)$$

The quantity on the right is definitely negative, which means the frequency measured on Earth is less than the frequency measured on the sun. This corresponds to a redshift in the light. This effect has been verified to high accuracy, including experiments done with lasers in tall towers, a' la Galileo!

6.2 The Perihelion shift of Mercury

Mercury, like most planets, has an elliptical orbit with a closest point of approach, or perihelion, and a furthest point of recession, the aphelion. Because of perturbations caused by other planets, principally Jupiter, this egg-shaped orbit precesses–as time goes by, the perihelion advances. This means it takes Mercury slightly more than 360 degrees to get back to its closest point of approach.

The advance in perihelion is very slow– on the order of 600 seconds of arc per century. All but 43 seconds of arc can be attributed to perturbations by the other planets. Explaining this discrepancy was a very big problem at the turn of the century. As it turns out, General Relativity can account for these additional 43 seconds. There is some uncertainty, of course–because there are other effects that could contribute, such as the sun's quadrupole moment, which have not been determined as yet. And if these other effects could be determined, it may be that GR would turn out to predict the perihelion advance incorrectly.

To determine the orbits in a Schwarzchild spacetime, it is necessary to calculate the extremal curves:

$$\delta \int ds = 0 \tag{6.1}$$

This variational calculation gives the same result as

$$\delta \int \left[-\left(1 - \frac{2MG}{c^2r}\right)\dot{t}^2 + \left(1 - \frac{2MG}{c^2r}\right)^{-1}\dot{r}^2 + r^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right)\right] ds = 0 \tag{6.2}$$

hence the latter will be used, since it simplifies the calculations. Three of the Euler-Lagrange equations, together with the metric (instead of the Euler-Lagrange equation for r, will be used to determine the orbit. Recall that the Euler-Lagrange equation is given by

$$\frac{d}{ds}\frac{\partial L}{\partial \dot{x}^a} = \frac{\partial L}{\partial x^a}$$

The Euler-Lagrange equations are

$$\frac{d}{ds}\left(r^2 \dot{theta}\right) = r^2 \sin\theta\cos\theta\dot{\phi}^2 \tag{6.3}$$

$$\frac{d}{ds}\left(r^2 \sin^2\theta\dot{\phi}\right) = 0 \tag{6.4}$$

$$\frac{d}{ds}\left[\left(1 - \frac{2MG}{c^2 r}\right)\dot{t}\right] = 0 \tag{6.5}$$

In addition, the metric divided by ds^2 will be used:

$$-\left(1 - \frac{2MG}{c^2 r}\right)c^2\dot{t}^2 + \left(1 - \frac{2MG}{c^2 r}\right)^{-1}\dot{r}^2 + r^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right) = -1 \tag{6.6}$$

Without loss of generality, equation 6.3 may be specialized to the case $\theta = \pi/2$, with initial conditions chosen so that $\dot{\theta} = 0$. This corresponds to choosing the orbital plane. Equation 6.4 can be integrated, giving

$$r^2\dot{\phi} = h \tag{6.7}$$

where h is a constant similar to the areal velocity. Recall, however, that the dot refers to a derivative with respect to the path length parameter, ds, not to coordinate time. Similarly, equation 6.5 leads to

$$\left(1 - \frac{2MG}{c^2 r}\right)\dot{t} = b = constant \tag{6.8}$$

Substitute these expressions into equation 6.6:

$$1 = \left(1 - \frac{2MG}{c^2 r}\right)^{-1} c^2 b^2 - \left(1 - \frac{2MG}{c^2 r}\right)^{-1}\dot{r}^2 - \frac{h^2}{r^2} \tag{6.9}$$

As in standard celestial mechanics, let $r = r(\phi)$, so that

$$r' = \frac{dr}{d\phi} = \frac{dr}{ds}\frac{ds}{d\phi} = \frac{\dot{r}}{\dot{\phi}} \tag{6.10}$$

Solving this for $\dot{r}$ and substituting yields:

$$\left(1 - \frac{2MG}{c^2 r}\right) = c^2 b^2 - \frac{h^2}{r^4}r'^2 - \frac{h^2}{r^2}\left(1 - \frac{2MG}{c^2 r}\right) \tag{6.11}$$

Next, again following standard celestial mechanics, put $r = 1/u$:

$$\left(1 - \frac{2MG}{c^2}u\right) = c^2 b^2 - h^2 u'^2 - h^2 u^2\left(1 - \frac{2MG}{c^2}u\right) \tag{6.12}$$

This equation could be solved for u' and integrated. It's more useful, however, to take a further derivative:

$$2u'u'' = \frac{2MG}{c^2 h^2}u' - 2uu' + \frac{6MG}{c^2}u^2 \tag{6.13}$$

From this equation (or the previous), it is evident that $u' = 0 \rightarrow r = constant$ solutions are possible here as in standard orbital mechanics. The other possibility is:

$$u'' + u = \frac{MG}{c^2h^2} + 3\frac{MG}{c^2}u^2 \tag{6.14}$$

This looks like standard mechanics with two differences. First, there is the additional term on the right hand side. Second, instead of the Areal velocity $H = r^2 d\phi/dt$, there is the similar quantity $h = r^2 d\phi/ds = r^2 d\phi/dtdt/ds$, where t is of course the coordinate time and s is the arclength parameter. Perturbation theory must be used to make further progress. This requires the choice of a small parameter, but since 'small' depends on the units used, the above equation must first be made dimensionless. Since MG/c^2h^2 has dimensions of m^{-1}, a natural choice would be

$$u = \frac{MG}{c^2h^2}x$$

Substituting gives:

$$x'' + x = 1 + \epsilon x^2 \tag{6.15}$$

where

$$\epsilon = 3\frac{M^2G^2}{c^4h^2} \tag{6.16}$$

It is important to verify that ϵ is small. Since the equation will be applied to planetary orbits, especially Mercury, values for h for such bodies must be found.

$$h = r^2\frac{d\phi}{ds} = r^2\frac{d\phi}{dt}\frac{dt}{ds} \approx rv\frac{1}{c}$$

since special relativity is approximately valid for typical planetary orbits, and $dt/ds = (1/c)\gamma$. Inserting all these numbers for Mercury yields about $2.5x10^{-9}$, justifying the use of ϵ as a small parameter. Next, the variable x is expanded in terms of this small parameter:

$$x = x_0 + \epsilon x_1 + \epsilon^2 x_2 + ... \tag{6.17}$$

Here the $x_0, x_1, x_2...$ are all functions associated with the different powers of ϵ in the original equation. Substituting and retaining only first order terms:

$$x_0'' + x_0 + \epsilon\,(x_1'' + x1) = 1 + \epsilon x_0^2 \tag{6.18}$$

Equating the different powers of ϵ gives two equations:

$$x_0'' + x_0 = 1 \tag{6.19}$$

$$x_1'' + x_1 = x_0^2 \tag{6.20}$$

The first equation can be solved easily, and represents the standard solution:

$$x_o = 1 + A\cos(\phi + \delta) \tag{6.21}$$

The phase shift factor can be set equal to zero. Any variance from standard orbits will be caused by solutions to the second equation, which can be written

$$x_1'' + x_1 = 1 + 2A\cos\phi + A^2\cos^2\phi = (\left(1 + \frac{A^2}{2}\right) + 2A\cos\phi + \frac{A^2}{2}\cos 2\phi \quad (6.22)$$

where the identity $cos^2\phi = (1/2)(1+\cos 2\phi)$ was used. This differential equation is easily solved by inspection:

$$x_1 = B\cos(\phi + \delta_2) + \left(1 + \frac{A^2}{2}\right) + A\phi\sin\phi - \frac{A^2}{6}\cos 2\phi \quad (6.23)$$

Hence the full solution can be written

$$x = x_0 + \epsilon x_1 = 1 + A\cos(\phi) + \epsilon\left(B\cos(\phi + \delta_2) + \left(1 + \frac{A^2}{2}\right) - \frac{A^2}{6}\cos 2\phi + A\phi\sin\phi\right) \quad (6.24)$$

The $\cos 2\phi$ and $\cos(\phi + \delta_2)$ will cause a small back-and-forth shifting of the perihelion as time goes on, but nothing that would be noticeable, given the size of ϵ. The last term, $\phi\sin\phi$, however, will get larger and larger as time progresses. Using the identity

$$\cos(\phi - \epsilon\phi) = \cos\phi + \epsilon\phi\sin\phi$$

The solution becomes

$$x = 1 + a\cos(\phi - \epsilon\phi) + \text{small periodic terms} \quad (6.25)$$

This corresponds to $1/r$, so maximums in x will correspond to minimums (perihelia) in r. These perihelia will occur when

$$\phi(1 - \epsilon) = 2\phi n \quad (6.26)$$

Hence ϕ is given approximately by

$$\phi = \frac{2\pi n}{1 - \epsilon} \approx 2\pi n(1 + \epsilon) \quad (6.27)$$

The change in the angle from one orbit to the next is:

$$\Delta\phi = 2\pi(1 + \epsilon) \quad (6.28)$$

so the shift per orbit is given by

$$\delta\phi = 2\pi\epsilon = \frac{6\pi M^2 G^2}{c^4 h^2} = \frac{6\pi M^2 G^2}{c^4 r^4 (d\phi/dt)^2 (dt/ds)^2} \approx \frac{6\pi M^2 G^2}{c^2 r^4 v^2} \quad (6.29)$$

For Mercury, this result is 43.03 seconds of arc per century, in excellent agreement with the observed discrepancy.

6.3 Bending of light by the sun.

One of the original predictions made by Einstein concerned the bending of light by the sun. Light is massless, and should therefore follow a null geodesic in spacetime. Starlight passing near the sun should be bent toward the sun, so that it appears shifted in position away from the sun when observed by astronomers on Earth. An eclipse expedition led by Sir Arthur Eddington verified this prediction soon after the Great War. Einstein, understandably, was elated at the result.

Light travels on null geodesics, so given a curve parameter λ, the Lagrangian is given by

$$\delta \int \left[-\left(1 - \frac{2MG}{c^2 r}\right) \dot{t}^2 + \left(1 - \frac{2MG}{c^2 r}\right)^{-1} \dot{r}^2 + r^2 \left(\dot{\theta}^2 + \sin^2\theta \dot{\phi}^2\right) \right] d\lambda = 0 \tag{6.30}$$

The parameter is λ rather than the arclength, ds, because for null geodesics, no proper time passes. In addition to the Lagrange equations for ϕ and t, the metric relationship can be used (instead of the Lagrange equation for r):

$$-\left(1 - \frac{2MG}{c^2 r}\right) c^2 \dot{t}^2 + \left(1 - \frac{2MG}{c^2 r}\right)^{-1} \dot{r}^2 + r^2 \left(\dot{\theta}^2 + \sin^2\theta \dot{\phi}^2\right) = 0 \tag{6.31}$$

where a dot is a derivative with respect to the parameter λ.

Following the same steps as for the perihelion shift calculation, the following equation for $u = 1/r$ can be derived:

$$u'' + u = 3\frac{MG}{c^2} u^2 \tag{6.32}$$

where the prime denotes a derivative with respect to *phi*. As usual, it is necessary to rescale the variable u so as to make it dimensionless, and then pick out a natural small quantity with which to conduct the perturbation calculation. The natural choice $v = (MG/c^2)u$ doesn't work out well, as can be verified. Instead, use

$$v = r_0 u \tag{6.33}$$

where r_0 is the light ray's closest point of approach to the sun. It's easy to see, that with the sun's radius on order of a billion meters, that this quantity will be small. Substituting, the equation becomes

$$v'' + v = \epsilon v^2 \tag{6.34}$$

where ϵ is given by

$$\epsilon = \frac{3MG}{c^2 r_0} \tag{6.35}$$

Carrying out the perturbation expansion as before and solving results in

$$u = \frac{1}{r_0}\sin\phi + \frac{3MG}{2c^2r_0^2}\left(1+\frac{1}{3}\cos 2\phi\right) \tag{6.36}$$

The first term on the right is a straight line in polar coordinates. Without loss of generality, this may be taken to be the straight line parallel to the x-axis and passing through $y = r_0$. As $r \to \infty$, $u \to 0$, so set $u = 0$ and solve for ϕ, which is the asymptotic angle through which the light is bent as it goes to infinity, and one-half the total angle bent through (since the ray came from negative infinity, during which it bent by an identical amount).

$$\frac{\sin\phi}{r_0} \approx \frac{\phi}{r_0} \approx -\frac{3MG}{2c^2r_0^2}\left(1+\frac{1}{3}\right)\phi = -\frac{2MG}{c^2r_0} \tag{6.37}$$

The minus sign means the light bends toward the sun. The total angle the light is bent through, Δ, is therefore given by

$$\Delta = \frac{4MG}{c^2r_0} \tag{6.38}$$

. This quantity corresponds to 75 seconds of arc, and was verified by the famous eclipse expedition to Africa led by Sir Arthur Eddington in about 1919.

6.4 The reflection of radar by Venus

A more modern test of general relativity involves measuring how long it takes radar to bounce off Venus and return to Earth. Einstein's theory predicts a delay in the return of the wave, and this turns out to be the case. Since radar is a frequency of light, it will travel on null geodesics. A ray of light traveling between Earth and Venus will travel approximately on a straight line. The path may be taken as parallel to the x-axis in the x-y plane. In spherical coordinates, the line will therefore be given by $r\sin\phi = r_o$ and $\theta = \pi/2$

The line element for an infinitesimal displacement on this line is

$$0 = -\left(1-\frac{2MG}{c^2r}\right)c^2dt^2 - \frac{dr^2}{\left(1-\frac{2MG}{c^2r}\right)} + r^2d\phi^2 \tag{6.39}$$

From the equation of the path of the ray, obtain

$$d\phi = \left(\frac{d}{dr}\arcsin(r_o/r)\right)dr = \frac{dr}{\left(1-{r_o}^2/r^2\right)^{1/2}}$$

This can be used to eliminate $d\phi$ from the previous equation, resulting in

$$\begin{aligned} c^2dt^2 &= \tfrac{dr^2}{(1-2MG/c^2r)^2} + \tfrac{r_o^2dr^2}{(1-2MG/c^2r)(r^2-r_o^2)} = \\ &= \tfrac{dr^2\left(1-2mr_o^2/r^3\right)}{(1-r_o^2/r^2)(1-2MG/c^2r)^2} \end{aligned} \tag{6.40}$$

Taking the square root and expanding to first order in MG/c^2 yields

$$cdt = \frac{dr}{(1 - r_o^2/r^2)^{1/2}} \left(1 + \frac{2MG}{c^2 r} - \frac{MGr_o^2}{r^2}\right) \tag{6.41}$$

Despite appearances, this can be readily integrated:

$$ct = \left(\sqrt{r_p^2 - r_o^2} + \sqrt{r_e^2 - r_o^2}\right) + \frac{2MG}{c^2} log \left(\frac{\left(\sqrt{r_p^2 - r_o^2} + r_p\right)\left(\sqrt{r_e^2 - r_o^2} + r_e\right)}{ro^2}\right) - \\ - \frac{MG}{c^2} \left(\frac{\sqrt{r_p^2 - r_o^2} + r_e}{r_p} + \frac{\sqrt{r_e^2 - r_o^2}}{r_e}\right) \tag{6.42}$$

The first term on the right represents the usual flat space result, while the next two terms are additional contributions from GR, increasing the effective length. This extra distance means that signals bounced off Venus will be delayed, arriving later than expected, approximately $200\mu s$ at superior conjunction.

6.5 Exercises

1. Complete the perturbation calculations for the bending of light by the sun.

Chapter 7

Some Basic Exact Solutions

Einstein, when he first came up with general relativity, thought there would never be any exact solution to the theory. Today there are hundreds of solutions, and more are found every year. Nevertheless, very few of these exact solutions might be considered important or basic. Solutions that yield immediate insight into how the universe actually works. Aside from the Schwarzschild exterior and interior solutions, there are the Friedmann and Robertson-Walker solutions for cosmology, which are taken up in a later chapter. In this chapter, solutions describing a radiating body, an electrostatically charged body, and a rotating body will be taken up. The latter solution, called the Kerr solution, is no doubt the most important exact solution in this chapter, as it describes a rotating black hole.

7.1 De Sitter Space

7.2 The Reissner-Nordstrom Solution

An exact solution for the Einstein-Maxwell system for an idealized charged point particle yields another important exact solution of interest, called the Reissner-Nordstrom solution. The equation for a particle exhibiting a spin-1 long-range field in flat space is Maxwell's equation:

$$\partial_b F^{ab} = 4\pi j^a \tag{7.1}$$

where

$$F_{ab} = \nabla_a A_b - \nabla_b A_a \tag{7.2}$$

In flat space, various gauges can be chosen, each of which yields a different field equation and Lagrangian. Only the "gauge-free" equation has an exact solution, as far as anyone knows. The Lagrangian shall be therefore chosen to be:

$$\pounds = \sqrt{-g} F_{ab} F^{ab} \tag{7.3}$$

In solving the coupled Einstein-Maxwell system, a suitable stress-energy tensor is needed. This can be obtained by varying the Lagrangian density with respect to the metric:

$$T_{ab} = -\frac{\alpha_M}{8\pi}\frac{1}{\sqrt{-g}}\frac{\delta}{\delta g^{ab}}\sqrt{-g}\pounds \tag{7.4}$$

$$T_{ab} = \frac{\alpha_M}{16\pi}\left(F_a^d F_{bd} - \frac{1}{4}g_{ab}F_{cd}F^{cd}\right) \tag{7.5}$$

This stress energy is traceless, hence Einstein's equations read

$$R_{ab} = \kappa\left(T_{ab} - \frac{1}{2}g_{ab}T\right) = \kappa T_{ab} \tag{7.6}$$

The metric for static spherical symmetry can be taken to have the form

$$ds^2 = e^\nu dt^2 - e^\lambda dr^2 - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) \tag{7.7}$$

It is advantageous to recast Maxwell's equation in terms of ordinary partial derivatives:

$$\frac{1}{\sqrt{-g}}\partial_a\left(\sqrt{-g}F^{ab}\right) = 0 \tag{7.8}$$

This is possible only because F^{ab} is a second rank antisymmetric tensor, which is not hard to show. First, the last term in the covariant derivative vanishes because Γ^a_{bc} is symmetric in its lower indices while F^{bc} antisymmetric:

$$\nabla_a F^{ab} = \frac{\partial F^{ab}}{\partial x^a} + \Gamma^a_{ac}F^{cb} + \Gamma^b_{ac}F^{ac} = \frac{\partial F^{ab}}{\partial x^a} + \Gamma^a_{ac}F^{cb} \tag{7.9}$$

Second, it can be shown that the other Christoffel symbol can be written as

$$\Gamma^a_{ac} = \frac{\partial}{\partial x^c}log\sqrt{-g} = \frac{1}{2g}\partial_c g \tag{7.10}$$

where g is the determinant of the matrix of the metric tensor components. Inserting this expression and doing some algebra results in equation 8.14. Equations 8.13 and 8.14 shall now be solved. We search for a solution where F_{ab} is of the form

$$F_{ab} = \begin{pmatrix} 0 & -A_0' & 0 & 0 \\ A_0' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{7.11}$$

With this choice, the stress-energy tensor becomes

$$T_{ab} = \frac{\alpha_M}{32\pi}A_0'^2\begin{pmatrix} -e^{-\lambda} & 0 & 0 & 0 \\ 0 & e^{-\nu} & 0 & 0 \\ 0 & 0 & r^2e^{-\lambda-\nu} & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta e^{-\lambda-\nu} \end{pmatrix} \tag{7.12}$$

Einstein's equations then can be written down as

$$R_{00} = e^{\nu-\lambda}\left(\frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} + \frac{\nu'}{r}\right) = \frac{1}{4}\frac{G\alpha_M}{c^4}A_0'^2 e^{-\lambda} \tag{7.13}$$

$$R_{11} = \left(-\frac{\nu''}{2} + \frac{\nu'\lambda'}{4} - \frac{\nu'^2}{4} + \frac{\lambda'}{r}\right) = -\frac{1}{4}\frac{G\alpha_M}{c^4}A_0'^2 e^{-\nu} \tag{7.14}$$

$$R_{22} = e^{-\lambda}\left(-1 - \frac{r\nu'}{2} + \frac{r\lambda'}{2}\right) + k = \frac{1}{4}\frac{G\alpha_M}{c^4}A_0'^2 e^{-\lambda-\nu} r^2 \tag{7.15}$$

In the last equation, k=1 for spherical symmetry, k=0 for plane symmetry, and k=-1 one for hyperbolic symmetry. Multiplying equation 8.18 by $e^{-\nu+\lambda}$ and adding to equation 8.19yields

$$\nu' + \lambda' = 0 \rightarrow \nu = -\lambda \tag{7.16}$$

Using this result, it is now possible to simplify equation 8.14 to

$$\frac{\partial}{\partial r}\left(r^2 \sin\theta F^{01}\right) = 0 \tag{7.17}$$

which further reduces to

$$\frac{\partial^2 A_0}{\partial r^2} + \frac{2}{r}\frac{\partial A_0}{\partial r} = 0 \tag{7.18}$$

The solution, of course, is

$$A_0 = \frac{a}{r} + b \tag{7.19}$$

Inserting this solution into equation 8.20and using the fact that $\lambda = -\nu$ results in

$$1 + e^{\nu}\left(-1 - r\nu'\right) = 1 - \left(re^{\nu}\right)' = \frac{a^2 G\alpha_M}{4c^4}\frac{1}{r^2} \tag{7.20}$$

This equation can be immediately integrated to yield

$$e^{\nu} = 1 - \frac{2MG}{c^2 r} + \frac{a^2 G\alpha_M}{4c^4}\frac{1}{r^2} \tag{7.21}$$

A cosmological constant term can be added in, which results in an additional term. The constant a can be taken as proportional to the charge. Similar solutions can be found for plane and hyperbolic symmetry. One interesting feature of the Reissner-Nordstrom metric is that the force can be repulsive. It's unlikely, however, that enough charge could be accumulated to observe such an effect, since the large charges necessary would arc to ground.

