Lectures on Introduction to General Relativity

(SERC Preparatory School, IIT Madras, 2009)

Ghanashyam Date

The Institute of Mathematical Sciences

C.I.T. Campus, Tharamani, Chennai, 600 113.

shyam@imsc.res.in

Draft Copy Oct 29, 2009

Contents

Chapter 1

Introduction

1.1 Space-time, general relativity and gravitation

We begin with a quote of Einstein: *theory of relativity is intimately connected with a theory of space and time* [1]. The primary concern of theory relativity thus seems to be space-time and there is no mention of gravity to begin with. In the final formulation of a theory of space-time though, gravity is an integral part.

We have here three seemingly unrelated ideas/concepts – the idea of a space-time, the idea of 'democracy of observers' ("relativity") and the *phenomenon* of gravity – which are very tightly intertwined. Let us trace through the arguments that lead to the synthesis.

The idea of space: Intuitively, space is something in which things happen - bodies can be (and do) moved around. The space of everyday experience is such that what we perceive as rigid bodies can be moved around, well, rigidly without any distortions. One uses this fact of experience to set up coordinate systems to label points that could be occupied by bodies, particles etc. The most familiar coordinate system one sets up is the Cartesian system of orthogonal axes. An important property of the space one notices is that if one has assigned coordinates (x, y, z) and (x', y', z') to two ends of a rigid rod, then its length is given by:

$$\text{Length}^2 \ = \ (x' - x)^2 + (y' - y)^2 + (z' - z)^2. \tag{1.1}$$

This follows from noting that the coordinates are assigned by counting the numbers of unit rods needed along each axis and the Pythagoras theorem of Euclidean geometry.

Notice however that there are infinitely many Cartesian systems – each observer can choose his/her orientation of axes and of course the origin. Every one of these coordinate systems will give the same expression for lengths and the length of a given rod computed by different observer will turn out to be equal (assuming same units are used!). We know that all these coordinate systems with common origin are related to each other by rotations or the orthogonal transformations. Also observe that if a freedom loving observer decides to use non-orthogonal axes, the expression for length in terms of his/her coordinates will be different.

We can summarize by saying that *space is something that exists and is made 'manifest' by using coordinate systems set up by observers by using physical objects and processes.* The space of everyday experience is three dimensional and is such as to distinguish a class of coordinate systems - the Cartesian ones - in which the lengths of rigid rods are given by the specific expression. This distinguished class is generated from any one system, by orthogonal transformations.

The idea of relativity of 'observers': We see immediately where an 'observer' enters a theory of space and time. The procedure of making the space and time manifest involves setting up of coordinate systems which is done by real observers using real rods and clocks and using real physical processes. This procedure thus can not be independent of properties of physical objects and therefore the (metrical) properties of space time should not be mandated *ab initio* but should be inferred. An obvious question then is whether there are any criteria to be stipulated for preference for some observers. Just as one can have non-Cartesian coordinate systems which are equally good as far as assignments of coordinates goes, but they are not 'preferred' or naturally singled out because of the expression for the lengths. Identification of a distinguished (equivalence) class of observers corresponds to a (restricted) 'Principle of Relativity'. This class of Cartesian coordinate systems (or observers) can be regarded as a *'principle of relativity of orientation'*.

Time and Galilean Relativity: Now one can ask if there is a relativity with regards to states of motions of observers. Our experience with mechanics (equations of motion) leads us to identify the so called 'inertial observers' as a distinguished class of observers. Recall that an inertial observer is one who will verify the Newton's first law of motion namely, in the absence of an agent of force a body continues its state of uniform motion. Such observers are realized in practice by being far away from all known agents of forces (eg shield electromagnetism and/or use neutral test bodies and be far away from a massive body). We still have to identify relations between two inertial observers analogous to the orthogonal transformations between Cartesian systems. The Galilean Relativity makes an explicit statement about it as:

Time is absolute (independent of observer) and is same (up to shifts of origins) for all inertial observers while space coordinates are related by a time dependent translation i.e.

$$t' = t + a \quad , \quad \vec{r'} = R(\vec{\hat{n}})r + \vec{v}t + \vec{c} \tag{1.2}$$

Clearly, these transformations leave the acceleration and hence equation of motion invariant. Despite the somewhat circular nature of definition of an inertial system, in practice Galilean Relativity worked very well as far as *mechanical phenomena* were concerned. It failed for electromagnetism, Maxwell's theory being not invariant under the Galilean transformations.