7.3 The Vaidya Radiating Metric

7.4 Dust Solutions

7.5 The Tolman Metric

7.6 The Kerr Solution

In 1967 Kerr, as a graduate student, came upon his exact solution. At first its physical meaning wasn't clear, but soon it was found to describe a rotating black hole.

7.7 Projects

7.7.1 Interior Electrostatic Solution

In this project, the idea is to develop interior solutions that correspond to the Reissner-Nordstrom exterior. On the inside of a body, which might be subatomic in size, there is a charge density and corresponding electric field, held together by gravitation and possibly by other forces, which could be modeled by negative pressures (tensions). The stress energy is given by a sum of the stellar structure stress-energy plus the Maxwell stress-energy, supplemented by Maxwell's equation in the presence of charged material. The basic equations are (with tons of luck):

$$R_{00} = e^{\nu-\lambda}\left(\frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} + \frac{\nu'}{r}\right) = \Lambda e^{\nu} + \frac{4\pi G}{c^2}\left(\rho + 3\frac{P}{c^2}\right)e^{\nu} + + \frac{4\pi G}{c^4}\epsilon_0 {A_0'}^2 e^{-\lambda} \tag{7.22}$$

$$R_{11} = \left(-\frac{\nu''}{2} + \frac{\nu'\lambda'}{4} - \frac{\nu'^2}{4} + \frac{\lambda'}{r}\right) = -\Lambda e^{\lambda} + \frac{4\pi G}{c^2}\left(\rho - \frac{P}{c^2}\right)e^{\lambda} - - \frac{4\pi G}{c^4}\epsilon_0 {A_0'}^2 e^{-\nu} \tag{7.23}$$

$$R_{22} = e^{-\lambda}\left(-1 - \frac{r\nu'}{2} + \frac{r\lambda'}{2}\right) + 1 = -\Lambda r^2 + \frac{4\pi G}{c^2}\left(\rho - \frac{P}{c^2}\right)r^2 + + \frac{4\pi G}{c^4}\epsilon_0 {A_0'}^2 r^2 e^{-\lambda-\nu} \tag{7.24}$$

$$\frac{\partial}{\partial r}\left(e^{-\nu/2-\lambda/2} r^2 \frac{\partial A_0}{\partial r}\right) = e^{\lambda/2} r^2 \frac{\rho_e}{\epsilon_0} \tag{7.25}$$

Adding $e^{\lambda-\nu}R_{00} + R_{11}$ gives the usual

$$\frac{8\pi G}{c^2}\left(\rho + \frac{P}{c^2}\right)e^{\lambda} = \frac{\nu' + \lambda'}{r} \tag{7.26}$$

A cheap shot idea here would be to look at charged textures, which are extended defects thought to result during the big bang. These structures are thought to have negative pressures, which are interpreted as tensions. Assume, therefore, that

$$P = -\rho c^2 \tag{7.27}$$

Then $\lambda' + \nu' = 0$. It can then be shown that the following equations hold:

$$e^{-\lambda} = 1 - \frac{2m}{r} + \frac{1}{3}\Lambda r^2 - \frac{2\epsilon}{r} \tag{7.28}$$

$$m' = \frac{G}{c^2}4\pi\rho r^2 \tag{7.29}$$

$$\epsilon' = \frac{G}{c^4}4\pi\left(\frac{1}{2}\epsilon_0 {A_0'}^2\right)r^2 \tag{7.30}$$

$$q = 4\pi\int_0^r e^{\lambda/2}r^2\rho_e dr = 4\pi r^2\epsilon_0 {A_0}' \tag{7.31}$$

Believe it or not, these equations are all that's left of Einstein-Maxwell. Numerically integrate out from the center, assuming $\Lambda = 0$, say, and $e^{-\lambda} = 1$ at the core. As an alternative, following precisely the steps for the derivation of the TOV equation, and making use of a couple of similar definitions, the following equations can be derived:

$$\frac{8\pi G}{c^2}\rho = e^{-\lambda}\left(\frac{\lambda'}{r} - \frac{1}{r^2}\right) + \frac{1}{r^2} + \Lambda - \frac{4\pi G}{c^4}\epsilon_0 {A_0'}^2 e^{-\lambda-\nu} \tag{7.32}$$

$$\frac{8\pi G}{c^4}P = e^{-\lambda}\left(\frac{\nu'}{r} + \frac{1}{r^2}\right) - \frac{1}{r^2} - \Lambda + \frac{4\pi G}{c^4}\epsilon_0 {A_0'}^2 e^{-\lambda-\nu} \tag{7.33}$$

$$\frac{e^{\lambda}}{r^2} = \left(\frac{1}{r^2} - \frac{\nu'^2}{4} + \frac{\nu'\lambda'}{4} + \frac{\nu' + \lambda'}{2r} - \frac{\nu''}{2}\right) + \frac{8\pi G}{c^4}\epsilon_0 {A_0'}^2 e^{-\nu} \tag{7.34}$$

Painstakingly following the previously developed technique, the TOV equivalent can be derived:

$$P' = -\frac{\nu'}{2}\left(\rho c^2 + P\right) + \frac{2\epsilon_0 {A_0'}^2 e^{-\nu-\lambda}}{r} + \left(\frac{1}{2}\epsilon_0 {A_0'}^2 e^{-\nu-\lambda}\right)' \tag{7.35}$$

with (cross fingers)

$$\nu' = \frac{\frac{8\pi G}{c^4}Pr^3 + 2m + 2\epsilon + \frac{2}{3}\Lambda r^3}{r\left(r - 2m - 2\epsilon + \frac{1}{3}\Lambda r^3\right)} - \frac{4\pi G}{c^4}\epsilon_0 {A_0'}^2 e^{-\nu} r \tag{7.36}$$

Couple all this with the Maxwell equation and an equation of state, and you've got a system of equations.

$$q = 4\pi \int_0^r e^{\lambda/2} r^2 \rho_e dr = 4\pi\epsilon_0 r^2 \frac{\partial A_0}{\partial r} e^{-\nu/2-\lambda/2} \tag{7.37}$$

$$P = P(\rho) \tag{7.38}$$

The charge density and mass density should probably be proportional. Exact solutions may be possible, as well. Doubtless the basic equations, here, could use some more grinding. Again, you can start in the interior and work your way out. Note that the definitions for the previous texture case hold here, as well.

7.7.2 New and probably useless solutions to Einstein's equations

The intrepid student could take other combinations of the basic anti-symmetric tensors divised by Vuille and attempt to derive additional exact solutions in the spherically-symmetric case. I can assure you these solutions exist, and I've already found some, but there is no reason to think they are of value. One never knows. See the handout on the quark solution.

7.7.3 Homework 6

1. Calculate the perijovian shift of Io. Would this be observable?

2. Fill out the perturbation calculations for the bending of light by the sun.

3. Suppose you observed that the 434 nm line of hydrogen in light from a star was shifted to 520 nm. If, through analysis of an orbiting companion, you determined the mass of the star to be 100 solar masses, compute the radius of the star using gravitational red shift.

4. Suppose the action integral for a theory of gravity is given by

$$I = \int (\eta_{ab} - h_{ab}) \frac{dx^a}{d\lambda} \frac{dx^b}{d\lambda} d\lambda$$

where η_{ab} is the usual Minkowski metric and h_{ab} is given by

$$\frac{2MG}{c^2 r} \begin{pmatrix} c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Calculate the perihelion shift of Mercury in this theory, and the bending of light by the sun.

Chapter 8

Approximation Methods

8.1 Metric Perturbations

In many instances Einstein's equation is intractable, and it is necessary to resort to perturbation theory, often in tandem with numerical techniques. This process is fraught with peril, of course, since in General Relativity a spacetime is an equivalence class, under diffeomorphism, of an infinite number of different spacetimes. There is no guarantee, however, that all perturbations will yield the same result if every case. Similar problems arise in numerical solutions. Guidance must be sought in physical reasonableness and special cases.

The basic perturbation method is straightforward. Let g^0_{ab} be a known exact solution to Einstein's equations that is thought to be close to true metric, g_{ab}, of the spacetime. Then the true metric might be rewritten as

$$g_{ab} = g^0_{ab} + \lambda g^1_{ab} + \lambda^2 g^2_{ab} + ... \tag{8.1}$$

Most of the time only the first order term is retained, and lambda is absorbed into g^1_{ab}. Using this truncated series, the Christoffel symbols and curvature terms may then be calculated, with terms quadratic in g^1_{ab} subsequently dropped while terms of zero and first order terms separated out. Generally speaking, it's still a very tough calculation. An example might be a quantum field that is thought to be close to a Schwarzschild solution.

The most common use of metric perturbation theory is the derivation of the so-called Linearized Equations. In this technique, the metric is approximated by

$$g_{ab} = \eta_{ab} + h_{ab} \tag{8.2}$$

where h_{ab} is considered small compared to the Minkowski metric, η_{ab}. Raising and lowering of indices will be effected with η. The inverse of the metric, to first order in h_{ab}, is

$$g^{ab} = \eta^{ab} - h^{ab} \tag{8.3}$$

Next, the Christoffel symbols must be calculated. Because of the raising and lowering convention, all these symbols will be linear in h_{ab}:

$$\Gamma^a_{bc} = \frac{1}{2}\eta^{ad}\left(h_{bd,c} + h_{cd,b} - h_{bc,d}\right) \tag{8.4}$$

where the comma in, for example, $h_{bc,d}$ indicates a partial derivative of h_{bc} with respect to x^d. It is straightforward, then, to calculate the Riemann tensor, which to first order is:

$$R_{abcd} = \frac{1}{2}\left(h_{ad,bc} + h_{bc,ad} - h_{ac,bd} + h_{bd,ac}\right) \tag{8.5}$$

The Einstein tensor, again to first order, may now be written as

$$G_{ab} = R_{ab} - \frac{1}{2}Rg_{ab} \tag{8.6}$$

where the various contractions have, of course, been effected with η_{ab}. This expression is complicated, but can be simplified with a gauge transformation.

8.1.1 The Einstein-Proca System in Perturbation

In this section, perturbation technique will be illustrated by applying it to the Einstein-Proca system. The calculation will be developed in three different ways, with details left for the reader. In the first method, the equations shall be set up, variables will be made dimensionless, and the resulting equations solved. The second approach will involve a perturbation of Minkowski space, while the third method will use a perturbation of the Schwarzschild exterior solution.

An exact solution for the Einstein-Proca system for an idealized point particle has yet to be found [1], [2]. Such systems have been occasionally discussed in the literature, for example in Dereli et al. [3], and have been invoked by Tucker and Wang [4] in connection with dark matter gravitational interactions, where it was shown that such fields could explain in part the galactic rotation curves. Numerical solutions were found independently by Obukov and Vlachynsky [5]and Toussaint [6]. These latter two papers demonstrated the existence of naked singularities in this system. In this section, the system will be solved up to a final integral, which will then be subjected to perturbation analysis.

Consider a force modeled as a Proca interaction. During gravitational collapse, the equivalent of the force charge, referred to here as the Proca charge, would not be cancelled by an accumulation of opposite charges, as in electromagnetic interactions. The stress energy of the force field would therefore be expected to make contributions to the gravitational field of the spacetime surrounding the collapsed object. Because both the force and the associated gravitational field fall off exponentially, the effect on the spacetime surrounding a stellar-size black hole would be completely negligible.

On the other hand, it is thought that microscopic black holes may have been created in vast numbers during the Big Bang. These micro black holes would be expected to have a variety of different sizes, including, conceivably, some on the order of a femtometer across. For such objects, there is the possibility that associated fields of Proca-type would prevent the formation of event horizons, leaving a (short-lived) naked singularity. This, then, might be considerred a counter-example to Penrose's cosmic censorship conjecture.

The equation for a particle exhibiting a spin-1 short or intermediate-range field in flat space is Proca's equation [7], which in the absence of currents is

$$\partial_a F^{ab} + \mu^2 A^b = 0 \tag{8.7}$$

where

$$F_{ab} = \nabla_a A_b - \nabla_b A_a \tag{8.8}$$

The metric will be taken to have diagonal form $c^2, -1, -1, -1$. The quantity μ is a constant, interpreted as being proportional to the mass of the field quanta and inversely proportional to the range of the interaction.

Traditionally, the form of equation 8.7 was chosen for several good reasons. First and foremost, it gives an intuitively correct answer, which is a potential that rapidly falls off as r gets large. Second, it can be realized by adding a linear term to Maxwell's equations. Third, the equation is covariant, and finally, a Lagrangian exists, meaning this equation is extremal in a more general function space.

The Lagrangian density for the classic Proca system is:

$$\pounds = \sqrt{-g}\left(\alpha F_{ab} F^{ab} + \beta A_a A^a\right) \tag{8.9}$$

where g is the determinant of the metric, and α and β are constants. Varying this equation with respect to A^c returns equation 8.8, provided that $\beta/2\alpha = -\mu^2$. It turns out that the last term on the right in 8.9 , which distinguishes the standard Proca from Maxwell, causes considerable difficulties in finding the solution to the general relativistic problem. These difficulties are absent in the Reissner-Nordstrom problem primarily due to the antisymmetry of F_{ab}. Nonetheless, considerable progress can be made, as will be demonstrated in the next section.

Derivation and Solution of the Field Equations

The metric for static spherical symmetry can be taken to have the form

$$ds^2 = e^{\nu} dt^2 - e^{\lambda} dr^2 - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) \tag{8.10}$$

Similar forms can also be written down for plane and hyperbolic symmetry: all subsequent steps in this paper could equally well be taken in those two cases . The Proca stress-energy tensor can be obtained from

$$T_{ab} = -\frac{\alpha_M}{8\pi}\frac{1}{\sqrt{-g}}\frac{\delta}{\delta g^{ab}}\sqrt{-g}\pounds \tag{8.11}$$

For a given field, the constant α_M is a parameter that tells how strongly the stress-energy of the field creates gravitation. This gravitational strength is so weak compared to the other forces that it is impractical to determine experimentally. Again for convenience, this constant and the factor of 8π shall be rolled into the constants α and β. Applying this formula to equation 8.9 results in

$$T_{ab} = 2\alpha F_a{}^d F_{bd} + \beta A_a A_b - \frac{1}{2} g_{ab} \left(\alpha F_{cd} F^{cd} + \beta A_c A^c \right) \tag{8.12}$$

The Proca stress energy, unlike the Maxwell stress-energy, is not traceless. Einstein's equations read

$$R_{ab} = \kappa \left(T_{ab} - \frac{1}{2} g_{ab} T \right) \tag{8.13}$$

It is advantageous to recast the Proca equation in terms of ordinary partial derivatives:

$$\frac{1}{\sqrt{-g}} \partial_a \left(\sqrt{-g} F^{ab} \right) - \frac{\beta}{2\alpha} A^b = 0 \tag{8.14}$$

The Proca system corresponds to a choice of

$$\frac{\beta}{2\alpha} = -\mu^2 \tag{8.15}$$

We search for a solution of equations 8.9-8.14 where F_{ab} is of the form

$$F_{ab} = \begin{pmatrix} 0 & -A_0' & 0 & 0 \\ A_0' & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{8.16}$$

With these choices, the stress-energy tensor becomes

$$T_{ab} = \alpha A_0'^2 \begin{pmatrix} -e^{-\lambda} & 0 & 0 & 0 \\ 0 & e^{-\nu} & 0 & 0 \\ 0 & 0 & -r^2 e^{-\lambda-\nu} & 0 \\ 0 & 0 & 0 & -r^2 \sin^2\theta e^{-\lambda-\nu} \end{pmatrix} +$$

$$+ \frac{\beta A_0^2}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{\lambda-\nu} & 0 & 0 \\ 0 & 0 & r^2 e^{-\nu} & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta e^{-\nu} \end{pmatrix} \tag{8.17}$$

Einstein's equation then can be written down as

$$R_{00} = e^{\nu-\lambda} \left(\frac{\nu''}{2} - \frac{\nu'\lambda'}{4} + \frac{\nu'^2}{4} + \frac{\nu'}{r} \right) = -\kappa\alpha {A_0'}^2 e^{-\lambda} + \kappa\beta {A_0}^2 \tag{8.18}$$

$$R_{11} = \left(-\frac{\nu''}{2} + \frac{\nu'\lambda'}{4} - \frac{\nu'^2}{4} + \frac{\lambda'}{r} \right) = \kappa\alpha {A_0'}^2 e^{-\nu} \tag{8.19}$$

$$R_{22} = 1 + e^{-\lambda} \left(-1 - \frac{r\nu'}{2} + \frac{r\lambda'}{2} \right) = -\kappa\alpha r^2 {A_0'}^2 e^{-\lambda-\nu} \tag{8.20}$$

Of course, $R_{33} = R_{22}\sin^2\theta$. Finally, the equation 8.7for the massive vector field is given by

$$A_0'' + \frac{2}{r}A_0' - \left(\frac{\lambda'}{2} + \frac{\nu'}{2}\right)A_0' + \frac{\beta}{2\alpha}e^{\lambda}A_0 = 0 \tag{8.21}$$

On the face of it, these equations are not dissimilar to Einstein-Maxwell, differing only by the inclusion of two rather innocuous terms. In fact, these small changes result in a tremendous complications, as will soon be seen. In the first place, unlike Einstein- Maxwell, the enormous simplification of $\lambda' + \nu' = 0$ does not occur. Indeed, multiplying equation 8.18 by $e^{-\nu+\lambda}$ and adding to equation 8.19 yields

$$\frac{\nu'}{r} + \frac{\lambda'}{r} = \kappa\beta {A_0}^2 e^{\lambda-\nu} \tag{8.22}$$

Solving this equation for λ' and substituting into equation 8.20 results, after some algebra, in:

$$e^{\lambda} = \frac{1 + r\nu' - \kappa\alpha r^2 {A_0'}^2 e^{-\nu}}{1 + \frac{1}{2}\kappa\beta r^2 {A_0}^2 e^{-\nu}} \tag{8.23}$$

So the function e^{λ} has been solved in terms of the other two functions. This result, when substituted into the 00 and 11 equations, makes them identical. Using the last two equations, the remaining equations for ν and A_0 can be written as:

$$\nu'' + \nu'^2 + \frac{2\nu'}{r} = -2\kappa\alpha {A_0'}^2 e^{-\nu} + \left(2 + \frac{r\nu'}{2}\right)\kappa\beta {A_0}^2 e^{-\nu}\frac{1 + r\nu' - \kappa\alpha r^2 {A_0'}^2 e^{-\nu}}{1 + \frac{1}{2}\kappa\beta r^2 {A_0}^2 e^{-\nu}} \tag{8.24}$$

$$A_0'' + \frac{2}{r}A_0' = \frac{\beta}{2\alpha}A_0\left(-1 + \alpha\kappa r A_0 A_0' e^{-\nu}\right)\frac{1 + r\nu' - \kappa\alpha r^2 {A_0'}^2 e^{-\nu}}{1 + \frac{1}{2}\kappa\beta r^2 {A_0}^2 e^{-\nu}} \tag{8.25}$$

The equation for ν can be significantly simplified by the substitution

$$e^{\nu} = f \tag{8.26}$$

where $f = f(r)$. Substituting this into equation 8.24 results in

$$f'' + \frac{2}{r}f' = -2\kappa\alpha {A_0'}^2 + \kappa\beta {A_0}^2\left(2 + \frac{rf'}{2f}\right)\left[\frac{f + rf' - \kappa\alpha r^2 {A_0'}^2}{f + \frac{1}{2}\kappa\beta r^2 {A_0}^2}\right] \tag{8.27}$$

Similiarly, in equation 8.25:

$$A_0'' + \frac{2}{r}A_0' = \frac{\beta}{2\alpha}A_0\left(-1 + \frac{\alpha\kappa r A_0 A_0'}{f}\right)\left[\frac{f + rf' - \kappa\alpha r^2 {A_0'}^2}{f + \frac{1}{2}\kappa\beta r^2 {A_0}^2}\right] \tag{8.28}$$

It may be there is an exact solution for these two equations, however finding it would be a matter of experimentation and luck, given the cubic nonlinearities. A perturbative approach, on the other hand, has good chances of success, and can be quite accurate for reasonable values of the parameters of the theory. The procedure involves redefining all quantities so that they are dimensionless, using naturally-occurring parameters.

First, to get the Proca, it is necessary to define α and β. Let these be

$$\alpha = -\frac{1}{2}\epsilon_0 \tag{8.29}$$

and

$$\beta = \mu^2 \epsilon_0 \tag{8.30}$$

The quantity ϵ_0 fulfills the same function as the permittivity of free space in electromagnetism, but in this context pertains to the Proca interaction. μ is, of course, the standard range parameter. Next, set

$$x = \mu r \tag{8.31}$$

This redefines the r-coordinate in terms of a dimensionless parameter. The metric function f is already dimensionless; however A_0 has dimensions of Joules per Proca charge. Denote the Proca charge by q, in analogy with electromagnetism. Next, set

$$A = su \tag{8.32}$$

where

$$s = \epsilon_0^{-1} q\mu \tag{8.33}$$

The parameter s carries all the units of A. Substitute all these into the above equations and obtain the following two equations in terms of dimensionless variables only:

$$\left(u'' + \frac{2}{x}u'\right)\left(f + \frac{1}{2}\epsilon x^2 u^2\right) f = u\left(f + \frac{1}{2}\epsilon u u'\right)\left(f + xf' + \frac{1}{2}\epsilon x^2 u'^2\right) \tag{8.34}$$

$$\left(f'' + \frac{2}{x}f' - \epsilon u'^2\right)\left(f + \frac{1}{2}\epsilon x^2 u^2\right) f = \epsilon u^2 \left(2f + \frac{1}{2}xf'\right)\left(f + xf' + \frac{1}{2}\epsilon x^2 u'^2\right) \tag{8.35}$$

where

$$\epsilon = \frac{\kappa q^2 \mu^2}{\epsilon_0} \tag{8.36}$$

is a small, dimensionless perturbation parameter, with $\kappa = G/c^4$. For a scale similar to that of the strong force, the factor μ^2 is quite large, $\approx 10^{30}$, and $\kappa \approx 10^{-44}$. The remaining term, q^2/ϵ_0, is analogous to electromagnetic quantities, where the term would have magnitude of about 10^{-27}. Since the strong force is about 100 times stronger than the electromagnetic force, it follows that this combination of terms should be around 10^{-25} in the case under consideration. It appears therefore well justified to consider ϵ a small quantity for a wide range of scale. The functions u and f may therefore be expanded:

$$f = f_0 + \epsilon f_1 + \epsilon^2 f_2 + .. \tag{8.37}$$

$$u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + ... \tag{8.38}$$

Inserting these expressions, the following zeroth order equations are obtained:

$$\left({f_0}'' + \frac{2}{x} {f_0}' \right) {f_0}^2 = 0 \tag{8.39}$$

$$\left({u_0}'' + \frac{2}{x} {u_0}' \right) {f_0}^2 = u_0 f_0 \left(f_0 + x {f_0}' \right) \tag{8.40}$$

Equation 8.39 has the solution

$$f_0 = a + \frac{b}{x} \tag{8.41}$$

The second term on the right will be the usual Schwarzschild term, but will evidently be small, and more appropriately first order. Hence b will be taken to be zero, with $a = 1$, giving Minkowski space as the lowest order in the metric. With this choice, equation 8.40 has the usual flat space solution, which is

$$u_0 = c_0 \frac{e^{-x}}{x} + c_1 \frac{e^{x}}{x} \tag{8.42}$$

It is evident that $c_1 = 0$ in this case. The first order equations may be written:

$$\left({u_1}'' + \frac{2}{x} {u_1}' \right) {f_0}^2 + \left(u_0'' + \frac{2}{x} u_0' \right) f_0 \left(2 f_1 + \frac{1}{2} x^2 u_0^2 \right) =$$

$$= u_0 f_0 \left(f_1 + x {f_1}' + \frac{1}{2} x^2 {u_0'}^2 \right) + \left(f_0 + x {f_0}' \right) \left(f_0 u_1 + u_0 f_1 + \frac{1}{2} u_0^2 u_0' \right) \tag{8.43}$$

$$\left({f_1}'' + \frac{2}{x} {f_1}' - {u_0'}^2 \right) {f_0}^2 + \left({f_0}'' + \frac{2}{x} {f_0}' \right) \left(2 f_0 f_1 + \frac{1}{2} x^2 u_0^2 \right) =$$

$$= {u_0}^2 \left(2 f_0 + \frac{x}{2} {f_0}' \right) \left(f_0 + x {f_0}' \right) \tag{8.44}$$

The focus here is on equation 8.44, which yields the first-order correction to the metric. The homogeneous solution is again given by equation 8.41, except this time the constant solution will be discarded and the b/x term retained. This can be identified with the standard Schwarzschild term. In addition, a particular solution is needed. After substituting the functions f_0 and u_0, the equation for f_1 becomes

$${f_1}'' + \frac{2}{x} {f_1}' = c_0^2 \left(3 \frac{e^{-2x}}{x^2} + 2 \frac{e^{-2x}}{x^3} + \frac{e^{-2x}}{x^4} \right) \tag{8.45}$$

The particular solution of this equation is

$$f_{1p} = c_0^2 \left(\frac{1}{2} \frac{e^{-2x}}{x} + \frac{1}{2} \frac{e^{-2x}}{x^2} + \int \frac{e^{-2x}}{x} dx \right) \tag{8.46}$$

This expression is positive-definite, which will be important in the subsequent interpretation. The last term can be integrated by parts to give a slight simplification, which is

$$f_{1p} = \frac{c_0^2}{2}\left(\frac{e^{-2x}}{x^2} + \int_x^\infty \frac{e^{-2x}}{x^2}dx\right) \tag{8.47}$$

The metric function e^ν , with appropriate renormalization of the constants, can then be written in the form

$$e^\nu = 1 - \frac{2MG}{c^2 r} + \frac{q^2 G}{\epsilon_0 c^4}\left(\frac{e^{-2\mu r}}{r^2} + \mu^2 \int_r^\infty \frac{e^{-2\mu r}}{r^2}dr\right) \tag{8.48}$$

In the above equation, it has been assumed that the total classical energy of the field contributes to the gravitational field. In the limit as $\mu \to 0$, corresponding to an infinite range for the vector potential, a Reissner-Nordstrom spacetime is recovered.

In the early universe, it is thought, numerous micro black holes may have been created. These black holes would be expected to evaporate over time due to emission of thermal radiation. The positive Proca terms in the above metric suggest the possibility that some of these objects might be devoid of event horizons, in agreement with the earlier numerical solutions of Obukov and Vlachynsky and Toussaint.

Another interesting property of the above solution is that the gravitational field is repulsive when the constants take on suitable values, since as r gets very small the exponential terms will dominate. One is left to speculate whether such repulsive effects could prevent complete catastrophic gravitational collapse.

Method 2: Perturbation of Flat Space

Method 3: Perturbation of the Exterior Schwarzchild Solution

Bibliography

[1] Robin Tucker, private communication.

[2] D. Kramer, H. Stephani, E. Herlt, M. MacCallum, "Exact Solutions of Einstein's Field Equations", Cambridge University Press, Cambridge, 1980.

[3] T. Dereli, M. Onder, Jorg Schray, Robin W. Tucker, and Charles Wang, "Non-Riemannian gravity and the Einstein-Proca system", Classical and Quantum Gravity 13 (1996) L103-L109.

[4] R.W. Tucker and C. Wang, "An Einstein-Proca-fluid model for dark matter gravitational interactions", Nuclear Physics B (Proc.Suppl.), 57 (1997), 259-262.

[5] Obukov, Y.and Vlachynsky, E.J. ,arXiv:gr-qc/0004081 v1 28 April 2000.

[6] Toussaint, M.,arXiv:gr-qc/991042 v1 12 October 1999.

[7] Proca, Le Journal de Physique et le Radium 7, 347 (1936).

[8] Adler R., Bazin M., and Schiffer M., "Introduction to General Relativity, 2nd Edition",McGraw-Hill, New York, 1975.

[9] Wald, R., "General Relativity", University of Chicago Press, Chicago, 1984.

8.2 The PPN Formalism

8.2.1 Brans-Dicke

8.2.2 Rosen Bimetric Theory

8.3 Numerical Techniques

Chapter 9

Gravity Waves

The generation of gravity waves and their direct detection is one of the greatest test of general relativity. The LIGO project, championed and godfathered by Kip Thorne, is an ambitious project with the goal of detecting directly the emission of gravity waves by systems of collapsed objects.

9.1 Einstein-Rosen metric

One of the first wave solutions in general relativity was the famous Einstein-Rosen metric. The trial metric is taken to have the form

$$ds^2 = e^{2\gamma-2\psi}\left(dt^2 - d\rho^2\right) - e^{-2\psi}\rho^2 d\phi^2 - e^{2\psi}dz^2 \tag{9.1}$$

Where $\psi = \psi(t,\rho)$ and $\gamma = \gamma(t,\rho)$. This is, of course, not the only possible trial metric for cylindrical symmetry. Einstein's equations then yield:

$$\frac{\partial^2\psi}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial\psi}{\partial\rho} - \frac{\partial^2\psi}{\partial t^2} = 0 \tag{9.2}$$

$$\frac{\partial\gamma}{\partial t} = \rho\left[\left(\frac{\partial\psi}{\partial\rho}\right)^2 + \left(\frac{\partial\psi}{\partial t}\right)^2\right] \tag{9.3}$$

9.2 The Binary Pulsar

In 1975 Brian Hulse, a graduate student working at Arecibo, detected a series of curious signals from a pulsar in deep space. At first he and his research advisor, Taylor, thought it was only a glitch in the system, but Hulse kept hacking away at the equipment. Finally, it came out that the pulsar signal was coming from two distinct pulsars tightly orbiting each other.

Any pair of stars bound to each other gravitationally represents a system with a non-zero quadrupole moment, and if the stars are close enough together, there is the possibility that significant energy in the form of gravitational waves will be emitted. This loss of energy will result in the stars approaching each other, which in turn will shorten the period of the motion. And over a period of years, this shortening of the period was measured precisely, and found to be in agreement with the prediction of general relativity. This observation is the most accurately measured of any in physics, which makes GR the best-tested theory in all of physics.

The derivation begins with an analysis of plane waves propagating through vacuum far from the source. Obviously, $R_{ab} = 0$ in this region. In addition, since the waves are weak, it is possible to make some approximations. The analysis begins with

$$\nabla^c \nabla_c \bar{h_a} b = 16\pi \frac{G}{c^2} T_{ab} \tag{9.4}$$

9.3 LIGO

9.4 Exercises

1. Compute the Christoffel symbols and verify the equations for the Einstein-Rosen Metric.
2. Estimate the power flux per unit length of gravitational radiation emitted by an Einstein- Rosen system.

Chapter 10

Stellar Structure

10.1 Newtonian Stars

Finding the structure of a star means finding the mass, radius, and the density and pressure as a function of position. In addition, quantities such as the moment of inertia, temperature, and compression modulus are interesting and useful. Since the equations are generally intractable, numerical calculations are required, and are best illustrated with simple Newtonian models. Consider a thin box of gas at a radius r inside a star. In equilibrium, there are three forces on this box: gravity, acting inwards; $P_1$the pressure on the inside face, acting outwards; and P_2, the pressure on outside face, acting inwards (pressure on the sides of the box balance and can be safely forgotten). The derivation of the central equation follows from elementary principles:

$$P_1 A - P_2 A - \frac{GMm}{r^2} = 0 \rightarrow P_2 - P_1 = -\frac{GM\rho A \Delta r}{Ar^2}$$

M is the total mass inside a given radius r, whereas m is the mass of the material in the box, with ρ its density and A the cross-sectional area. In the above equation, we used the fact that $m = 4\pi r^2 \rho \Delta r$. This leads to the system of equations

$$\frac{dP}{dr} = -\frac{GM\rho}{r^2}$$

$$\frac{dM}{dr} = 4\pi r^2 \rho$$

To obtain the structure of a star, we have only to integrate this equation outwards from the core until the pressure is zero. This can be done with Euler's method, in which we basically back up a couple steps and replace the derivatives with discrete differences. Call the pressure at the center P_0 . Label the pressure one step further out by P_1, and n-steps out by P_n. From the above equation, we can derive the following difference equations:

$$P_{n+1} = P_n - \frac{GM\rho}{r^2}\Delta r$$

$$M_{n+1} = M_n + 4\pi r^2 \rho \Delta r$$

Notice that if at the point r_o we knew all the values on the right hand side, we would then be able to easily compute the pressure one step further out from the center, r_1. Once we know the values at r_1, we repeat the process until the pressure drops below zero, which indictates we have left the star. . Before we can start this process of interation, however, it is necessary to prescribe an equation of state linking the pressure and density and possibly other thermodynamic and physical quantities. $PV = nRT$ is an example of such an equation of state. To solve the structure of the star, we use the following prescription:

1. Specify a central density 2. Obtain and from the equation of state. 3. Plug into the pressure equation. This gives us , since . 4. Obtain and from the mass equation and the equation of state. 5. Plug into the pressure equation, and get . 6. Continue this process until the pressure first drops below zero. At this point, M will be equal to the stellar mass, and r to the stellar radius.

All you need is a do-loop that will march you out to the periphery. Experiment with step size for the highest value, starting crude and going to a smaller step. Caution: a counter in your program, which has the function of preventing infinite loops in the event of a bug or failure to satisfy a conditional statement, should be increased for very small step size. This is especially crucial for the white dwarf or a larger star, else the program will never reach the periphery where the pressure drops to zero. The code can easily be checked by comparing the numerical constant density solution with the exact solution.

10.2 Relativistic Stars

In stars the pressures and densities can become very high, so it is appropriate to use General Relativity to describe the equilibrium configurations. The operative equation is the Tolman- Openheimer-Volkoff equation derived in Chapter 5, together with supporting equations. To recap, these are:

$$p' = - - \frac{GM\rho}{r^2}\left(1 + \frac{P}{\rho c^2}\right)\left(1 + \frac{4\pi p r^3}{Mc^2}\right)\left(1 - \frac{2MG}{c^2 r}\right)^{-1} \tag{10.1}$$

$$\nu' = -\frac{2p'}{\rho c^2 + p} \tag{10.2}$$

$$e^{-\lambda} = 1 - \frac{2MG}{c^2 r} \tag{10.3}$$

$$M' = 4\pi r^2 \rho \tag{10.4}$$

$$p = p(\rho) \tag{10.5}$$

The equations have been arranged so as to explicitly show factors of G and c. These five equations must be solved when finding the structure of static stars. A considerably more challenging set must be solved in the rotating case, where a single equation can take several lines to write down. The mass and radius are of the greatest interest, since these are measurable quantities. If there is no particular interest in the metric potentials ν and λ, equations 10.2 and 10.3 can be completely ignored.

Some care must be taken in the initial step of any numerical routine, since in some terms a zero appears in the denominator. These problems don't occur in the limit as r goes to zero, so a simple expedient is to start the integration one step away from the center of the star. Since it can be shown that a relative maximum exists in the pressure at r=0, there is no change in moving from the center to the first grid point, anyway.

10.3 The Equation of State

The only equation left to determine is the equation of state, 10.5. This equation is generally derived and put in tabular form, since most of the time it isn't possible to express the results in closed form. The simplest example is

$$\rho = constant \tag{10.6}$$

This EOS can be used to test a numerical routine, since an exact solution is available. Another often-used EOS, especially in cosmology, is

$$P = \alpha \rho c^2 \tag{10.7}$$

where $0 \leq \alpha \leq 1$. The case $\alpha = 0$ corresponds to dust–effectively pressure-free, while for 'stiff matter' $\alpha = 1$. In this latter extreme case, the adiabatic speed of sound, given by

$$\frac{dp}{d\rho} = v_s^2 \tag{10.8}$$

equals the speed of light. Superluminal sound speed is generally regarded as physically impossible, though little is really known of the nature of matter at such extreme densities. In fact, no actual particle would travel faster than light in this case, only a signal, and it may be there would be some bizarre quantum phenomena associated with tunneling.

A common EOS that is frequently used is

$$P = kN^{\Gamma} \tag{10.9}$$

N is the number density, and the constant K is usually chosen so as to make the pressure continuous when matching up with some other equation of state, say at lower density. Γ is called the adiabatic index, the heat capacity at constant pressure divided by the heat capacity at constant volume. It takes the value of 5/3 for typical gas, and goes to 4/3 for a relativistic gas. At ultra high density it tends to increase to about 2, and might go as high as 3 in extreme cases, though a higher value than that would be considered unlikely. For this equation of state, it's not hard to derive an expression for the energy density. The derivation starts from the first law of thermodynamics,

$$dQ = d\frac{\rho}{n} + Pd\left(\frac{1}{n}\right) \tag{10.10}$$

where ϵ is the energy per baryon and n is the number density, while Q is the heat exchanged and P is the pressure. Notice that $1/n$ is a volume per baryon. In the adiabatic case, dQ=0, and a straight-forward integration gives

$$\rho = \frac{P}{c^2\left(\Gamma - 1\right)} + DN \tag{10.11}$$

D is again a constant that would be chosen by matching the pressure and energy densities to some known true or trusted values. When matching to nuclear densities, this value typically turns out to be equal to one AMU.

A number of different equations of state have been derived, and are operative in different density regimes. One of the original is the Feynman-Metropolis-Teller (aka FMT) equation of state, which works for lower densities up to about $1x10^4 gm/cm^3$ The Harrison-Wheeler EOS is based on models of nuclear matter, and is valid up to nuclear densities. This has been supplanted by other models, most notable BPS (Baym-Pethick-Sutherland) and BBP (Baym-Bethe-Pethick) equations of state. All of these may be used with reasonable confidence up to nuclear density, which occurs at about $2x10^14gm/cm^3$. Beyond that, it's anyone's guess, though that hasn't stopped numerous researchers from deriving candidates.