One had two options now: either Galilean relativity is applicable only to mechanical phenomena or that Galilean transformations need to be modified. If the former is valid, then earth's velocity relative to an absolute space or ether or whatever should be detectable, say by doing experiments with light. All such attempts failed. Speed of light was firmly constant independent of earth's motion. The conflict between electromagnetism and Galilean transformations must be faced.

Einstein believed that relativity of inertial observers should *not* be confined only to mechanical phenomena. There is also the implicit assumption in the Galilean transformations that time assignments are independent of observers which could be possible if clocks could be synchronized by sending instantaneous signals. However *if instantaneous transmission of signals is not possible* then Galilean transformations, particularly $t' = t + a$ will be fictitious.

Hence, there is a case to doubt Galilean transformations. While relativity of inertial observers may still be maintained, Galilean transformations need not be. What should replace these? Whatever these are, these should lead to same speed of light measured by all inertial observers.

As we all know, the new set of transformations are the Lorentz transformations which look like,

$$\begin{aligned} x' &= \gamma(x - \beta t) \qquad \beta := \frac{v}{c} \quad , \quad \gamma := \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \\ t' &= \gamma(t - \frac{\beta}{c}x) \\ y' &= y \\ z' &= z \end{aligned} \tag{1.3}$$

These leave invariant the new *space-time intervals*:

$$(\Delta s)^2 := c^2(\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \tag{1.4}$$

We may summarize now: A principle of relativity with respect to the state of motion of observers can be formulated by asserting that space and time, now to be regarded as a single entity space-time, is such as to admit a distinguished class of frames (or observers) called the 'inertial frames'. These are obtained from any one member by the Lorentz transformations which leave the space-time intervals invariant. Neither mechanical nor electromagnetic phenomena can single out any one inertial frame.

Einstein was still not satisfied. Newtonian gravity did not conform to the principle of special relativity. There is also no conceivable reason as to *why* inertial frames whose definition itself is somewhat circular, are preferred. While the space-time was such as to imply modifications of physical properties eg length contraction, time dilation etc but there is no provision in the theory to incorporate effects of material bodies on the space-time. Such one way influencing, being against the Machian view point, was deeply unsatisfying to Einstein.

At this point Einstein observes the striking numerical equality of the 'inertial mass' and the 'gravitational mass'. It is striking because both the notions are defined so differently. Ratios of inertial masses is defined via the ratios of the accelerations suffered by two bodies subjected to the same force of whatever kind. The gravitational mass on the other hand is something characteristic of gravitational force between two bodies much the same way as electric charges are characteristic of electro-static forces. There is no *reason* for these to be equal. However *if* these are exactly equal, then it follows that gravitational effects can be interpreted in terms of acceleration and hence can be made to 'disappear' by referring to an 'accelerated observer'. Conversely, an observer accelerated relative to an inertial observer can equally well describe motion by postulating a gravitational field.

One now sees a way to resolve one of the puzzles. Inertial frames are preferred because there are no gravitational fields. If these are present, then these could be interpreted in terms accelerated observers. So if one includes gravitational phenomena as well then one might as well propose a principle of relativity of *all observers* regardless of the their state of motion.

Einstein then considers an inertial observer K and another observer K' rotating uniformly with respect to K. Using properties of Lorentz contraction, he argues that spatial geometry as determined by the rotating observer should deviate from the Euclidean one as determined by K since the ratio of circumference to radius will be smaller for K'. Thus geometry as inferred by an accelerated observer is non-Euclidean in general. But since K' will perceive a gravitational field, he should conclude that gravity affects geometry. This is very satisfying since now one sees the possibility of material bodies affecting the geometry of space-time. Since Newtonian gravity is determined by material bodies and gravity can affect geometry, it follows that material bodies can affect the geometry of space-time. How exactly this happens is of course the content of the Einstein Field equations.

But if gravity can be 'gotten rid off' by going to a freely falling lift, is gravity completely fictitious? No! One can nullify effects of gravity only in small portions of space-time. One can experience weightlessness in a freely falling lift but on earth. Earth itself is freely falling in the gravitational field of the Sun but tides do occur. This in fact suggests that gravitation is really manifestation of *tidal forces* which in the geometrical set up will turn out to be the effects of the curvature. Thus preferred status of inertial frames is not completely discarded but its applicability is limited to small regions of space-time. Since in such small portions gravity can be nullified, we can safely stipulate that laws of physics take a form consistent with *special theory of relativity* in such *locally inertial frames*.