10.3.1 EOS for a degenerate, ideal Fermi gas

White dwarfs and neutron stars cool to T=0 after some considerable time. At such a temperature, the matter would be a completely degenerate collection of fermions. **Completely degenerate** means they fill all the lowest available energy states. In order to calculate thermodynamic quantities, it is necessary to find an expression for the number of occupied states in a given cell in phase space. The total volume of phase space is given by

$$V_{tot} = V_x V_p \tag{10.12}$$

where V_x is the ordinary volume, and V_p is the volume of momentum space. Particles occupy a certain volume in phase space that can be determined from the DeBroglie relationship:

$$p\lambda = h \tag{10.13}$$

A particle will occupy a space about as large as its wavelength, so its position volume is λ^3, while the momentum volume is p^3. Larger momentum means smaller wavelength and vice versa, so this means a typical cell will have volume h^3. Each cell, in addition, can have more than one particle, depending on the spin g of the particle. For massive particles $g = 2S + 1$, while for photons $g = 2$

and for neutrinos $g = 1$. Then the total number of available states in the phase space, N, is given by

$$N = \frac{gVV_p}{h^3} \tag{10.14}$$

This doesn't say anything about the number of particles in a given cell, or state. To get the average number of particles occupying a cell with energy given between E and $E + dE$, it is necessary to multiply by a distribution function, f, which gives the average number of particles oer cell. Dividing by the total volume of phase space then gives a phase space number density, $\aleph$:

$$\frac{d\aleph}{d^3xd^3p} = \frac{g}{h^3}f \tag{10.15}$$

From this phase space number density, the ordinary number density, n, can be calculated, as well as the energy density ρ and the pressure, p. These are given by

$$n = \int \frac{d\aleph}{d^3xd^3p}d^3p \tag{10.16}$$

$$\rho = \int E\frac{d\aleph}{d^3xd^3p}d^3p \tag{10.17}$$

and

$$P = \frac{1}{3}\int pv\frac{d\aleph}{d^3xd^3p}d^3p \tag{10.18}$$

In the above equations,

$$E = \left(p^2c^2 + m^2c^4\right)^{1/2} \tag{10.19}$$

and

$$v = \frac{pc^2}{E} \tag{10.20}$$

For fermions, the distribution function takes the Fermi-Dirac form,

$$f(E) = \frac{1}{exp((E-\mu)/kT) + 1} \tag{10.21}$$

where k is Boltzman's constant and μ is the chemical potential. As $T \to 0$, $\mu \to E_F$, the Fermi energy, and the distribution function takes the form

$$f(E) = \begin{cases} 1 & E \leq E_F \\ 0 & E > E_F \end{cases} \tag{10.22}$$

Finding the equation of state consists of carrying out the integrals. For a gas of electrons, the number density becomes

$$n_e = \frac{2}{h^3}\int_0^{p_F} 4\pi p^2 dp = \frac{8\pi}{3h^3}p_F^3 \tag{10.23}$$

where the Fermi momentum is defined by

$$E_F = \left(p_F^2 c^2 + m_e^2 c^4\right)^{1/2} \tag{10.24}$$

Define a dimensionless Fermi momentum by

$$x = \frac{p_F}{m_2 c} \tag{10.25}$$

Then

$$n_e = \frac{1}{3\pi^2 \lambda_e^3} x^3 \tag{10.26}$$

The energy density is

$$\rho_e = \frac{2}{h^3} \int_0^{p_F} \left(p^2 c^2 + m_e^2 c^4\right)^{1/2} r\pi p^2 dp = m_2 c^2 \lambda_e^3 \chi(x) \tag{10.27}$$

where

$$\chi(x) = \frac{1}{8\pi^2} \left\{ x\left(1+x^2\right)^{1/2}\left(1+2x^2\right) - \ln\left(x + \left(1+x^2\right)^{1/2}\right) \right\} \tag{10.28}$$

The pressure is given by

$$P_e = \frac{1}{3}\frac{2}{h^3} \int_0^{p_F} \frac{p^2 c^2}{\left(p^2 c^2 + m_e^2 c^4\right)^{1/2}} 4\pi p^2 dp = \frac{8\pi m_e^4 c^5}{3h^3} \int_0^x \frac{x^4 dx}{\left(1+x^4\right)^{1/2}} =$$

$$= \frac{m_e c^2}{\lambda_e^3} \phi(x) = 1.4218x10^{25} \phi(x) dyne - cm^{-2} \tag{10.29}$$

with

$$\phi(x) = \frac{1}{8\pi^2} \left\{ x\left(1+x^2\right)^{1/2}\left(2x^2/3 - 1\right) + \ln\left(x + \left(1+x^2\right)^{1/2}\right) \right\} \tag{10.30}$$

While the pressure may be primarily supplied by the degenerate electrons, the rest mass density is still mainly due to the ionized atoms. The density is given by

$$\rho_0 = \sum n_i m_i \tag{10.31}$$

with n_i the number density and m_i the mass of the i^{th} atomic species. The baryon number density is given by

$$n = \sum n_i A_i \tag{10.32}$$

where A_i is the atomic weight. An average baryon rest mass can be defined by

$$m_B = \frac{1}{n} \sum n_i m_i \tag{10.33}$$

Then the rest mass density can be written as

$$\rho_0 = n m_B = \frac{n_e m_B}{Y_e} \tag{10.34}$$

where Y_e is the average number of electrons per baryon. These equation give the EOS parametrically in x. The number density of electrons is evidently related to the number density of baryons. Given n_e, it is possible to compute x corresponding to it, then also to obtain the energy density and pressure and average baryon number density. In a numerical routine, once the new pressure is found, it will be necessary to invert the equation through some root-finding method, such as the bisection method, to find the new value of x. This value can then be plugged in to find n and ρ for the next loop.

10.3.2 Harrison-Wheeler EOS

10.3.3 BPS/BBP

10.3.4 BJ EOS

10.4 Exercises

1. Write a program solving a constant density Newtonian star, and then compare your results to the exact solution, which can be easily derived

2. Generalize Exercise 1 to the case of relativistic stars, and solve in the case of constant density, again comparing it to the exact solution.

3. Show that

Chapter 11

Cosmology

Cosmology is the study of the structure and evolution of the universe. In the early twentieth century, the universe was thought to be finite, consisting mainly of the Milky Way Galaxy. Various nebula, from irregulars to elliptics and dramatic spirals, were thought to be clouds of gas within the galaxy. Around 1930, when Hubble discovered the red shift of these nebulas, and others noted the appearance of supernova, was it found that these nebulae were in fact galaxies in their own right, and that they were distributed across vast distances. Currently, the microwave background appears as strong evidence that the universe began in a gigantic explosion some fifteen billion years ago. It useful to develop mathematical models of the universe as a whole in order to better understand its structure and evolution.

11.1 Newtonian Cosmology

Newton thought that the universe was infinite and filled with stars. If it were not infinite, he reasoned that the stars would eventually collapse together, and that was not apparently the case. With Newton's law of gravitation it is possible to derive a model of an expanding universe that has many features in common with the solutions of general relativity (see Guth and Steinhardt,1989). First, we start with Newton's law of gravity, in particular the total kinetic and potential energy of a portion of the universe, m:

$$E_{tot} = \frac{1}{2}mv^2 - \frac{GMm}{r} \tag{11.1}$$

Assuming the density is constant, we can use

$$M = \frac{4\pi}{3}r^3\rho \tag{11.2}$$

together with Hubble's law,

$$v = Hr \tag{11.3}$$

Substituting these last two equations into 11.1results in

$$E_{tot} = \frac{1}{2}mr^2\left(H^2 - \frac{8\pi}{3}G\rho\right) \tag{11.4}$$

As in classical mechanics, the test particle can escape whenever $E_{tot} \geq 0$. The critical mass density at which escape is possible, therefore, is given by

$$\rho_c = \frac{3H^2}{8\pi G} \tag{11.5}$$

This result would be modified if a cosmological constant term were added, corresponding to an additional potential energy. Equation 11.1 would then be modified to read

$$E_{pot} = -\frac{GMm}{r} - \frac{1}{6}\Lambda mc^2r^2 \tag{11.6}$$

This corresponds to a repulsive force; changing the sign would make it attractive.

An expanding universe model can be obtained from this Newtonian model by supposing that the variable r scales with time. Let s be a standard distance between two galaxies at a given time. If the universe is expanding with time, then for some function $R(t)$

$$r(t) = R(t)s \tag{11.7}$$

$R(t)$ is called the scale factor. Hubble's law follows from the assumption of expansion, since taking the derivative of 11.7 with respect to time (indicated with a dot) gives:

$$v = \dot{r} = \dot{R}s = \frac{\dot{R}}{R}Rs = \frac{\dot{R}}{R}r \tag{11.8}$$

Hubble's contant is therefore given by

$$H = \frac{\dot{R}}{R} \tag{11.9}$$

Inserting this expression into the total energy, equation 11.4, results in

$$E_{tot} = \frac{1}{2}ms^2R^2\left(H^2 - \frac{8\pi}{3}G\rho\right) \tag{11.10}$$

Next, define a parameter k by

$$k = -\frac{2E_{tot}}{mc^2s^2} \tag{11.11}$$

With this definition, equation 11.10 can be rearranged to read

$$k = \frac{1}{c^2}R^2\left(\frac{8\pi}{3}G\rho - H^2\right) \tag{11.12}$$

Note that k, as defined, is a constant, is the same for all test particles, and is independent of time and position. This means that its value determines properties valid for the entire space-time. Indeed, from equation 11.11, it is clear that if $k > 0$ then $E_{tot} < 0$ and the spacetime will collapse. If $k \leq 0$, on the other hand, then $E_{tot} \geq 0$ and the universe will expand forever. The size of k is immaterial, because by definition it depends on a chosen, fixed value of s, which is arbitrary. By convention, then, k is chosen to be -1,0, or 1. $k = 1$ corresponds to a closed universe that expands and then falls back on itself, while for $k = 0$ or 1 the universe is open, meaning it will expand forever.

To obtain a differential equation for R, substitute equation 11.9 into equation 11.12:

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3}G\rho(t) - \frac{kc^2}{R^2} \tag{11.13}$$

This gives us one equation with two unknown function, the density, $\rho(t)$, and the scale factor, $R(t)$. A relationship between them can be obtained from equations 11.2and 11.7:

$$\rho = \frac{3M}{4\pi s^3 R^3} \tag{11.14}$$

These two equations can now be solved for the three different values of k. In the case of $k = 0$, the solution is

$$R = b_o t^{2/3} \tag{11.15}$$

where b_o is a constant. The Hubble constant can then be determined using equation 11.9:

$$H = \frac{2}{3t} \tag{11.16}$$

For the case of $k = \pm 1$, the solution is more complicated, and must be given in terms of a parametrization rather than an explicit function of cosmic time.

If the cosmological constant is non-zero, the same analysis leads from equation 11.6 to

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3}G\rho(t) + \frac{1}{3}\Lambda c^2 - \frac{kc^2}{R^2} \tag{11.17}$$

The original form of the equation can be regained by making the following definition:

$$\rho_{eff} = \rho + \frac{\Lambda c^2}{8\pi G} = \rho + \rho_{vac} \tag{11.18}$$

With this vacuum energy density, equation 11.17 becomes

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3}G\rho_{eff}(t) - \frac{kc^2}{R^2} \tag{11.19}$$

This background mass density must be provided by one or another quantum fields. The empirical bound on Λ is about $3x10^{-52}m^{-2}$, hence the bound on ρ_{vac} is about $1.6x10^{-26}kg$. ρ_{vac} can supercharge the rate of inflation in early universe, which, as will be seen, can provide answers to a number of problems.

11.1.1 The deceleration parameter

Measuring the rate of change of expansion, the deceleration, is key to answering the question of whether or not the universe is going to contract. The deceleration is proportional to the mass density, and the mass density, in turn, determines the geometry of the universe.

First, differentiating equation 11.14 yields

$$\dot{\rho} = -3\frac{\dot{R}}{R}\rho \tag{11.20}$$

Multiplying equation 11.13 by R^2 and differentiating, and then substituting equation 11.20 results in

$$\ddot{R} = -\frac{4\pi}{3}G\rho \tag{11.21}$$

The deceleration is proportional to the mass density, hence measurement of this parameter will yield the density, which determines whether the universe is open or closed. Some modification of this equation is necessary in the relativistic case, and while this can be inferred from

11.2 Einstein Cosmologies

The universe may be considered a homogeneous, isotropic gas, where the particles of gas are in fact galaxies. Homogeneous means that the universe is essentially the same everywhere, whereas isotropic means that the universe, from a given vantage point, looks the same in all directions. These conditions give us exactly three trial metrics, corresponding to flat, spherical, and hyperbolic symmetry:

$$ds^2 = -d\tau^2 + a(\tau)^2 d\sigma^2 \tag{11.22}$$

where

$$d\sigma^2 = \begin{cases} d\psi^2 + \sin^2\psi\left(d\theta^2 + \sin^2\theta d\phi^2\right) & \text{k=1} \\ dx^2 + dy^2 + dz^2 & \text{k=0} \\ d\psi^2 + \sinh^2\psi\left(d\theta^2 + \sin^2\theta d\phi^2\right) & \text{k=-1} \end{cases} \tag{11.23}$$

The first expression corresponds to spherical symmetry–actually $R\bigotimes S^3$, where R stands for the real line and S^3 is a three-dimensional sphere (imbedded in a four-dimensional space-time). The second expression is a flat space, which has topology R^4. The last expression is a three-dimensional hyperbola crossed with R. As in the Newtonian analysis, the first corresponds to $k = 1$, the second to $k = 0$, and the third to $k = -1$. $a(\tau)$ is the scale factor. The g_{00} metric term can be taken to be equal to -1 because a trivial rescaling of the time coordinate can make it disappear (see Exercise 3).

Of the three symmetries, we will completely solve the case k=0. The Christoffel symbols are given by

$$\begin{gathered} \Gamma^0_{11} = \Gamma^0_{22} = \Gamma^0_{33} = a\dot{a} \\ \Gamma^1_{10} = \Gamma^1_{01} = \Gamma^2_{02} = \Gamma^2_{20} = \Gamma^3_{03} = \Gamma^2_{30} = \dot{a}/a \end{gathered} \tag{11.24}$$

with all other components equal to zero. The nonzero Ricci tensor components are

$$R_{00} = -3\ddot{a}/a$$
$$R_{11} = R_{22} = R_{33} = \ddot{a}/a + 2\dot{a}^2/a^2 \tag{11.25}$$

Computing the Ricci scalar and then assembling Einstein's equations results in:

$$G_{00} = 3\frac{\dot{a}^2}{a^2} = 8\pi\rho \tag{11.26}$$

$$G_{**} = -2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} = 8\pi P \tag{11.27}$$

Substituting equation 11.26 into equation 11.27 yields

$$3\frac{\ddot{a}}{a} = -4\pi\left(\rho + 3P\right) \tag{11.28}$$

The cases for spheroidal and hyperboloidal symmetry are handled similarly. The resulting equations are

$$3\frac{\dot{a}^2}{a^2} = 8\pi\frac{G}{c^2}\rho - \frac{3k}{a^2} \tag{11.29}$$

$$3\frac{\ddot{a}}{a} = -4\pi\frac{G}{c^2}\left(\rho + 3\frac{P}{c^2}\right) \tag{11.30}$$

These equations can be solved exactly for a variety of special equations of state. P=0 gives the Friedmann solutions, identical to what was obtained in the previous section. Another common equation of state is

$$P = \alpha\rho c^2 \qquad 0 \le \alpha \le 1 \tag{11.31}$$

where α is just a parameter and ρ is, naturally, the energy density (as opposed to the mass density of the previous section). $\alpha = 0$ models dust, which may be thought of as pressure-free matter, while $\alpha = 1/3$ corresponds to radiation. The condition $\nabla^a T_{ab} = 0$ yields the dynamics of the cosmological fluid:

$$\frac{d}{d\tau}\left(\rho c^2 a^3\right) \quad = \quad -\frac{p}{c^2}\frac{d}{d\tau}a^3 \tag{11.32}$$

After the initial expansion, the pressure drops effectively to zero, yielding dust. The above equation can then be integrated to give

$$\rho = \frac{c_0}{a^3} \tag{11.33}$$

This can be substituted into Einstein's equations:

$$G_{00} = 3\frac{\dot{a}^2}{a^2} = 8\pi\rho = 8\pi\left(\frac{c_0}{a^3}\right) \tag{11.34}$$

This can be easily solved, giving, as before

$$a \approx t^{2/3} \tag{11.35}$$

For radiation-dominated, $P = \rho c^2/3$. Substituting into the equations of motion and integrating results in

$$\rho = \frac{\alpha}{R^4} \tag{11.36}$$

This, too, can be easily integrated in the flat space case.

11.3 Red Shift and Acceleration

The Robertson-Walker metric, derived in the last chapter, can also be written as

$$ds^2 = d\tau^2 - R(\tau)^2\left(\frac{dr^2}{1-Kr^2} + r^2 d\Omega^2\right) \tag{11.37}$$

The red shift formula uses a parameter called z, and is defined by

$$1 + z = \frac{\Delta\lambda}{\lambda_{emitted}} = \frac{R_{observed}}{R_{emitted}} \tag{11.38}$$

To see how this is derived, start with the metric in the form

$$ds^2 = d\tau^2 - a^2 d\sigma^2 \tag{11.39}$$

Light travels on null geodesics. So a photon emitted at τ_e and observed at a later time τ_0 will traverse spacetime according to a parameter σ, given by

$$\int_{\tau_e}^{\tau_0} \frac{d\tau}{a} = \sigma \tag{11.40}$$

If the observed quantity is periodic, the next observation will occur $\Delta\tau_0$ later, due to emission occuring $\Delta\tau_e$ later than the first. If these differences, which are equal to the periods at the respective times, are small compared to the total travel time, then they can be represented as

$$\int_{\tau_e+\Delta\tau_e}^{\tau_0+\Delta\tau_0} \frac{d\tau}{a} = \sigma \tag{11.41}$$

Subtract these two expressions:

$$\int_{\tau_e}^{\tau_0} \frac{d\tau}{a} - \int_{\tau_e+\Delta\tau_e}^{\tau_0+\Delta\tau_0} \frac{d\tau}{a} = 0 \tag{11.42}$$

With a little inspection, and maybe a cosmological number line, this is seen as equal to

$$\int_{\tau_e}^{\tau_e+\Delta\tau_e} \frac{d\tau}{a} - \int_{\tau_0}^{\tau_0+\Delta\tau_0} \frac{d\tau}{a} = 0 \tag{11.43}$$

Dividing by $\Delta\tau_e\Delta\tau_0$, and denoting the anti-derivative by F, it then follows from the definition of derivative that

$$\frac{1}{a(\tau_e)\Delta\tau_0} = \frac{1}{a(\tau_0)\Delta\tau_e} \tag{11.44}$$

This results immediately in

$$\frac{\lambda_0}{\lambda_e} = \frac{a_0}{a_e} \tag{11.45}$$

The z-factor is then defined as

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{a_0}{a_e} - 1 \tag{11.46}$$

11.4 Inflationary Cosmologies

11.5 Exercises

.1. Starting from 11.6, derive 11.17.

2. Suppose the metric is taken to be $ds^2 = -A(t)^2 dt^2 + B(t)^2 d\sigma^2$. Show that by coordinate transformation this metric is the same as given in 11.22 and 11.23.

3. Solve equation 11.32 for the general case, with $0 \leq \alpha \leq 1$. Solve for the metric function a in the case of k=0.

4. Find the equation for the field and the stress-energy for the following Lagrangian density:

$$\pounds = \left(\nabla_a \psi \nabla^a \psi - \frac{m^2 c^2}{\hbar^2} \psi^2 \right) \sqrt{-g}$$

don't do the next one!! 5. Repeat the analysis of section 2, assuming gravity is given by a massive scalar field, with potential energy $E_{pot} = GMme^{-\mu r}/r$.

Chapter 12

Exotic Structures

12.1 Geometric Relics of the Big Bang

Particle physics, when applied to the early universe, yields the possibility of dramatic and exotic structures that may have a major impact on the evolution and structure of the universe. These structures are related to symmetry breaking, and result in exotic defects in the structure of spacetime in zero, one, two, and three physical dimensions. In this section, these defects will be reviewed, then presented in mathematical detail in subsequent sections. Defects in the structure of spacetime result from something called spontaneous symmetry breaking, and come in several sizes and shapes. First of all, what is spontaneous symmetry breaking? A good example of symmetry breaking is what happens when water turns to ice. As liquid water, the molecules can swirl around any way they wish. You can think of the water as having a 'rotational symmetry', since any way you look at it, it looks the same. When the water freezes to ice, however, you suddenly have a crystal with a regular structure. The lattice of crystal looks different when viewed from various directions. The former symmetry, as liquid water, is broken. Furthermore, ice crystals may start forming in different regions, with different orientations. When these regions grow together, defects will be observed between them: points, lines, and planes, which can be seen in ice cubes right out of your refrigerator. Now, as the universe expanded, it cooled, just as the freon in a refrigerator cools as it expands. And just as cooling water can turn to ice with defects, many scientists believe that interesting relic defects would have 'frozen out' during the initial phase of expansion. The relic defects predicted include: a point-like defect, called a magnetic monopole, a string-like defect called a cosmic string, a sheet-like defect, called a domain wall, and an extended defect, called a texture. Magnetic monopoles were first postulated by Paul Dirac in the thirties, and physicists have been looking for them unsuccessfully ever since. Imagine taking a magnet and cutting it in half. Ordinarily, you get two magnets, each of them having a north and south pole. This is because magnetism is caused by a circulation of electrons at the atomic level. A monopole, on the other hand, would be either a north or south pole in isolation. This is a rather radical idea, and goes contrary to our everyday experience. And such monopoles, if they exist, would be quite remarkable. First of all, though they'd be submicroscopic in size–far smaller than a proton–they'd be enormously heavy: a hundred thousand of them would weigh in at about a microgram! That doesn't sound like much, but comparing the weight of a hydrogen atom to the weight of a monopole is like comparing the weight of a child's paper boat to that of a battleship. For something so tiny, monopoles would be absurdly massive. A useful feature of monopoles would be their ability to annihilate all other forms of matter. They would eat everything in their path, but rather than keeping most of it down for a later grand burp like a microhole, they would convert the matter immediately into energy by stimulating proton decay. A car constructed around a closed cycle monopole steam engine could cruise for decades on one tank of water, with no pollution at all. Of course, they could also be used in highly efficient space drives, surpassing even antimatter propulsion. Unfortunately, only a single experiment has ever detected a candidate monopole, and subsequent efforts have failed. This is bad news for particle theorists, since there ought to be lots and lots of them, everywhere. Rapid universal expansion–particularly the turbocharged inflation model–may account for their scarcity, however, so all is not lost. A second type of monopole, the global monopole, also known as a hedgehog, would be so enormous that it s unlikely astronomers could possibly overlook it, no matter where it was in the universe. The mass energy of the hedgehog is thought to increase linearly with distance from the core. Hedgehogs and anti-hedgehogs would attract each other strongly, moving towards mutual annihilation at relativistic speeds. Because of this, not many are expected to be hanging around today. Some scientists think they may have created perturbations in the homogeneity of the universe, leading to large scale structure formation. Cosmic strings, not to be confused with the strings of string theory, are again predicted as forming during the inflationary period just after the Big Bang. Whereas ordinary magnetic monopoles are submicroscopic, cosmic strings, like hedgehogs, would be enormous, far larger than galaxies, long, twisting tubes of 'stress-energy' stretching for thousands, even millions of light years. Right after the Bang there might have been great networks of them, all tangled together like a ball of yarn the cat got into. Wherever they intersected themselves they'd tend to join and form loops, which would then spin like gigantic hula hoops, rapidly radiating away their substance in the form of gravity waves. One interesting property of cosmic strings is that the spacetime around them would be very nearly–but not quite–flat. This is paradoxical for such massive objects, but follows from solutions of Einstein s field equations. The only difference from flat space would be a curious deficit angle: the distance around a circle would be less than times the radius! A cosmic string, passing through an observer, would cause the opposite sides of that observer to approach each other at about a kilometer per second, not a healthy prospect. Cosmic strings and loops are thought to have formed the seeds of galaxies. Just how and why galaxies formed is one of the big unsolved mysteries in astrophysics. The problem is, despite all its violence, the

Big Bang was apparently a very smooth and uniform event, a kind of perfect explosion that was the same in every direction. The evidence for this is very strong in the microwave background, which has a uniform temperature across the sky, though recently faint ripples have been detected in this distribution. Without any lumps, wrinkles, or other imperfections in the expanding gas of the universe there wouldn't have been any reason for galactic-sized clumps to form, and the universe would have turned out to be a very boring place: just an expanding cloud of gas–no galaxies, no stars, no planets, no sentient beings to ever get interested in the origins of things. The cause of the recently-detected ripples in the microwave background has yet to be explained, though perhaps cosmic strings or other defects are involved. Still, there's the same problem with cosmic strings as with magnetic monopoles: they haven't been observed. But astronomers are still looking for them, and they may yet turn up if we look deep enough into space. Like cosmic strings, two-dimensional defects, called domain walls, are supposed to be fantastic, galaxy-spanning structures. They are thought to take either the form of sheets or of gigantic bubbles. The bubbles expand, contract, or oscillate in and out, while the sheets often move at relativistic speeds. The exact behavior depends on the nature of the spacetime on either side of the wall. Domain walls have a very intriguing property: if you're close to one, they push you away. That's right. They repell you, drive you further away. Domain walls exhibit repulsive gravity. If you could rig a small one to the back of a space ship, you could use it to travel anywhere you wanted, like an antigravity space drive. The tension of the wall, according to Einstein's equations, is what creates the repulsive effect. It acts like a source of negative mass, a hypothetical substance with remarkable properties. There are two problems with such a space drive, however. First, domain walls are under such incredible tension that any normal matter would be torn to pieces long before the repulsive gravity effect could be obtained. Ordinary everyday tensions, such as that created by stretching a rubber band, are too weak by about twenty powers of ten! Furthermore, the repulsive gravity effect is actually limited in range: at a distance, the wall becomes attractive again. Only universe-spanning walls would have universal repulsive gravity. Domain walls, like monopoles and strings, have not been observed, though my personal suspicion is that many of them may have turned into quasars. There are a tremendous number of quasars in deep space, and they are postulated to be driven by black holes. My graduate thesis advisor, Jim Ipser at the University of Florida, together with Pierre Sikivie, showed that spherical domain walls with a flat spacetime interior always collapse into black holes, and they would be whopping big ones, big enough to eat solar systems for breakfast just as quasars are purported to do. Textures are a three-dimensional defect–an extended region of false vacuum–and are candidates for the creation of structure in the universe. This type of defect, unlike the others, is expected to be unstable, rapidly collapsing in on itself at the speed of light, while spewing out tremendous amounts of energy in the form of Goldstone bosons, massless particles associated with quantum scalar fields. In the process of self-destructing, however, the texture—also called a knot –would create fluctuations in the density of matter, including galactic spheroids, which could then evolve into quasar-containing galaxies. So far, textures, like the other defects, haven t been detected. This may in part be due to the fact that thicker defects tend to unravel as the universe expands. Thinner ones would persist, though, and of course even the thicker ones could still be seen in deep space, which is equivalent to looking backwards in time. Some physicists are looking elsewhere for the perturbations needed to seed not only structures such as galaxies, but also, on a larger scale, the great voids and walls of galaxies. Currently, theories based on quantum fluctuations, caused by the Heisenberg Uncertainty Principle, are getting popular.

12.2 Magnetic Monopoles

12.3 Cosmic Strings

12.4 Domain Walls

12.5 Hedgehogs

12.6 Lorentzian Wormholes

Chapter 13

Singularity Theorems

Roger Penrose and Steve Hawking developed mathematical tools to study the structure of spacetime, and under fairly reasonable assumptions proved that the spacetime manifolds of classical general relativity had to have singularities. Intuitively, a singularity is a bad spot, a place where physical theories breakdown and information cannot be extracted.

A quantum theory of gravity would no doubt give entirely different results. What is also not universally appreciated is that the singularity theorems were reverse-engineered: all depend on conditions put on the stress-energy, conditions that are essential in order to get the desired results. And it turns out that many very common spacetimes violate those conditions. Nonetheless, it is of interest to see how these problems are addressed for classical general relativity. Good references include Penrose, Hawking and Ellis, Beam and Ehrlich, Wald, and Barrett.

13.1 Elementary Topology

13.2 Spacetime Preliminaries

1. Definition: A space-time M is a real, 4-dimensional connected C^∞ Hausdorff manifold, without boundary, together with a globally defined C^2 tensor field g of type (0,2) which is non-degenerate and Lorentzian. In addition, M will be taken to be time-oriented.

2. Explanation of terms in Definition 1.

A real, 4-dimensional manifold is a space which, locally, looks like 4-dimensional Euclidean space, R^4. Mathematically, this means that for every point p in M there is a neighborhood U and a homeomorphism $\phi : U \to \phi(U) \subset R^4$. These homeomorphisms are called charts, or coordinate systems. C^∞ means that when two coordinate systems overlap, then the composition of one with the inverse of the other is continuously differentiable to all orders.

Hausdorff means that any two points of M may be separated by a pair of disjoint open sets, i.e. there exists two open sets U,V in the topology of M such that $x \in U$,$y \in V$, with $U \cap V = \emptyset$. Locally this condition is satisfied automatically, since Euclidean space is Hausdorff. Globally, this condition may be suspended in certain examples (see Hawking and Ellis, page 177).

M is connected if it cannot be written as a disjoint union of two open sets, i.e. $M \neq U \cup V$ where $U \cap V = \emptyset$, with neither U nor V empty. This hypothesis is reasonable, since we could have no knowledge of a disconnected component.

"Without boundary" means that there are no neighborhoods homeomorphic to R^{4+}, where $R^{4+} = \left\{x \in R^4 s.t. x^1 \geq 0\right\}$. Points of M having an image on $x^1 = 0$ are called boundary points of M. The assumption that there are no such points is made on the basis of physical reasonableness (i.e. no spacetime deadends in the middle of nowhere!). A notion of singular boundary has been defined, however, whereby one puts in points representing the singularity and then tries to define a manifold with boundary structure.

The tensor field g, called the metric tensor, is C^2 in order to allow the Einstein field equations, which involve second derivatives, to be defined at every point. Non-degenerate means that if $g(V,W) = g_{ab}V^aW^b = 0$ for all W, (V and W being elements of a tangent vector space), then V=0. This condition ensures that g_{ab} has an inverse at every point. Lorentzian means that at each point p in M there is a basis for the Tangent space $T_p(M)$ in which g is Minkowskian, i.e. has matrix form given by the flat space metric tensor, η_{ab}.

A vector $V \in T_p(M)$ is said to be time-like, space-like, or null according as $g(V,V) = g_{ab}V^aV^b$ is negative, positive, or zero. The null cone at p is the collection of null vectors in $T_p(M)$. This null cone separates the time-like vectors into two componenets. If one can choose, in a continuous fashion, one of these components at each point of M, M is said to be time-orientable. To label these chosen components the future cones, and all the others the past cones, is to time-orient M.

3. Definition: A curve in a manifold M is a smooth mapping $\alpha : I \to M$, where I is an open interval in R.

Generally we will consider two curves differing only by a smooth parameter change to be the same curve. Intuitively this means we are taking a curve to be its underlying point set in M. A curve is regular ("kink-free") if its tangent vector is non-vanishing. A curve is time-like if its tangent vector is timelike at every point, and future-oriented if the vector is future-pointing at every point. A causal curve may have tangent vectors which are either null or time-like.

4. Definition: The future endpoint of a curve γ is that point $p \in M$ such that for all sequences $\{u_i\} \to b = \sup I, \gamma(u_i) \to p$. (Past endpoints are similarly defined)

All timelike or causal curves will be required to contain their endpoints (if they exist), or else be future and/or past endless. This excludes the curves depicted in figure 2 from being time-like curves: in the first case, the tangent vector waggles back and forth, so the curve cannot be smooth at the future endpoint, and in the second case, the curve becomes null only at the future endpoint. In addition, this requirement means hat curves containing both endpoints wil have a closed interval of R as their domain of definition.

5. Definition. Let $\bigtriangledown$ denote the unique torsion-free connection on M under which$g$ is covariantly constant, i.e. $\bigtriangledown g = 0$. Then $\nabla_X Y = X^a \bigtriangledown_a Y^b$ is the covariant derivative of the vector field Y in the direction of X— in other words, the usual directional derivative, generalized to manifolds. In terms of components,

$$X^a \bigtriangledown_a Y^b = X^a \left(\frac{\partial Y^b}{\partial x^a} + \Gamma^b_a c Y^c \right)$$

with the usual summation convention. Torsion-free means that $\Gamma^b_{ac} = \Gamma^b_{ca}$.

7. Definition. Let $p \in M$, $V \in T_p(M)$. Let γ^V be the geodesic with tangent vector V at p, with $\gamma_V(0) = p$. Then the exponential map, $exp : T_p(M) \to M$, is given by $exp_p(V) = \gamma_V(1)$, if it exists.

Let $\lambda \in [0, 1]$. Intuitively, one expects $exp_p(\lambda V)$ to be on the same geodesic, but not as far along if $\lambda < 1$.

On a sphere the exponential map is badly behaved. All vectors of sufficient length, say at the Tangent space of the north pole, could be mapped to the sout pole, giving a map which could not be one-to-one, obviously. However, for a small enough neighborhood it is a diffeomorphism from a region of $T_p(M)$ to a neighborhood on the manifold M. This neighborhood is called a normal neighborhood.

8.Definition: If Q is a star-shaped region containing the origin in $T_p(M)$, and if $exp_p \mid_Q$ (that is, the exponential map restricted to Q) is a diffeomorphism on Q, then $exp_p[Q]$ is called a normal neighborhood of p.

It can be shown that such a neighborhood exists for every point p. (See Postnikov, O'Neill, etc) Mathematically, this models the idea that spacetime in sufficiently small regions looks like flat space, i.e. Minkowski space.

"Star-shaped", in definition 8, means simply that $V \in Q \Rightarrow \lambda \in Q, \lambda \in [0, 1]$. A simply convex neighborhood is a normal neighborhood of each of its points (not just for a given point p). Such a neighborhood exists at each point of M. (See Postnikov for the proof).

9. Definition. A simple region N is a simply convex open set such that N closeure is compact and is itself contained in a simply convex open set.

10. Properties of simple regions 1. If $p, q \in N$, there exists a unique geodesic in N connecting them.
2. ∂N, the boundary of N, is compact, and closed subsets of N are also compact.
3. M can be covered by a locally finite system of simple regions. Also, finite collections cover compact subsets.

Locally finite means that every $x \in M$ has a neighborhood U which intersects only finitely many elements of the cover. This property really follows from the paracompactness of manifolds. The usefulness of simple regions will be evident at various junctures.

11. Definition: Let γ be a geodesic with tangent vector field T. A solution V of the geodesic deviation equation,

$$\nabla_T \nabla_T = -R(V, T)T$$

is called a Jacobi field on γ. In terms of components,

$$T^a\nabla_a\left(T^b\nabla_b V^c\right) = -R^c_{abd}V^bT^aT^d$$

Example: Find the Jacobi Fields in R^2 along the t-axis. Solution: $R^a_{bcd} = 0$ in flat space, hence the Jacobi Equation reduces to

$$\frac{\partial^2 V^c}{\partial t^2} = 0 \Rightarrow V^c = a^c t + b^c$$

with a^c and b^c being constant vectors. An example is illustrated in figure 5, for $V = t\hat{x}$. If the geodesics along t=x and the t-axis represent the worldlines of particles, it's clear that Jacobi Fields are relative position vectors pointing from one particular geodesic to neighboring geodesics. The first and second derivatives of V are the relative velocity and acceleeration, respectively.

12. Definition: Let γ be a geodesic, p,q distinct points on γ, and V a Jacobi Field vanishing at p and q but otherwise not identically zero. The p and q are called conjugate points on γ.

Intuitively, conjugate points represent places where neighboring geodesics intersect, though since the field is defined on γ no other such geodesic need actually exist.

As will be seen in a later section, a conjugate point is a place where light rays enter the chronological future of their point of origin. This phenomena is the well-known bending of light rays by gravitation (though mathematically obtained in a less usual way).

13. Example. The longitudinal lines on S^2 are geodesics. It is left as an exercise to show that $\partial/\partial\phi$ satisfies the Jacobi equation, and that the north and south poles are conjugate points. From Figure 6, it can be seen that $\partial/\partial\phi$ indeed vanishes at the poles, as required. Conjugate points will play a very primary role in the proving of the singularity theorems.

14. Definition. A smooth 1-parameter system of afinely-parameterized geodesics is a smooth map $\mu : (t_o, t_1)x(-\epsilon, \epsilon) \to M$such that for every $v_o \in (-\epsilon, \epsilon)$, $\mu(t, v_o)$ is an afinely parametrized (a.p.) geodesic.

A simple example is $\mu(\theta, \phi) =< \sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta >$, for $-\epsilon < \phi < \epsilon$ and $0 < \theta < \pi$, which parametrizes the geodesics following longitudinal lines on the sphere in a neighborhood of $\phi = 0$.

15. Proposition. Let$\gamma(t) = \mu(t, 0)$, and let $\partial/\partial t$ and $\partial/\partial v$ denote the coordinate vectors on M associated with $\mu(t, v)$. Then $partial/\partial v$ is a Jacobi Field on γ.

Discussion: μ maps a piece of R^2 onto M, with $\partial/\partial t$ and $\partial/\partial v$ the tangent vectors on R^2. The actual vectors on M should be denoted "$d\mu(\partial/\partial t)$" or "$d\mu_\star(\partial/\partial t)$" , for example. Calling them $\partial/\partial t$ and $\partial/\partial v$ is a common abuse of notation. ("$d\mu$" is the differential map of μ It's defined by $d\mu(\partial/\partial t)f = \partial/\partial t\,(f \circ \mu)$, where $\to R$ is a real valued function.)

For convenience we may choose $\partial/\partial t$ and $\partial/\partial v$ to be elements of the basis $\{\partial/\partial x^a$ of coordinate vectors. Also, set $\partial/\partial t = T = T^a\partial/\partial x^a$, and $\partial/\partial v = V = V^a\partial/\partial x^a$.

Proof: $[T,V] = \nabla_T V - \nabla_V T = 0$since in components,

$$T^a \nabla_a V^b - V^a \nabla_a T^b = T^a \left(\frac{\partial V^b}{\partial x^a} + \Gamma^b_{ac} V^c \right) - Va \left(\frac{\partial T^b}{\partial x^a} + \Gamma^b_{ac} T^c \right) =$$

$$= T^a \frac{\partial V^b}{\partial x^a} - V^a \frac{\partial T^b}{\partial x^a} + \Gamma^b_{ac} V^c T^a - \Gamma^b_{ac} V^a T^c = 0$$

Note that $T^a = V^a = 1$, and $\Gamma^a_{bc} = \Gamma^a_{cb}$ so the first two are zero and the last two cancel.

Hence $\nabla_T V = \nabla_V T$. Differentiate again, and add two terms which are both zero:

$$\nabla_T \nabla_T V - \nabla_T \nabla_V T = \nabla_T \nabla_V T - \nabla_V \nabla_T T + \nabla_{[V,T]} T = -R(V,T)T$$

$\nabla_T T = 0$ since $\mu(t, v_o)$ is a geodesic. $[V,T] = 0$, so $\nabla_{[V,T]} T = 0$. In the definition of R(X,Y)Z on page 5, set X=V, Y=T, and Z=T.

This completes the derivation of the Jacobi equation.

16. Proposition: Let T be the longitudinal vector field of a one-parameter system of geodesics. If $g(T,T) = g_{ab} T^a T^b$ is the same for each geodesic of the system, then $g(T,V) = g_{ab} T^a V^b$ is constant along γ.

Proof:

$$\nabla_T [g(T,V)] = \nabla_T g(T,V) + g(\nabla_T T, V) + g(T, \nabla_T V) = g(T, \nabla_T V) = g(T, \nabla_V T) = \frac{1}{2} \nabla_V [g(T,T)] = 0$$

This shows that g(T,V) is constant along the geodesics γ in the one-parameter family with tangent field T.

The above propositions will be used in the next section.

13.3 Causality and Chronology

1. Definition: A trip is a curve which is piece-wise a future-oriented timelike geodesic.

For a trip from x to y, write "$x << y$", which is read "x chronologically precedes y". Penrose is the originator of the idea: nearly everyone else uses timelike curves. For most purposes the two concepts are identical, and equivalent in the sense that whenever a time-like curve connects two points, there exists a trip also and vice versa. In many cases trips make for easier proofs. Both trips and timelike curves will be employed subsequently, according to convenience.

2. Definition: A causal trip is the same as a trip, except that causal geodesics, possibly degenerate, replace timelike geodesics. Write $p < q$ if "p causally precedes q".

Note that$p < p$ always makes sense, , degenerate meaning the trip consists of a single point. However $p << p$ is a violation of causality: one may return to the same oint in spacetime while traveling into the future. As will be shown later, compact spacetimes always admit such trips.

The requirement that curves contain their endpoints prevents the accumulation of an infinite number of joints, otherwise known as a bad trip. Hence, unless a trip is future or past endless, it will be composed of a finite number of "joints".