One can appreciate the grand synthesis now. Space-time is not some *inert*, arena in which things happen but is a *dynamical entity*. This comes about because space-time must be manifested via frames of references or coordinate systems to be constructed in conformity of properties of real physical objects (no fictitious assumptions of infinite speeds). Here in enters principle(s) of relativity of classes of observers. The phenomenon of gravity is such that one can simultaneously bypass the vexed question of singling out inertial observers and non-conformity of Newtonian gravity to special relativity with the additional bonus of space-time and matter both influencing each other.

We see the conceptual scheme of the synthesis. But now we need to make it quantitative and precise. We need a suitable mathematical framework.

From the example of special relativity, we already observe that the new framework should be generalization of the Minkowski space-time i.e. it should be a space on which is defined a notion of an *invariant interval* which however is not a *fixed* expression. Since all observers are to be treated on the same footing and each one sets up a coordinate system, *all* coordinate systems should be transformable into each other and physical quantities should be analogues of the Lorentzian *four vectors*, transforming under these coordinate transformations. Unlike the Lorentz transformations which are *linear*, these are arbitrary and this feature needs to be taken care of while taking derivatives and setting up the differential equations of physics.

We will see that the mathematical framework is that of *Riemannian geometry*.

Chapter 2

Mathematical digression

While we won't need all the machinery in this course, this is good opportunity to get an exposure to the hierarchy of structures one introduces in arriving at the desired Riemannian geometry. There are several excellent books available [2]. The aim here is to introduce structures in stages to see what they enable us to do. Only basic ideas are discussed.

2.1 Sets, Metric Spaces and Topological Spaces

The absolute minimum to begin with is a set or a well defined collection of elements. We can consider subset of a set, a collection (or set) of subsets of a given set, can construct new sets by defining *Cartesian product* of two sets X, Y as the set of ordered pairs whose first entry is an element of X and the second entry is an element of Y. There are two notions that we need, that of a *mapping* between two sets and that of a *binary relation* on a set.

The notion of a *mapping*, $f : X \to Y$, associates a unique element of Y to every element of X. (It can be represented also as a subset of the Cartesian product $X \times Y := \{(x, f(x))/f(x) \in Y, \ \forall\ x \in X\}$). Some features immediately arise. A mapping f is one-to-one, or *injective*, if $f(x) = f(y) \Rightarrow x = y$; it is on-to or *surjective*, if for every $y \in Y$, $\exists\ x \in X$; it is *bijective*, if it is one-to-one and on-to.

For every map $f : X \to Y$, we can define *inverse image* of $y \in Y$ to be the subset: $Inv_f(y) := \{x \in X/f(x) = y\}$. For an injective map we can define an *inverse map* f^{-1} :Range(f) $\subset Y \to$ Domain(f) $\subset X$. Notice that if there is a bijective map from X to Y, then inverse map is also bijective and *all* set theoretic properties of the two sets X and Y are identical - the only difference between the two is the labels on their elements. The two sets are then said to be *equivalent*. Examples: two finite sets containing the same number of elements are equivalent; set of even integers is equivalent to set of all integers; set of rational numbers is equivalent to the set of integers; an open interval $(0, 1)$ is equivalent to the set of all real numbers etc.

The notion of a binary relation R on a set X is simply that it is a subset $R \subset X \times X$. Some particular subsets deserve special names eg *equivalence relation, partial order, . . . etc.* For us, the first one is more relevant. It is defined by the following conditions: (i) Reflexivity: $(x, x) \in R, \forall\ x \in X$, (ii) Symmetry: $(x, y) \in R \Rightarrow (y, x) \in R$, and (iii) Transitivity: $(x, y) \in R, (y, z) \in R \Rightarrow (x, z) \in R$. Innocuous as these may look, one has an important result that *Every equivalence relation partitions the set and conversely, every partition defines* *an equivalence relation.* Here, partition of a set X means X can be expressed as $X = \cup_i X_i$ such that $X_i \cap X_j = \Phi$. The proof is very simple and is left as an exercise.

As an example, consider X = set of all sets. On this, define a relation xRx' iff there exists a bijective map between x and x'. Show that this is an equivalence relation. Define the *equivalence class of* x, $[x] := \{y/yRx\}$. Show that $[x] = [y]$ iff $y \in [x]$ (or $x \in [y]$). Otherwise $[x] \cap [y] = \Phi$. Thus, the set of all sets is partitioned into classes consisting of equivalent sets.

Both these notions are used repeatedly to organize various structures.