3. Proposition: $a << b \Rightarrow a < b$
$a << b, b << c \Rightarrow a << c$
$a < b, b < c \Rightarrow a < c$
Proof: This is immediate, from the definitions.

"Mixed transitivity" is more difficult (e.g. $a < b, b << c \Rightarrow a << c$). Proving this will be one of the major results of this section.

4. Definition: The set $I^+(x) = (y \in M \,|x << y)$ is called the chronological future of x. $I^-(x) = (y \in M \,|y << x)$ is called the chronological past of x. The set $J^+(x) = (y \in M \,|x < y)$ is called the causal future of x. $J^-(x) = (y \in M \,|y < x)$ is called the causal past of x.

The chronological future of a set is defined in the natural way: If $S \subset M$, then $I^+[S] = \cup_{x \in S} I^+(x)$. $I^-[S]$, $J^+[S]$, and $J^-[S]$ are defined similarly.

5. Examples

6. Proposition. $I^+(p)$ is open for any $p \in M$.

Proof: Let $x \in I^+(p)$. Then there is a trip γ from p to x. Let N be a simple region containing x, and let $y \in N$ be on the terminal segment of γ. Let $V = exp_y^{-1}(x)$, Then V is timelike adn belongs to the open set Q of future-pointing timeslike vectors in $exp_y^{-1}[N]$. But the exponential map is a homeomorphism when restricted to $exp_y^{-1}[N]$, so $exp_y(Q)$ is open in M and contains x. But $exp_y(Q) \subset I^+(y) \subseteq I^+(p)$, meaning that $I^+(p)$ must be open, since every point $x \in I^+(p)$ is contained in an open set, and $I^+(p)$ is the union of these open sets.

Note that the proof depends on the openness of Q. This fact follows from N and hence $exp_y^{-1}N$ being open (exp is a diffeomorphism), and the openness of the collection of future-ointing timelike vectors in $T_y(M)$. $T_y(M)$ and R^4 are homeomorphic spaces and the interior of the upper hypercone $t^2 = x^2 + y^2 + z^2$ is clearly open. Q is the intersection of the set of vectors homeomorphic to this interior, and the open set $exp_y^{-1}[N]$. The intersection of two open sets is open.

7. **Corollary.** The relation , , is open, i.e. if $p << q$ om $<$. there are disjoint open sets $\mathcal{U}$ and $\mathcal{V}$, $p \in \mathcal{U}$ and $q \in \mathcal{V}$, such that for all $p' \in \mathcal{U}$ and $q' \in \mathcal{V}$, $p' << q'$.
Proof: Exercise.

8. **More Corollaries, and exercises**
A. $x \in I^+(y) \iff y \in I^-(x)$
$x \in J^+(y) \iff y \in J^-(x)$

B. $I^+[S] = I^+[\bar{S}]$
C. $I^+[S] = I^+[I^+[S]] \subset J^+[S] = J^+[J^+[S]]$

The above corollaries will be used routinely when needed. The next series of propositions will lead to proof of "mixed transitivity" : $p << q, q < r \Rightarrow p << r$.

9. **Definition.** Let N be a simple region, and let $p, q \in N$. The world function $\Phi : N \times N \to R$ is defined by

$$\Phi(p,q) = g(\exp_p^{-1}(q), \exp_p^{-1}(q))$$

i.e. $\Phi(p,q)$ is the squared length of the geodesic pq from p to q (that is, of the unique geodesic lying in N). $\Phi(p,q) = \Phi(q,p)$, and is positive, negative, or zero according as pq is spacelike, timelike, or null. Continuity follows from the existence and uniqueness of the geodesic connecting p and q.

10. **Lemma.** Let NB be a simple region and let $p \in N$. The hypersurfaces $H_{p,k} = (q \,|\, \Phi(p,q) = K)$ are smooth in N (except at q=p) and are spacelike, timelike, or null according as the constant K is negative, positive, or zero. In addition, the geodesic pq is normal to $H_{p,k}$ at q.

Proof: In Minkowski Normal coordinates the equation for $H_{p,K}$ is $-t^2 + x^2 + y^2 + z^2 = K$, which is smooth except at the origin when K=0. A smooth hypersurface is spacelike, timelike or null according as its normal vectors are timelike, spacelike, or null at every point.

Now let $q \in H_{p,K}$, and let V be a tangent vector to $H_{p,K}$ at q. Vary q along a curve with tangent vector field V in $H_{p,K}$. Then the geodesics pq form a 1-parameter system (see figure 9). So V belongs to a Jacobi Field vansihing at p. Since on each of the geodesics $g(T,T) = K \;\; (T = exp_p^{-1} q)$, proposition I.16 holds, so g(T,V)=0 all along pq (being evidently zero at p, where V vanishes). Hence pq is perpendicular to $H_{p,K}$at q.

11. **Lemma.** Let N be a simple region, suppose $a, b, c \in \bar{N}$ are such that ab and bc are both future-causal, having distinct directions at b if both are null, or suppose a timelike curve or trip γ existsin $barN$ from a to c. Then ac is future-timelike.

Proof: Consider $\Phi(p) = \Phi(a,p)$, as p travels from a to c along $ab \cup bc$, or along γ, Let T be the tangent vector to the curve. The rate of change of $\Phi(p)$ is given by $\nabla_T \Phi$.

Now, the previous lemma indicates that $\nabla\Phi = g^{ab}\nabla_b\Phi$ must be future-causal (this also follows on direct calculuation of $\nabla\Phi$ from $\Phi(p) = -t^2 + x^2 + y^2 + z^2, p = (t,x,y,z))$, Since T is also future-causal, we have that $\nabla_T\Phi = T^a\nabla_a\Phi$ is negative, unless both vectors are null and proportional (see the next proposition for a proof of this).

Because of the joint at b, p must enter $I^+(a)$ (if it has not already done so). Once in $I^+(a)$ it must remain, since escape would require a velocity faster than light. But then $\nabla_T\Phi$ will be strictly negative after leaving b, so Φ will be a decreasing function from b to c. With $\Phi(p)$ non-decreasing on ab (i.e. in the event T and $\nabla\Phi$ are null and proportional), decreasing on bc, and $\Phi(a,a)=0$, it follows that $\Phi(a,c)<0$. So the geodesic ac is future timelike.

Next, some unfinished business in Lemma 11 will be addressed.

12. **Proposition:** Causal vectors V and W, not proportional if both are null, are in the same causal cone if and only if g(V,W)<0.

Proof: Let V be, say, a future-pointing causal vector, W an arbitrary causal vector. Chooose a basis so that the metric g is Minkowskiian. Let v and w be spacelike vectors in $(\partial/\partial t)^{\perp}$, the subspace of vectors perpendicular to $(\partial/\partial t)$, such that

1. $V = a\frac{\partial}{\partial t} + v\ ; W = b\frac{\partial}{\partial t} + w$

Since V and W are causal, obtain:

2. $g(V,V) = -a^2 + g(v,v) \leq 0 \iff g(v,v) \leq a^2$
3. $g(W,W) = -b^2 + g(w,w) \leq 0 \iff g(w,w) \leq b^2$

Restricted to $(\partial/\partial t)^{\perp}$, g becomes the Euclidean dot product. Hence the usual Schwarz inequality holds, giving:
4. $|g(v,w)| \leq |v||w| \leq |ab|$
5. $g(V,W) = -ab + v\cdot w = -ab + |v||w|\cos\theta$
where *theta* is the angle between v and w. Now, if either $|a|>|v|$ or $|b|>|w|$, or if $\theta \neq 0\ or \pi$, then by 5,
6. $sgn(g(V,W)) = sgn(-ab) = sgn(-b)$

Since $a>0$ by supposition, then $sgn(g(V,W))<0 \Rightarrow b>0 \iff$ W is in the future-causal cone, i.e., the same cone as V. So the theorem follows, modulo the special cases.

Case 1: $|a|=|v|, |b|=|w|, b>0, \theta=0$**.**
b>0 implies W is in the same causal cone as V. $\theta=0$ implies w and v are in the same direction. Under these conditions,

$$V = \left|\frac{a}{b}\right| W$$

, and of course V and W are both mull, which is excluded in the statement of the theorem.

Case 2: $|a|=|v|, |b|=|w|, b<0, \theta=0$**.**

Then $-ab+|v||w|>0$, which is okay since W is in the opposite causal cone.

Case 3.$|a|=|v|, |b|=|w|, b>0, \theta=\pi$**.**
Then W is in the same causal cone as V, and

$$-ab + |v||w|\cos\pi < 0$$

, fulfilling the requirements of the proposition.

Case 4.$|a| = |v|, |b| = |w|, b < 0, \theta = \pi$**.**
Then

$$V = -\left|\frac{a}{b}\right| W$$

, excluded as in case 1.

At last we're ready to prove the other important result of this section (the first being $I^+(p)$ is open).

Note: insert figure in proposition 13

13. Proposition. If $a << b$ and $b < c$, then $a << c$. Similarly, if $a < b$ and $b << c$, then $a << c$. **Proof:** Suppose $a << b$ and $b < c$. Let α be a trip

from a to b. and let γ be a causal trip from b to c. Cover γ with a collection of simple regions, (N_i). Since γ is compact, there is a finite subcollection which also covers γ, say $(N_1, ..., N_r)$. Set $b = x_0$ in N_{i_0}. Choose y_1 in N_{i_0} on the terminal segment of α, $y_1 \neq x_0$. If $c \in N_{i_o}$, then by Lemma 11 $y_1 c$ will be future timelike, so $a << c$ and we're done. Otherwise, choose x_1 to be the future end point of $\bar{N_{i_0}} \cap \gamma$. If $x_1 = c$ then again we're done, since $y_1 x_1$ is future timelike. Otherwise, choose y_2, $y_2 \neq x_1$, in N_{i_0}, on the geodesic $y_1 x_1$, x_2 the endpoint of $\bar{N_{i_0}} \cap \gamma$. If $c \in N_{i_1}$, then again we're done, since $y_2 c$ is timelike, and $a << y_1 << y_2 << c$. If $x_2 = c$, again done. Otherwise, continue in this manner. Since there are only finitely many $N_i's$, the process must terminate, with $a << c$.

13.4 Past, Future, and Achronal Sets

1. Definition F is a **future set** if $F = I^+[S]$ for some $S \subset M$. **P** is a **past set** if $P = I^-[S]$ for some $S \subset M$.

By previous results, it's clear that $F = I^+[F]$ and $P = I^-[P]$. In the following, results for future sets will have dual results for past sets, so only the former will be stated and proved.

2. Proposition: If F is a future set then $\bar{F} = (x \,| I^+(x) \subset F\,)$.

Proof: Let $x \in \bar{F}$. Note that $I^+[\bar{F}] = I^+[F] = F$ by previous results, (I.8.b), so $I^+(x) \in F$.

Conversely, note that if $I^+(x) \subset F$, then $\overline{I^+(x)} \subset \bar{F}$. Since it is always true that $x \in \overline{I^+(x)}$, then $x \in (x \,| I^+(x) \subset F\,)$ implies $x \in \bar{F}$.

The only possible glitch in the above argument might be $I^+(x) = \emptyset$. This possibility is excluded since x is contained in a simple region, homeomorphic to a neighborhood of the origin in $T_x(M)$, which contains future-pointing vectors. So neither $I^+(x)$ nor $I^-(x)$ can be empty for any $x \in M$.

3. Proposition: Let F be a future set. Then:
A. $\bar{F} = \sim I^-[\sim F]$
B. $\partial F = (x \mid I^+(x) \subset F \; and \; x \notin F)$
C. $\partial F = \sim F \cap \sim I^-[\sim F]$
D. $F = I^+[\bar{F}]$

Note: $\sim A = M - A$, the complement of the set A. ∂F is the boundary of F.

Proof: Part D was proven in proposition 2. For A, we need to prove mutual containment.

Let $x \in \bar{F}$. Then by proposition 2 $I^+(x) \subset F$. But then $\sim I^+(x)$ contains $\sim F$, so $I^-[\sim I^+(x)] \supset I^-[\sim F]$ and $\sim I^-[\sim I^+(x)] \subset \sim I^-[\sim F]$. (Taking complements reverses the inclusion symbol).

Case 1: $x \notin I^+(x)$ **and** $x \notin I^-(x)$ Then $x \in \sim I^+(x)$ and $x \notin I^-[\sim I^+(x)]$, so $x \in \sim I^-[\sim I^+(x)]$ hence $x \in \sim I^-[\sim F]$.

Case 2:$x \in I^+(x)$ **and** $x \in I^-(x)$ Then $x \in \sim I^+(x)$. But then $x \notin I^-[\sim I^+(x)]$ for otherwise there would be a trip γ from x to $y \in \sim I^+(x)$, a contradiction, since y cannot be in $I^+(x)$ and $\sim I^+(x)$ at the same time. So $x \notin I^-[\sim I^+(x)]$, which means $x \in \sim [\sim I^+(x)]$, and consequently $x \in \sim I^-[\sim F]$.

Now let $x \in \sim I^-(F)$. Then $x \notin I^-(\sim F)$. This means $I^+(x) \cap \sim F = \emptyset$. $I^+(x)$ is nonempty, hence it must intersect F, i.e. $I^+(x) \subset F$. Then by Proposition 2, $x \in \bar{F}$.

The proofs of C and D are left as exercises.

4. Proposition. Let $Q \subset M$. Then the following are equivalent:

$I^+(Q) \subset Q$
$I^-[\sim Q] \subset \sim Q$
$I^+[Q] \cap I^-[Q] = \emptyset$
$intQ = I^+[Q]$
$\partial Q = (\sim I^+[Q]) \cap (\sim I^-[\sim Q])$

Proof: Exercise. Note that the conditions on Q, in each case, make $a \in Q, b \in \sim Q$ and $a << b$ impossible.

5. Proposition If $I^+[Q] \subset Q$ and Q is open, then Q is a future set.

Proof: By proposition 4, $\int Q = I^+[Q]$, ad since Q is open, $Q = \int Q$, hence $Q = I^+[Q]$ and Q is a future set.

6. Proposition. The union of any system of future sets is a future set. The intersection of any finite system of future sets is a future set.

Proof: Let (F_i) be a system of future sets. Then $\cup_i F_i = \cup_i I^+[F_i] = I^+[\cup_i F_i]$ is immediate, so $\cup_i F_i$ is a future set. ($x \in \cup F_i \iff x \in F_{i0}$ for some $i_o \iff$

$x \in I^+[F_{i_o}] \iff x >> y$ for some $y \in F_{io} \iff x \in I^+[\cup_i F_i]$ since $y \in \cup F_i$.) Let Q, R be future sets. Then $I^+[Q \cap R] \subset conI^+[Q] \cap I^+[R] \subset conQ \cap R$ and by proposition 5, $Q \cap R$ is a future set.

7. Definition A set $S \subset M$ is **achronal** if no two points of S are chronologically related. (i.e. if $x, y \in S$, then neither $x << y$ nor $y << x$ holds.)

8. Examples. Null geodesics are achronal sets, as are light cones and the planes t=0 and t=x in Minkowski space.

A set can be space-like—having only spacelike tangent vectors—and yet still not be an achronal set. For example, take the set $(t, x, y) = (t, 2\cos t, 2\sin t)$ in Minkowski 3-space. The tangent vector field to this spiral is $< 1, -2\sin t, 2\cos t >$ which is spacelike for all t. But if $p = (0, 2, 0) andq = (2\pi, 2, 0 >$ then $p << q$.

9. Definition A set $B \subset M$ is called an **achronal boundary** if $B = \partial I^+[S]$ for some set $S \subset M$. (i.e. the boundary of a future set.)

Note: If A is a set , ∂A is the boundary of A. It is defined as the set of points x such that every neighborhood intersects both A and $\sim A$.

An example would be the light cone enclosing $I^+(p)$. The next proposition shows that the concept is time-symmetric.

10. Proposition. B is an achronal boundary if and only if $B = \partial I^-[T]$;for some $T \subset M$.

Proof: We'll prove both directions of the implication, starting with $\Rightarrow$. Set $T = \sim F$, where F is the future set with $B = \partial F$.

Claim: $B = \partial I^-[T]$. Let $x \in \partial F = B$. Then by proposition 3B $I^+(x) \subset F$ and $x \in \sim F$. Since $I^-(x) \neq \emptyset$, there exists $y \in I^-[\sim F]$ and a trip γ from y to x. Therefore every neighborhood N_x of x contains points of $I^-[\sim F]$. By proposition 3A, we have $\bar{F} = \sim I^-[\sim F]$ So since $x \in \bar{F}$ (boundary points are always in the closure of a set), every neighborhood N_x contains points of $\sim I^-[\sim F]$. This makes x a boundary point of $I^-[\sim F]$, i.e. $x \in \partial I^-[T]$.

Suppose, on the contrary, that $x \in \partial I^-[\sim F]$. Then every neighborhood N_x of x contains points of $I^-[\sim F] = \sim \bar{F}$ and $\sim I^-[\sim F] = \bar{F}$. Hence $x \in \partial \bar{F}$. But $\partial \bar{F} \subset \partial F$ (Proof:Exercise). So $x \in \partial F$.

reverse rightarrow. Suppose $B = \partial I^-[T]$ for some set T. It must be shown that B is also the boundary of a future set, i.e. an achronal boundary. To this end, set $F = \sim \overline{I^-[T]}$.

Claim: F is a future set with $B = \partial F$.

A. F is a future set.

Let $x \in I^+[F]$. Then there is a $z \in F$ with $z << x$, i.e. $z \in \sim \overline{I^-[T]}$ and precedes x. If x were not in F, then $x \in \overline{I^-[T]}$ and $z \in I^-(x) \subset I^-[\overline{I^-[T]}] = I^-[I^-[T]] = I^-[T] \subset \overline{I^-[T]}$, a contradiction, since z is in the complement of this last set. Hence $x \in F$ must hold, and $I^+[F] \subset F$. Further, F is open (it's the complement of a closed set). So by proposition 5 F is a future set.

B. $B = \partial F$. (i.e. $\partial I^-[T] = \partial(\sim \overline{I^-[T]})$. "$\subset$" Let $x \in \partial I^-[T]$. Then every neighborhood N_x contains points of $I^-[T]$ and $\sim I^-[T]$. $I^+(x) \subset \sim \overline{I^-[T]}$, for otherwise there would be a $y \in I^+(x) \cap \overline{I^-[T]}$ with $x << y$ and a sequence in $I^-[T]$ (y_i) with $y_i \to y$, giving $x << y_i << t$ for some $t \in T$, and hence $x \in I^-[T]$, which would preclude it from being a boundary point of $I^-[T]$ (which is open). so $I^+(x) \subset \sim \overline{I^-[T]}$, giving $x \in \partial F$, since in every neighborhood there will be found points of $\sim \overline{I^-[T]}$ and , of course, $\overline{I^-[T]} \supset I^-[T]$.

"$\supset$". If $x \in \partial\left(\sim \overline{I^-[T]}\right)$, then every N_x contains points of $\sim \overline{I^-[T]}$ and $\overline{I^-[T]}$. But if $p \in \overline{I^-[T]}$, then there is a sequence $\{p_i\}$ in $I^-[T]$ converging to p (or $p \in I^-[T]$), and so N_x will contain elements of $I^-[T]$. Furthermore, $\sim I^-[T] \supset \sim \overline{I^-[T]}$. Hence every neighborhood intersects both $I^-[T]$ and $\sim I^-[T]$, so $x \in \partial I^-[T]$, and $B = \partial F$.

11. Proposition. If $B(\neq \emptyset)$ is an achronal boundary, then there is a unique past set P and a unique future set F such that F,P, and B are disjoint with $M = P \cup B \cup F$. Furthermore, any trip or timelike curve from a point of P to a point of F must meet B in a unique point.

Proof: By the previous proposition, a future set F and a past set P exist satisfying $B = \partial F = \partial P$. With $F = \sim \overline{I^-[T]}$ as in that construction, it follows from proposition 3 that $\bar{F} = \sim I^-[\sim F] = \sim I^-[\overline{I^-[T]}] = \sim I^-[I^-[T]] = \sim I^-[T]$, making clear that $T = \sim F$. Then $P = I^-[\sim F]$, and again by proposition 3 $\sim P = \sim I^-[\sim F] = \overline{F}$, showing that $\bar{F} \cup P = F \cup B \cup P = M$.

If γ is a trip from P to F, then $\gamma \cap \sim F$ and $\gamma \cap \sim P$ are closed sets, containing between them containing all of γ. Thus there is overlap, which evidently must be in B. ($F \cup B = B^c$, so $\sim F \cap \sim P = B$). But B is achronal, so the intersection with γ must be a single point.

There remains the question of the uniqueness of the decomposition.

Let $P' = I^-[P'], F' = I^+[F']$ be sets such that $M = P' \cup B \cup F'$. Since $M = P \cup B \cup F$ also, it follows that either $P \cap F'$ or $F \cap P'$ is non-empty, say the former.

Any two points in M may be connected, by a curve γ. Let $x \in P \cap F'$. If γ leaves P, it must enter $\sim P = B \cup F$. But since x was also in F', γ must also enter $\sim F' = B \cup P'$. Since exiting any future or past set entails crossing a boundary, it follows that γ must be contained in $B \cup (P' \cap F) \cup (F' \cap P)$. M is connected, hence arc-wise connected, so from x one can draw a path to any other $p \in M$. But then:

$$M = B \cup (P' \cap F) \cup (F' \cap P)$$

$$\emptyset = \sim B \cap (\sim P' \cup \sim F) \cap (\sim F' \cup \sim P)$$

$$\emptyset = (F \cup P) \cap (F' \cup B \cup P) \cap (P' \cup B \cup F)$$

$$\emptyset = (F' \cup B \cup P) \cap [(F \cup P) \cap (P' \cup B \cup F)]$$
$$= (F' \cup B \cup P) \cap [((F \cup P) \cap P') \cup ((F \cup P) \cap B) \cup ((F \cup P) \cap F)]$$

$$= (F' \cup B \cup P) \cap [((F \cup P) \cap P') \cup \emptyset \cup F]$$

$$= [(F' \cup B \cup P) \cap F] \cup [(F' \cup B \cup P) \cap (F \cap P' \cup P \cap P')]$$
$$= (F' \cap F) \cup [F' \cap (F \cap P')] \cup [B \cap (F \cap P')] \cup [P \cap (F \cap P')] \cup [F' \cap (P \cap P')] \cup [B \cap (P \cap P')] \cup [P \cap (P \cap P')]$$

$$\emptyset = (F' \cap F) \cup \emptyset \cup \emptyset \cup \emptyset \cup \emptyset \cup \emptyset \cup (P \cap P')$$

The final result is, at last,

$$\emptyset = (F' \cap F) \cup (P \cap P')$$

This last equation holds true only if $F' \cap F = \emptyset$, and $P' \cap P = \emptyset$. But then $F' \subseteq P$ and $P' \subseteq F$. Since $M = F' \cup P' \cup B$, it follows that $F = P'$ (and $P = F'$). But then F is both a future and a past set, i.e. $I^-[F] = I^+(F) = F$. By exercise 26 (which is "true"), this implies F=M, a contradiction. It follows that $F' = F$ and $P' = P$, and uniqueness is proved.

In Minkowski space, $F = I^+[B]$ and $P = I^-[B]$ gives the decomposition, but mutilate the space and this no longer holds true, as in the follow two figures.

Proposition. Any achronal boundary is a topological 3-manifold.

Proof: Let P and F be as in the preceding proposition, and let $a \in B$. Let N be a simple region containing a, and choose Minkowski Normal coordinates for N. We would like curves with (x,y,z)= constant, "parallel" to the "t-axis" in N to be timelike curves. This isn't automatically guaranteed, since the metric fluctuates from point to point. Since exp^{-1} is continuous in its argumnents, one will always be able to find a neighborhood where this property holds. Hence let $Q = \{p \in N \mid |t| \le \rho; x^2 + y^2 + z^2 < \rho^2\}$ with rho chosen sufficiently small. Label the timelike curves $\eta_{x,y,z}$ for (x,y,z)= constant. Each such curve stretches from $(-\rho, x, y, z)$ in $I^-(a)$ to (ρ, x, y, z) in $I^+(a)$, and since B is achronal, must meet B at a single point, denoted b(x,y,z). b(x,y,z) is therefore a 1-1 mapping between $B \cap Q$ and an open ball in R^3 of radius ρ. Since we can obviously cover B with such coordinate neighborhoods, B will be a 3-manifold, provided it can be shown b(x,y,z) is continuous. But this follows from the achronality of B.

$$\lim_{\Delta x, \Delta y, \Delta z} = b(x_0, y_0, z_0)$$

necessarily, as achronality implies the difference in the time coordinate,

$$\Delta t^2 \le \Delta x^2 + \Delta y^2 + \Delta z^2$$

else the points $b(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z)$ and $b(x_0, y_0, z_0)$ would be chronologically related. So $\Delta t \to 0$ necessarily as $\Delta x, \Delta y, \Delta z \to 0$, showing that b is continuous.

Null hypersurfaces such as $E^+(a) = J^+(a) - I^+(a)$, γ a timelike curve, turn out to be achronal boundaries of particular interest. In Minkowski 3-space take η to be the closed, spacelike curve given by $t = 0 = x^2 + y^2 - 1$. (See figure 16). Note that every point of $B = \partial I^+[\eta]$ which is not on η is the future endpoint of some null geodesic in B. This property holds true in general.

13. Lemma Let F be a future set with $B = \partial F$. Let $x \in B$ and suppose an open set $Q \ni x$ exists such that:

(a) for any $y \in Q \cap F$ there is a trip γ from a point $z \in F - Q$ to y or, equivalently,

(b) $F = I^+[F - Q]$

Then B contains a null geodesic with future endpoint x.

Note that the condition on Q, $F = I^+[F - Q]$, excludes the possibility $J^+(x) \cap \overline{F} = \emptyset$, in which case the lemma couldn't hold.

Proof: The equivalence of (a) and (b) will be left as an exercise.

Let N be a simple region, $x \in N \cap F \subset Q, x \in B$. Let $\{y_i\}$ be a sequence in $N \cap F$ converging to x. By (a), each y_i is the future endpoint of a trip γ_i from some $x_i \in F - Q \subset F - N$. Let $v_i \in F \cap \partial N$ be the past endpoint of the connected component of $\gamma_i \cap \overline{N}$ which terminates at y_i. By Lemma I.11 the geodesics $v_i y_i$ are timelike. ∂N is compact, so $v_{i_j} \to v \in \partial N$. ($v_{i_j}$ is some infinite subsequence of the v_i). v_x must be timelike or null, since each $v_{i_j} x_{i_j}$ is timelike, and the world function is continuous, ($\Phi(v_i, x_i) < 0$ and $\Phi(v_i, x_i) \to \Phi(v, x)$).

But v_x cannot be timelike. $v \in \overline{F}$, and by proposition 2 $I^+(v) \subset F$, which would imply $x \in I^+(v) \subset F$, a contradiction, since $x \in B$ (disjoint from F by proposition 11.) Hence v_x is null, and by the same reasoning no point of v_x can reside in F. Further, each point $u \in v_x$ must be a limit point of F. (To see this rigorously, choose a smaller simple region N' containing x, with u on $\partial N'$). Hence v_x must like entirely in B.

14. Theorem. Let $S \subset M$ and set $B = \partial I^+[S]$. Then if $x \in B - \overline{S}$ there exists a null geodesic $\eta \subset B$ with future endpoint x and which is either past-endless or else has a past endpoint on $\overline{S}$.

Proof: Figure 16 is a good illustration for this theorem, where all null geodesics on B have past endpoints on $\overline{S}$.

$\overline{S}$ is closed and $x \notin \overline{S}$, so there an open set $Q \ni x$ and an open set $V \supset \overline{S}$ such that $V \cap Q = \emptyset$. $(I^+[S] \cap V) \cap \gamma \neq \emptyset$ for any trip γ with initial point on S, so clearly the condition $F = I^+[F - Q]$ of the previous lemma is satisfied. Thus there is a null geodesic on B with future endpoint x. Let η be the maximal extension into the past of this geodesic. If η is not past-endless (recall that past-endless means absence of a past endpoint, not necessarily that the curve extends indefinitely into the past) then it has past endpoint y, which is in B since B is closed. If $h \notin \overline{S}$, then apply the previous lemma again, obtaining another null geodesic, say γ, with future endpoint y, and with a distinct direction from η, since η was maximally extended into the past. But then by Lemma I.11 points of γ chronologically precede points of η, a contradiction since B is achronal. Hence either $y \in \overline{S}$ or else η is past-endless.

Figure 16 illustrates another general feature of achronal boundaries. Note that at the point p at the apex of the cone, different null geodesics intersect, and subsequently must exit B and enter $I^+[S]$. Thus we have the next proposition.

15. Proposition. Let $B = \partial I^+[S]$. Suppose $x \in B - \overline{S}$ is an endpoint of two null geodesics on B. Then:

(a) if x is a past endpoint of one or both geodesics, then their union is a null geodesic on B;

(b) If x is a future endpoint of both geodesics, then unless one is contained in the other, every extension of either geodesic into the future beyond x must leave B and enter $I^+[S]$.

Proof: (a) If x is the past endpoint of one geodesic and the future endpoint of another geodesic, then since B is achronal, the two must constitute a single geodesic, else Lemma 1.11 gives a contradiction to achronality. On the other hand, if x is the past endpoint of both geodesics, then by Lemma 13 there must exist another geodesic on B having x as its future endpoint. All three geodesics must constitute a single geodesic, else again the achronality of B is violated.

(b) Extend one of the geodesics into the future beyond x. If the geodesic remains on B, then achronality is violated unless one geodesic contains the other. Hence no extension can lie on B, except in this case. Now suppose the geodesics are distinct. Extend one of them, and call it η, the other *gamma*. Let $y \in \eta$ on the extension beyond x, and let $z \in \gamma$ preceding x. Lemma 1.11 gives $x << y$. $z \in \partial I^+[S]$, and by proposition 2, $y \in I^+[S]$ since z is in the closure of a future set.

13.5 Global Causality Conditions

In this section various causal conditions will be introduced, with theorems leading to the establishing of a topology based on causal structure, called the Alexandrov Topology after the person who first suggested it.

The grossest causality violation is considered closed timelike curves or trips.

There is a spectrum of other causality conditions, such as closed causal curves, and the following:

1. Definition. A spacetime M is **future distinguishing** at $p \in M$ if and only if $I^+(p) \neq I^+(q)$ for each $q \in M$ with $q \neq p$. **M** is **future distinguishing** if it is future distinguishing at each of its points. Similar definitions hold for **past distinguishing**.

Clearly no spacetime which has a closed trip or causal trip can be future- or past-distinguishing. Figure 18 gives an example of a spacetime which has no closed causal trip, yet is neither future- nor past-distinguishing.

Take Minkowski space with $|x| \leq 1$ and identify (t,1) and (t,-2). Let $ds^2 = dtdx + t^2dx^2$. The light cones are tangent to the null geodesics obtained by setting $ds^2 = 0 = dx(dt + t^2dx) \Rightarrow x = constant$ and $t = 1/(x+c)$. Now remove the origin. The spacetime is causal, but not distinguishing.

An equivalent definition is the following: M is future-distinguishing at p if an only if for every causal trip γ initiating at p there is a neighborhood of p which γ intersects at most once. The proof will be left as an exercise. This alternate definition makes clear the difference between causal and distinguishing spacetimes. In a distinguishing spacetime, causal curves come close to being closed, but never actually close. There will always be a neighborhood (albeit possibly very small) to which the departing trip cannot return.

2. Definition. An open set $Q \subset M$ is causally convex if and only if Q intersects no trip in a disconnected set.

The definition is equivalent to: for every $x, y \in Q, x << z << y \Rightarrow z \in Q$, which is reminiscent of the usual definition of convex set (i.e. for any $x, y \in Q$, the straight line connecting x and y is contained in Q). (Figure 19).

3.Definition. M is **strongly causal** at p if and only if p has arbitrarily small causally convex neighborhoods. M is strongly causal if and only if it is strongly causal at each point.

When strong causality is violated causal trips can leave the spacetime and later return, though an actual closed causal trip need not occur. Figure 20 gives an example of a spacetime which is causal but not strongly causal.

Minkowski two-space with $|t| \leq 1$, (1.x) and (-1,x) identified, and two half-infinte lines removed. Strong causality fails along the dotted null geodesic.

4. Definition. Let Q be an open subset of M and $x, y \in Q$. Write $x <<_Q y$ if a trip from x to y exists which is completely contained in Q. A similar definition holds for $x \prec_Q y$.

Since Q is open it is a spacetime manifold in its own right, so propositions proven for $<<$ and $\prec$ hold for $<<_Q$ and $\prec_Q$.

5. Defintion. $< x, y >_Q = \{x | s <<_Q z <<_Q y\}$. $< x, Y > = < x, y >_M$.

In particular, $< x, y >= I^+(x) \cap I^-(y)$ which is open since both $I^+(x)$ and $I^-(y)$ are. sets of this form will constitute a base for a topology on M, as we will see.

6. Proposition. Let N be a simple region and $x, y \in N$. Then the set $< x, y >_N$ has the property that no causal trip lying in N can intersect $< x. >_N$ is a disconnected set.

Proof: Let $\gamma \subset N$ be a causal trip intersecting $< x.y >_N$, and let u, v be in the intersection with $u \prec v$. (We assume $x << y$, else the proposition holds vacuously.) Let $w \in \gamma$, with $u \prec w \prec v$. Then $x << u \prec w \Rightarrow x << w$. and $w \prec v \prec y \Rightarrow w << y$. Furthermore, xw and wy are timelike geodesics which exist and lie entierly in side N, since N is a simple region. But this means $w \in < x, y >_N$. Inasmuch as u,v and w were chosen arbitrarily, it follows that $\gamma \cap < x, y >_N$ must be connected.

7. Proposition. If N is a simple region, Q an open set contained in N, and $p \in Q$, then there exist points $u.v \in Q$ such that $p \in < u, v >_N \subset Q$.

Proof: Choose Minkowski Normal Coordinates for N at p, and $\epsilon > 0$ so that the ball B, $t^2 + x^2 + y^2 + z^2 \leq \epsilon$ is entirely contained in Q. Make ϵ small enough so that any timelike curves in B are also timelike with respect to the flattened Minkowski metric $ds^2 = -4dt^2 + dx^2 + dy^2 + dz^2$ (this is possible because the manifold is Lorentizian). This ensures that the relevant portions of $I^-(v)$ and $I^+(u)$ and their boundaries will be contained entirely inside B. (see figure 21). Let $u = (-\epsilon/2, 0, 0, 0)$ and $v = (+\epsilon/2, 0, 0, 0)$.

By our choice of ϵ we may be confident that the actual timecones of u and v are contained in the modified cones, which themselves are in B. Now let $w \in < u, v >_N$ and let γ be the timelike geodesic uw. By construction, if γ reaches ∂B, then γ must have intersected the past nullcone of v. If q is the point of intersection, then $qwq \cup wv$ is a trip, which makes qv a timelike geodesic by proposition 2.11. But this is a contradiction, since qv is null. Hence $w \in < u, v >_N \subset N \subset B \subset Q$.

8. Proposition. Any simple region must be strongly causal.

Proof: Proposition 6 says $< x.y >_M$ is causally convex, and proposition 7 says that each point of N has arbitrarily small causally convex neighborhoods. Hence N is strongly causal.

9. Definition. A **local causality neighborhood** is a causally convex open set whose closure is contained in a simple region.

10. Proposition. M is strongly causal at p if and only if p is contained in some LCN.

Proof: If M is strongly causal at p then there exist arbitrarily small causally convex neighborhoods at p. Choose a simple region N containing p, an open set Q containing p inside N, and finally a causally convex set containing p inside Q. The closure of this last set will be inside N, hence will be a local causality neighborhood.

Conversely, if p belongs to an LCN then, by proposition 7, there exist arbitrarily small sets $< u, v >_N$ contained in the LCN which contain p. If a trip γ intersected $< u, v >_N$ in a disconnected set, then by proposition 6 γ cannot be contained enitrely in N. Thus it must leave and re-enter N, and so leave and re-enter the LCN, a contradiction. So $< u, v >_N$ is causally convex, and M is strongly causal at p.

By the above proposition, the collection of points at which M is strongly causal constitutes an open set.

11. Proposition. No local causality neighborhood can contain a future or past endless causal trip.

Proof: Exercise.

12. Lemma. Let $p \in M$. Then strong causality fails at p if and only if there exists $q \prec p, q \neq p$, such that:$x << p$ and $q << y$ together imply $x << y$ for all x,y.

Proof: Suppose strong causality fails at p. Let N be a simple region containing p, and let $Q_i =< u_i, v_i >_N$, a nested sequence of neighborhoods of p convergin on p, i.e. $Q_1 \supset Q_2 \supset Q_3 ..., \cap Q_i = p$, $\overline{Q_i} \subset N$. Then each Q_i fails to be causally convex, else it would be an LCN, violating proposition 10. Let γ_i intersect Q_i in a disconnected set. By proposition 6 γ_i cannot be completely contained in N. Let γ_i have past enpoint a_i in Q_i, exit N at $b_i \in \partial N \cap \gamma_i$, and re-enter N at $c_i \in \partial N$, terminating at $d_i \in Q_i$. ∂N is compact, so there is an accumulation point c of the $\{c_i\}$. $c_i d_i$ is future timelike for all i, and since $d_i \to p$, cp must be future causal. Choose $q \in cp$, $q \neq c$ and $q \neq p$. Suppose $x << p$ and $q << y$. (see figure 22.

$p \in I^+(x)$ and $I^+(x)$ is open, so $Q_i \subset I^+(x)$ for large enough i. But then $a_i \in I^+(x)$. $c \prec q << y$ gives $c \in I^-(y)$. $I^-(y)$ is open and so contains infinitely many c_i. So for large enough i, $x << a_i << b_i << c_i << y \Rightarrow x << y$ as required.

Conversely, assume $q \prec p, x << p, and q << y \Rightarrow x << y$ for all x,y. Let $P \ni p, Q \ni q$ be disjoint open sets. Then P cannot be causally convex (see figure 23).

If $z \in I^+(p)$, then $q < p << z$ gives $z \in I^+(q)$. Choose y on the trip from q to z. Then $x << y << z$ gives a trip which intersects P in a disconnected set. So strong causality fails at p.

13. Proposition. If M is stronly causal at P then M is distinguishing at p (future and past).

Proof: Exercise.

14. Stable causality, whereby a spacetime remains causal under arbitrarily small perturbations of the metric, is yet a stronger condition than strongly causal. Fig 24 gives an examples of a spacetime which is strongly causal but not not stably causal.

Minkowski space, with removed half-line and identification. a perturbation of g widening the light cone woudl result in closed trips. Otherwise, strong causality prevails.

Sets of the form $< u, v >$ cover M. If $p \in Q \cap R$, Q and R of this form, and if for any such p there exists a set $S =< w, z >$ such that $p \in S \subset Q \cap R$, then a theorem from general topology says that the collection of all such sets forms a base for a topology on M. Recall that the collection of open n-balls forms a base for a topology in R^n, or for that matter, Minkowski space. The diamond-shaped sets $< u, v >$ form an equivalent topology, called the **Alexandrov Topology**. Hence, we have the following propositions and theorem.

15. Proposition. If $p << q$ and $p << r$ then there exists a point w such that $w << qw << r$, and $p << w$.

Proof: Exercise.

16. Proposition. If $x, p, q, r, s \in M$ are such that $x \in< p, q > \cap < r, s >$, then there exist $u, v \in M$ such that $x \in< u, v >\subset< p, q > \cap < r, s >$

Proof: Exercise.

17. Theorem. The following three restrictions on a a spacetime M are equivalent:

(a) M is strongly causal;

(b) the Alexandrov topology agrees with the manifold topology;

(c) the Alexandrov topology is Hausdorff.

Proof: $(a) \Rightarrow (b)$. Suppose strong causality holds at p. Let $P \ni p$ be open in the manifold topology. It must be shown that an Alexandrov neighborhood containing p exists in P. Let N be a simple region in P containing p and $Q \ni p$, a causally convex open set contained in N (which exists by propositions 7 and 10). Also, by proposition 7, there exists a pair of points u, v in Q such that $p \in< u, v >_N \subset Q$. But $< u, v >_N =< u, v >$, since Q is an LCN. (otherwise, some trip could leave and re-enter N, hence leave and re-enter Q, a constradiction). Hence $p \in< u, v >\subset Q \subset P$, as required, showing that sets open in the manifold topology are also open in the Alexandrov Topology. Since sets of the form $< u, v >$ area already open in the manifold topology (as proved in proposition 2.13, together with the fact that finite intersections of open sets are open $\rightarrow I^+(u) \cap I^-(v) =< u, v >$.) Therefore we have equivalent topologies.