In order to generalise the familiar calculus, we need to suitably generalise the notions of limits of sequences, continuity of functions, their derivatives and integrals. To this end, let us recall the definitions of limit of a sequence of real numbers and continuity of a function at a point.

Definition: A sequence $\{x_n\}$ is said to converge to x if for every $\epsilon > 0$, $\exists\ N > 0$ such that $|x_n - x| < \epsilon\ \forall\ n > N$. This is denoted as $x_n \to x$. Likewise,

Definition: A function $f(x)$ is said to be continuous at a if for every $\epsilon > 0$, $\exists\ \delta > 0$ such that $|f(x) - f(a)| < \epsilon\ \forall\ |x - a| < \delta$.

In these definitions, the absolute value of the differences provides a notion of nearness. The generalization to sequences of points in n-dimensional spaces or functions of n-variables involves only using the corresponding definition of the absolute value, namely the length of the difference vector also called its Euclidean norm. This norm satisfies the following properties: (a) $|\vec{x} - \vec{y}|$ is always non-negative and vanishes only of the difference vector vanishes; (b) it is symmetric in $\vec{x}, \vec{y}$ and (c) $|\vec{x} - \vec{y}| \le |\vec{x} - \vec{z}| + |\vec{z} - \vec{y}|$ (triangle inequality). Interestingly, these properties are sufficient to prove all the results involving limits and continuity of real variables.

Now observe that, suppose we let the variables to be elements of an arbitrary set - not necessarily of numbers - but equip the set with *distance function* $d : X \times X \to \mathbb{R}$ i.e. $d(x, y) \in \mathbb{R}$ which precisely satisfies the three properties listed above. Then we *can* take over the definition of limit of a sequence of elements of X, $x_n \in X$! To define continuity of mapping $f : X \to Y$, we will need to introduce a distance function on X as well as on Y. The distance function is called a *metric* on the set X.

A set X together with a metric d defined on it, is called a *metric space.* Introducing this notion, we have managed to extend the notions of limit and continuity from sets of numbers to arbitrary sets which admit a metric. To be explicit, let us define an $\epsilon-$neighbourhood of $x \in X$ as: $N_\epsilon(x) := \{y \in X/d(y, x) < \epsilon\}$. The definition of limit $x_n \to x$ then becomes: for every $\epsilon > 0\ \exists\ N > 0$ such that $x_n \in N_\epsilon(x)\ \forall\ n > N$. In the definition of continuity, there will be $N_\epsilon(f(a))$ and $N_\delta(a)$ with two metrics on X and Y respectively. There are still the ϵ, δ, N which are real numbers.

Let us define $A \subset X$ to be an *open* set if for *every* $x \in A$, there exists an $\epsilon > 0$ such that $N_\epsilon(x) \subset A$. It follows that every $\epsilon-$neighbourhood is an *open set* of X; the set X itself is open and so is the empty set Φ. These open sets satisfy two crucial properties: (A) union of arbitrary number of open sets is an open set and (B) intersection of *finitely many* open sets is an open set. It turns out that these two properties together with X, Φ being open, are sufficient to deduce *all* properties/results pertaining to limits and continuity.

We can now free ourselves from the ϵ, δ numerical features from the notion of nearness - all we need to do is have a supply of proper subsets of X satisfying the properties (A) and (B). This leads to our final generalization which provides a satisfactory formulation of notion of nearness. Here is the definition.

Let X be a non-empty set and let T be a collection of subsets of X such that (1) $X, \Phi \in T$ (2) arbitrary unions of members of T is contained in T and (3) all intersections of finitely many members of T are contained in T. T is called a topology on X; members of T are called open sets and the set X together with a topology T is called a Topological space.

Exercise: Re-write the definition of limit of a sequence in a topological space.

There are three basic properties of topological spaces namely: (i) (local) connected-ness; (ii) separability and (iii) (local) compactness. These should be seen in the references. We will discuss these if and when needed.

Remark:

- On a given set, there can be several topologies and hence several different definitions of convergence of sequences.

  Two extreme examples are: (1) *trivial topology*, the only open sets are X and Φ and (2) *discrete topology*, every subset of X is an open set.

- Even a finite set can admit a topology and hence a corresponding notion of nearness.

- In metric spaces, there is a natural topology, namely that given by the $\epsilon-$neighbourhoods.

- Finally, *without a choice of a topology, it is meaningless to talk about limits.*

We can immediately define mappings between two topological spaces: $f : (X, T) \rightarrow (X', T')$. As a map between the two sets, F can be injective and/or surjective and/or bijective. The two sets can be equivalent as sets (there exist a bijective map). But is there a sense in which the map "preserves" also the topologies? The answer is yes.