$(b) \Rightarrow (c)$. Immediate, since the manifold topology was assumed to be Hausdorff.

$(c) \Rightarrow (a)$. Suppose by way of contradiction that strong causality fails at p. Let $q \prec p$ as in Lemma 12, with $p \in< x.u >$ and $q \in< v, w >$. $q \prec p << u$, so $q \in I^(u)$. Choose y just to the future of q, so $y \in I^-(u)$, and $y \in< v, w >$. By Lemma 12, $x << y$ so $y \in< x, u >$ also. Since $< x, u >$ and $< v, w >$ were arbitrary Alexandrov neighborhoods of p and q, and $< x, u > \cap < v, w > \neq \emptyset$, the Hausdorff property fails for p and q.

18. Proposition. If M is compact it must contain closed trips.

Proof: Cover M with Alexandrov Neighborhhoods. Since M is compact, this cover has a finite subcover, $\{< x_i, y_i >\}, i = 1, 2, ...n$. If x_i, say, is in $< x_i, y_1 >$, then $x_1 << x_1$ and we're done. So suppose $x \in< x_{i_1}, y_{i_1} >$ for some $i_1 \neq 1$. x_{i_1} must be contained in some element of the cover, say in $< x_{i_2}, y_{i_2} >$. Obtain an infinitsequence in this manner, with $... << x_{i_3} << x_{i_2} << x_{i_1} << x_1$. But there are only **finitely many** x_i's! This means the sequence must repeat itself, i.e. $x_{i_k} << x_{i_k}$ for some index, giving a closed trip.

Such a point x_{i_k} is called **vicious**. If V is the set of all vicious points in M, then clearly $V = \cup_{x \in M} < x, x >$, hence the set of points is open. In fact, V is the disjoint union of sets of the form $< x, x >$.

19. Proposition. If $< x, x > \cap < y, y > \neq \emptyset$, then $< x, x >=< y, y >$.

Proof: Let $z \in< x, x >, w \in< x.x > \cap < y, y >$. Then $z << x << w << y << w << x << z$, so $z \in< y, y >$ and $< x, x > \subset < y, y >$. Similarly $< y, y > \subset < x, x >$, so $< x, x >=< y, y >$.

Our last proposition in this section will be needed later for technical reasons. Penrose proves a lot more in his book, Techniques of Differential Topology in Relativity, much more than we will need, hence my abridgement.

20. Proposition. Suppose strong causality fails at p. Then either $p \in \overline{V}$ or p lies on an endless null geodesic γ, at every point of which strong causality fails, and such that if u and v are any two points of γ with $u \prec v$, $u \neq b$, then $u << x$ and $y << v$ together imply $y << x$.

Proof: Let Q_i be a nested sequence of LCN's as in Lemma 12, converging on p, and a_i, b_i, c_i, and d_i as in figure 22. $a_i << b_i$ for each i,$a_i \to p$ and $b_i \to b$, so pb must be causal, and similarly cp. There are 5 cases to consider.

Case 1: cp and pb are both timelike. then $p << b_i << c_i << p$, giving and $p \in V$.

Case 2: pb is timelike, cp null. Let x_i be a sequence of points on pb, $x_i \to p$. Then by Lemma 1.11 cx_i is timelike. $I^-(x_i)$ contains an infinite number of $c_i's$, while $I^+(x_i)$ contains an infinite number of $b_i's$. Choose i large enough so that $c_i << x_i << b_i << c_i$, relabelling the indices if necessary. Obtain in this way a sequence of closed trips through $x_i's$, and since $x_i \to p$, $p \in \overline{V}$.

Case 3: pb is null, cp timelike. This gives $p \in \overline{V}$ as in case 2.

Case 4: pb and cp are both null, with different directions at p. Choose $\{x_i\} \to p$ on pb as before, $\{y_i\} \to p$ on cp. By Lemma 1.11, $y_i x_i$ is timelike for each i. Pick $z \in y_i x_i$ for each i. Clearly $z_i \to p$ and $c << z_i << b$. As in case 2, it is possible to show that each z_i is in V, so again $p \in \overline{V}$.

Case 5: pb, cp are both mull, with the same direction a p. Then $c \prec b$, and cb is a single null geodesic. Choose $r \in cb, r \neq c$. If $x << r$ and $c << y$, then $x << b$ so $I^+(x)$ contains arbitrarily large numbers of the b_i, and $I^-(y)$ the c_i. Then $x << b_i << c_i << y$ for large enough i, and $x << y$, so by Lemma 12 strong causality fails at r. The time-reverse of Lemma 12 gives strong causality failure at c.

Now let N_1 be a simple region containing b, and let b' be an accumulation point of the b_i', the intersection of ∂N_1 with γ_i to the future of each b_i in N_1. If $b << b'$, or bb' is null with different direction than pb, then as in previous cases it can be obtained that $p \in \overline{V}$. So pb' may be taken as a single null geodesic γ. Obvioulsy strong causality fails along bb', by the time-reverse of Lemma 12. Continuing in this way both into the future and past, an endless null geodesic γ is obtained at every point of which strong causality fails.

Finally, let $u, v \in \gamma, u \neq v, u \prec v$, and let $u << x$ and $y << v$. By construction, $c \prec u << x$ so $c << x$ and $y << b^{(i)}$ for some i. But then $y << b^{(i)} << c_j <, x$ gives $y << x$ for a large enough j, as required.

Note that the proposition leaves open the possibility that p satisfies both conditions, i.e. that it lies on a specified null geodesic and is in $\overline{V}$. See figure 22 and Figure 25 (below).

At p strong causality fails, p is on endless null geodesic fulfilling proposition 20 and $p \in \overline{V}$. Closed trips approach p, for example abca.

13.6 Domains of Dependence

1. Definition: Let S be achronal. Define the future, past, and total domains of dependence, respectively, by:

$D^+(S) = \{x \mid$ every past-endless trip containing x meets S$\}$

$D^-(S) = \{x \mid$ every future-endless trip containing x meets S$\}$

$D^+(S) = \{x \mid$ every endless trip containing x meets S$\}$

Given initial data on an achronal set S, D(S) is that region of spacetime determined by that data. Hawking and Ellis use causal curves instead of trips, but the difference is negligible.

2. Definition. The future, past, or total Cauchy Horizon of a closed achronal set S is, respectively:

$H^+(S) = \{x \mid x \in D^+(S) \text{but} I^+(x) \cap D^+(S) = \emptyset\}$

$H^-(S) = \{x \mid x \in D^-(S) \text{but} I^-(x) \cap D^-(S) = \emptyset\}$

$H(S) = \{x \mid H^+(X) \cup H^{(S)\}}$

Figure 26.

3. Proposition. Let $S \subset M$ be achronal and closed. Then:

(A) $D^+(S)$ is closed.

(B) $H^+(S)$ is achronal and closed.

(C) $S \subset D^+(S)$

(D) $x \in D^+(S) \Rightarrow I^-(x) \cap J^+(S) \subset D^+(S)$

(E) $\partial D^+(S) = H^+(S) \cup S$

(F) $\partial D(S) = H(S)$

(G) $I^+[H^+[S]] = I^+(S) - D^+(S)$

(H) $intD^+(S) = I^+(S) \cap I^-(D^+(S))$

Proof: Exercise. Which of (A) through (H) do not require S to be closed? Find a corrected version of each in terms of "edge".

4. Definition. Let S be achronal. Then

$edge(S) = \{x \mid$ every neighborhood Q of x contains points y,z and two trips from y to z, just one of whic

Edge(S) is the set of limit points of S not in S, together with the set of points in the vicinity of which S fails to be a topological 3-manifold. In fact, if $p \ni edge(S)$, then there is a connected open set Q containing p such that $S \cap Q$ is an achronal boundary the spacetime Q, and conversely. This results in:

5. Proposition. Achronal boundaries are edgeless. If S is achronal, edge(S) is closed.

Proof: Exercise.

6. Proposition. (a) Let $x \in I^+(edge(S)) \cap D^+(S)$. Then there is a trip γ from $s \in edge(S)$ to x. Choose a neighborhood Q containing s such that $Q \subset I^-(x)$. By definition of edge(S), there exist y,z in Q and a trip μ from y to z, which misses S. But $z \in I^-(x)$, so there exists a trip into the past of x which misses S, which says that $x \ni D^+(S)$, a contradiction.

(b) Let $x \in edge(S)$, Q^{open} a neighborhood of x. By definition, there exists a pair y,z in Q and trips μ, γ from y to z, μ missing S and γ hitting S. Since $\partial D^+(S) = H^+(S) \cup S$, γ must intersect a boundary point b in reaching z (since $z \in \sim D^+(S)$). If $b \in H^+(S)$, we're satisfied. If $b \in S$ and $I^+(b) \cap D^+(S) = \emptyset$, then $b \in H^+(S)$. So suppose $p \in I^+(b) \cap D^+(S)$, close enough to b so that $p \in I^{(z)}$. Let γ' be the trip from z back to p, to b, then y. To reach p, γ' must intersect $\partial D^+(S)$, say at b'. But $b' \notin S$, since then $b << b'$, Contradicting the achronality of S. Hence $b' \in H^+(S)$.

Next, if μ hits $H^+(S)$ at, say, k, then μ cannot miss S, since $k \in D^+(S)$. Therefore μ misses $H^+(S)$, γ' hits $H^+(S)$, and and Q was arbitrary; so $edgeS \subset edgeH^+(S)$.

Now for the reverse inclusion. Let $x \in edgeH^+(S)$, Q y,z, μ, and γ as before, μ missing $H^+(S)$ γ hitting $H^+(S)$. Since γ hits $H^+(S)$, it must also meet S. Finally, suppose μ hit S at b. If $I^+(b) \cap D^+9S) = \emptyset$, then $b \in H^+(S)$, a contradiction. But then $I^+(b) \cap D^+(S) \neq \emptyset$, and as before obtain $p \in D^+(S)$ and a trip μ' which must intersect $H^+(S)$, again a contradiction. So μ misses S, γ hits S, and $edgeH^+(S) \subset edgeS$.

6.5. Theorem. Let S be achronal. Then every point of $H^+(S) - edgeS$ is the future endpoint of a null geodesic γ on $H^+(S)$ which is either pastendless or else has past endpoint on edge S.

Proof: Let $p \in H^+(S) - edgeS$. Then (a) $p \in I^+(S)$ or (2) $p \in S - edgeS$.

Case 1: Let $p \in I^+[S] \cap (H^+(S) - edgeS)$. Choose a simple region Q containing p such that $\overline{Q} \cap \partial I^+[S] = \emptyset$. This can be done by simply choosing $\overline{Q}$ inside $I^+[\beta]$ for a trip issuing from S and passing through p. Now let $F = I^+[S] - D^+[S] = I^+[H^+[S]]$ (prove it!) If $y \in F$, then a trip α going into the past from y must miss S. α encounters ∂Q, and subsequently $\partial I^+[S]$. Since the boundary of $I^+[S]$ was disjoint from ∂Q, α just enter $I^+[S] - Q$. But $z = \partial Q \cap \alpha$ is in F since α misses S, so $z << y, z \in F - Q$, satisfying the conditions of Theorem 3.13, giving a null geodesic γ with future endpoint p in $H^+(S)$. ($H^+(S)$ is clearly part of the achronal boundary of F). At any past endpoint on γ, simply repeat the argument, until γ is either past-endless or hits edge S. There can be no "Joint", else achronality is violated.

Case 2: If $p \in S - edgeS$, then there exists a simple region Q containing p such that every curve from a point of $I^+(p)$ to $I^-(S)$ in Q must hit S. Take a sequence of points q_n in $I^+(p)$ converging to p and repeat the argument in Theorem 3.13, or argue as above.

7.Proposition. If S is achronal and $x \in D^+(S) - H^+(S)$, then every past-endless causal trip with future endpoint x must intersect $S - H^+(S) - edgeS$ and must contain a point of $I^{[S]}$.

Proof: If $x \in S$ we are done. So suppose $x \in intD^+(S) = D^+(S) - H^+(S) - S$.

Let γ be a past-endless causal trip with future endpoint x. Cover γ with a locallly finite system of simple regions $N_a, N_2, N_3....$ Choose $y_1 \in D^+(S) \cap I^+(x) \cap \overline{N_{i_1}}$, where N_{i_1} is a simple region containing $x = x_1$. Let x_2 be the past endpoint of $\gamma \cap \overline{N_{i_1}}$. $x_2 \in \partial N_{i_1}$, $x_2 \prec x_1 << y_1$ so $x_2 << y_1$. Now $x_2 \in N_{i_2}$ for some i_2. Choose $y_2 \in I^+(x_2) \cap I^-(y_i) \cap N_{i_2}$. Let x_3 be the past endpoint of the connected component (as before) of $\gamma \cap N_{i_2}$, connecting $x_3 \in \partial N_{i_2}$ to x$_2$.$Choose$y$_3$ as before, in $N_{i_3} \cap I^+(x_3) \cap I^-(y_2)$. Continue indefinitely (see figure 27), obtaining a sequence y_i, with $.... << y_3 << y_2 << y_1; y_1 \in D^+(S)$, with $... \cup y_3y_2 \cup y_2y_1$ future timelike. Since the $N_i's$ are locally finite, they can't accumulate. Furthermore, no segment or part of γ can enter or leave any N_i more than a finite number of times, else a past-endless trip not intersecting S would be produced.m So the $x_i's$ must proceed indefinitely into the past on γ, at least until exiting $D^+(S)$, and $... \cup y_4y_3 \cup y_3y_{22}y_1$ constitutes a past-endless trip η which intersects S, since $y_1 \in D^+(S)$. Say η meets S at z on the segment y_ky_{k-1}. $x_k << z$, so $x_kD^+(S)$. Thus some point w of γ lies on $\partial D^+(S)$. If $w \in H^+(S)$, then by 3(g) $y_1D^+(S)$, a contradiction. Similarly, if $w \in edgeS$, then by proposition 6 we get the same contradiction. So $w \in S$, as required, but not in either $H^+(S)$ or $edgeS$. Furthermore, $x_k \in I^-(S)$.

8. Proposition. If $S \subset M$ is acrhonal and $p \in intD^+(S)$, then M is strongly causal at p.

Proof: Suppose first that some point $x \in D^+(S)$ lies on a cloased trip η. Such a trip is past-endless and so meets S at, say, w. But then $w << w$ contradicting the achronality of S. So $D^+(S) \cap V = \emptyset$, and $intD^+(S) \cap \overline{V} = \emptyset$. Now suppose strong causality fails at at some point $p \in intD^+(S)$. By proposition 4.20 there must be an endless null geodesic γ through p with the property that if $q \in \gamma$ and $q \prec p, q$, then every $y \in I^+(q)$ and $x \in I^-(p)$ must satisfy $x << y$. By the previous proposition, γ contains some point $q \in I^-(S)$. Let $y \in I^+(q) \cap I^-(S)$ and $x \in I^-(p) \cap intD^+[S]$. Then $x << y$. B?ut there exists $s_1, s_2 \in S$ such that $s_1 << x << y << s_2$, by construction, violating the achronality of S. Hence strong causality must hold on $D^+(S)$.

However, examples can be constructed where strong causality fails on S or $H^+(S)$ (i.e. on $\partial D^+(S)$.

9. Proposition. If S is achronal and $x \in intD^+(S)$, then $J^-(x) \cap J^+(S)$ is compact.

Proof: Let $\{N_i\}$ be a locally finite covering of $J^+(u) \cap J^-(v)$, and suppose by way of contradiction that $J^+(u) \cap J^-(v)$ is not compact. Then there will be a sequence $\{a_i\}$ in $J^+(u) \cap J^-(v)$ failing to have an accumulation point in $J^+(u) \cap J^-(v)$. Construct a future-endless trip γ starting at $y_o \in I^-(u) \cap N_{i_0}$, just as in proposition 9 but with time reversed.

you're at the top of page 59.

13.7 Exercises

13.7.1 Standard Point-Set Topology

1. Let X be a topological space, $A \subset X$. Suppose that for each $x \in A$ there exists an open set $U \ni x$ such that $U \subset A$. Show that A is open.

2. Let X be a topological space. Prove that (A) ϕ and X are closed (B) Arbitrary intersections of closed sets are closed (C) Finite unions of closed sets are closed.
3. Let Y be a subspace of X. If A is closed in Y and Y is closed in X, then A is closed in X.
4. Let $A \subset X$. Then $s \in \bar{A}$if and only if every open set U containing x intersects A.

5. Prove that $\bar{A} = A \cup A'$, where A' =set of all limit points of A

6. Prove: A subset B of a topological space X is closed if and only if it contains all its limit points.
7. Theorem: Every finite point set in a Hausdorff space is closed.

8. Find a homeomorphism of (-1,1) onto the real line.

9. Let $f : X \to Y$ and $g : Y \to Z$ be continuous. Show that $g \circ f : X \to Z$ is also continuous.

10. Show that the interval (a,b) (0,1) (and [a,b] [0,1]). " " means "homeomorphic".

11. Proposition: A space X is connected iff ("if and only if") the only subsets of X that are both open and closed in X are the empty set and X itself.

12. Proposition: the continuous image pof a connected space is connected.

13. Proposition: Every closed subset of a compact set is compact.

14. Proposition: the continuous image of a compact space is compact.

13.7.2 Topology of Spacetime

15. Let γ be a timelike geodesic or curve, V the unit tangent vector field to γ. Show that $V_a \nabla_b V^a = 0$

16. Find the geodesics in R^2 for the metric $ds^2 = dr^2 + r^2 d\theta^2$.

17. Characterize all static two-dimensional spacetimes with diagonal metric i.e. solve Einstein's equations for $ds^2 = -e^{A(x)}dt^2 + e^{B(x)}dx^2$. Note: recall that spacetimes that differ only by a coordinate transformation are identical.

18. (A) Show that any curve whose tangent satisfies $T^a \nabla_a T^b = \alpha T^b$, with α an arbitrary function on the curve, can be reparametrized to satisfy $T^a \nabla_a T^b = 0$. (B) Let λ be an affine parameter of a geodesic γ. Show that all other affine parameters of γ take the form $at + b$, where a, b are constants.

19. Verify that $\partial/\partial\phi$ satisfies the Jacobi equation, and that the north and south poles are indeed conjugate points.

20. Prove that $<<$ is open, that is, if $p << q$ in M, there are disjoint neighborhoods $\mathcal{U}$ of p and $\mathcal{V}$ of q such that if $p' \in \mathcal{U}$ and $q' \in \mathcal{V}$ then $p' << q'$.

21. Prove: (A) $I^+[S]$ is open for any set S.
(B)$x \in I^+(y) \iff y \in I^-(x)$ and $x \in J^+(y) \iff y \in J^-(x)$
(C) $I^+[S] = I^+[\bar{S}]$
(D) $I^+[S] = I^+[I^+[S]] \subset J^+[S] = J^+[J^+[S]]$

The following exercises, numbers 22-26, are "True or False". That means you must either prove them true, or show they're false by finding a counter-example.

22. If $I^-(p) \subset I^+(q)$, then there is a closed trip through p.

23. If $I^-(p) \cap I^+(q) \neq \phi$, then $q \in I^-(p)$.

24. If the chronological futures of two points do not intersect, then neither do their pasts.

25. If $I^-(p) \cap I^+(q) \neq \phi$, then $q \in I^-(p)$.

26. Let S be a nonempty subset of M. If $S = I^+[S] = I^-[S]$, then $S = M$.

27. On the domain of a coordinate system, if $V = \sum_i V^i \partial_i$ and $W = \sum_j W^j \partial_j$, then

$$[V, W] = \sum_{i,j} \left(V^i \frac{\partial W^j}{\partial x^i} - W^i \frac{\partial V^j}{\partial x^i} \right) \partial_j.$$

Note: $\partial_i = \frac{\partial}{\partial x^i}$

Chapter 14

Spinors in GR

Chapter 15

Quantum Gravity

15.1 Relativistic Quantum Fields

15.2 Basic Particle Theory

15.3 Supergravity

15.4 Twistor Theory

15.5 Canonical Variables

15.6 String Theory

15.7 Natural Strings