Topology allows us to introduce further attributes of a maps. $f : X \rightarrow Y$ is *open* if every open set of X is mapped to an open set of Y; it is *continuous* if inverse image, Inv_f of every open set of Y is an open set of X. f is an *homeomorphism* if it is bijective, open and continuous. Two topological spaces are homeomorphic if there exist a homeomorphism between them. This defines an equivalence relation which partitions the set of all topological spaces into mutually homeomorphic spaces. All set theoretic and topological properties for homeomorphic space are identical.

For finite sets one can easily display topologies and maps illustrating these definitions. The topology defined by the Euclidean norm on $\mathbb{R}^N$, is called the *usual topology* of $\mathbb{R}^N$.

Exercise: Convince yourself that the topological definition of continuity reduces to the usual definition for real function with usual topology.

We are now ready to go to the next step of generalising the notion of differentiation.

2.2 Manifolds and Tensors

We would like to see if the notion of differentiation can be imported to a general topological space. As before, let us recall the definition of derivative of a function. It is defined as

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} =: \frac{df}{dx}$$

While we can generalize the numerator and the denominator, how do we generalise the notion of *division* to non-numerical entities such as points of a topological space?

There is one way out of this, namely, *assign* numbers to points of the topological space. An immediate question is, how? This should be done in a "continuous manner" (recall that in the usual case, differentiation is defined only for functions which are at least continuous). This could be done, for example, by requiring suitable open sets of the topological space to be homeomorphic to suitable open sets of some $\mathbb{R}^n$. The integer n could provide the notion of 'dimensionality' (number of coordinates/number of independent variables in a function etc). For $n = 2$, this in turn can be imagined as sticking pieces of graph paper on the surface of some balloon (a topological space). But, clearly there are infinitely many ways of doing this and there is no way to make any natural choice. We can live with this freedom provided we can ensure that whatever we really want to do (define a derivative) does *not* depend on the choice of the labelling. This is done as follows. In anticipation, we denote a topological space as M from now on.

We first define an $n-$dimensional *Chart* around a point $p \in M$. This consists of an open set u_α containing p i.e. a neighbourhood of p, together with a homeomorphism $\phi_\alpha : u_\alpha \to \mathbb{R}^n$ i.e. $\phi_\alpha(q) \leftrightarrow (x^1(q), x^2(q), \ldots, x^n(q))$. Recall that homeomorphism is a one-to-one, on-to, open and continuous assignment. The $x^i(q)$ are called *local coordinates* of point $q \in u_\alpha$.

Introduce such charts around each point of M and choose a collection of charts covering all of M. Some of the charts may overlap: $u_\alpha \cap u_\beta \neq \Phi$. The common point then have two different coordinates, say $x^i(q)$ and $y^i(q)$ and due to the on-to-one assignments, we can use this to define a *coordinate transformation* $x^i \leftrightarrow y^i$. Clearly these are one-to-one, on-to (with respective domains and ranges) and continuous since the defining homeomorphisms are. We now require that $y^i(x^1, x^2, \ldots, x^n)$ *and* $x^i(y^1, y^2, \ldots, y^n)$ *are both infinitely many times differentiable functions.* The two conditions namely the collection of charts covering all of M and the smoothness of coordinate transformations in the overlap, implies that all charts must be of the same dimension, say, n. Such a collection of charts is called a *Smooth, $n-$dimensional Atlas*[1].

We can construct several different smooth atlases. Let us define a relation on the set of all atlases. We will say that two atlases, $\{(u_\alpha, \phi_\alpha)\}, \{(v_a, \psi_a)\}$, are *compatible* if their union is also an atlas. This requires that even for overlapping neighbourhoods from different atlases, the corresponding coordinate transformations are also smooth. This is an equivalence relation and the equivalence classes are called *differential structures* on the topological space. A topological space together with a given differential structure is called a *differential manifold.*

[1]Functions which k-times differentiable (partial derivatives in case of several variables) are said to be of class C^k. C^0 refers to continuous functions while C^∞ are termed *smooth*. One can also have real analyticity, complex analyticity classes etc. The atlases involving coordinate transformations of a given class are given the same adjective.

To appreciate the need for the smoothness of coordinate transformation consider a possible definition of differentiability of a real valued function $f : M \to \mathbb{R}$. The function itself can be defined independent of any atlas eg temperature on the surface of earth which does not need (longitude, latitude) to be chosen. Referring to a chart around some p, we convert the function to a function of x^i. We can now *define* f to be differentiable at p if $f(x^i)$ is differentiable(and we know what this means). But now the differentiability of a function seems to be tied with the particular chart chosen. If we choose a different chart, does the function still remain differentiable? Well, let us assume that $\partial f / \partial x^i$ exist. Let y^j denote another set of coordinates. By the chain rule, we expect that $\frac{\partial f}{\partial y^i} = \frac{\partial x^j}{\partial y^i}\frac{\partial f}{\partial x^j}$. Evidently, the left hand side will be well defined iff $\frac{\partial x^j}{\partial y^i}$ is well defined i.e. the coordinate transformation is well defined. Furthermore, f being smooth will be meaningless, unless the coordinate transformations are smooth. But this is precisely what is guaranteed by the condition on the atlas! So, although we need to use *arbitrary coordinates* to make sense of differentiability, the additional structure introduced, ensure that the property of differentiability is *independent* of the choice of coordinate. Our primary goal of importing notions of differentiation to topological spaces is achieved. The price to pay is the introduction of a differential structure and an implicit restriction to only those topological spaces which are locally $\mathbb{R}^n$.

As for topological spaces, there can be several different differential structures on the same topological space eg S^7 has 28 differential structure while $\mathbb{R}^4$ has infinitely many differential structures. For $\mathbb{R}^n$ with the usual topology and an atlas consisting of a just a *single* chart - the chart defined by the identity map, defines the "usual" differential structure. The analogue of homeomorphism in this case is called a *diffeomorphism*. Let M, N be two differential manifolds and let $f : M \to N$ be a map which is a homeomorphism of the underlying topological spaces. Under this, open sets of M go to open sets of N and this induces a corresponding coordinate transformation of local coordinates x^i on M going to local coordinates y^i on N. If these coordinate transformations ($x^i \leftrightarrow y^i$) are smooth, then f is called diffeomorphism and M, N are said to be diffeomorphic to each other. Again this is an equivalence relation and partitions the set of all differential manifolds into classes of mutually diffeomorphic manifolds.

On a manifold, several types of quantities can be defined in a natural manner. These can be defined in a *manifestly coordinate independent* manner or through use of coordinates such that the choice of coordinates does not matter. We have already seen the example of one such quantity, namely smooth, real valued functions $f : M \to \mathbb{R}$. Our next quantity is a smooth curve on a manifold.

A curve γ on M is a map $\gamma : (a, b) \subset \mathbb{R} \to M$ from an open interval into the manifold i.e. $t \in (a, b) \to \gamma(t) \in M$. Referring to local coordinates, this is represented by n functions of a single variable, $x^i(t), t \in (a, b)$. The curve is smooth, if these functions are smooth functions of t. Again, smoothness of λ is independent of the choice of local coordinates.

Let us assume for definiteness that $0 \in (a, b)$ and denote $p = \gamma(0)$. Every curve on a manifold gives rise to a *tangent vector* as follows. For any function $f : M \to \mathbb{R}$,

$$\left.\frac{d}{dt}f\right|_{\gamma} := \lim_{\epsilon \to 0} \frac{f(\gamma(\epsilon)) - f(\gamma(0))}{\epsilon} \tag{2.1}$$

Using a chart, $(u_\alpha, \phi_\alpha$, gives the function f as a function of the local coordinates as $f_\alpha(x^i(p)) :=$

$f(\ \phi_\alpha^{-1}(x^i)\ )$. In terms of this, we get,

$$
\begin{aligned}
\left.\frac{d}{dt}f\right|_\gamma &= \lim_{\epsilon\to 0}\frac{f_\alpha(x^i(\gamma(\epsilon))) - f_\alpha(x^i(\gamma(0)))}{\epsilon} && \text{But,}\\
x^i(\gamma(\epsilon)) - x^i(\gamma(0)) &\approx \epsilon\left.\frac{dx^i}{dt}\right|_{t=0} \\
\therefore\ \left.\frac{d}{dt}f\right|_\gamma &:= \lim_{\epsilon\to 0}\frac{f_\alpha(x^i(\gamma(0)) + \epsilon\frac{dx^i}{dt}) - f_\alpha(x^i(\gamma(0)))}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{\epsilon\frac{dx^i}{dt}\frac{\partial f_\alpha}{\partial x^i}}{\epsilon} \\
&= \left.\frac{dx^i}{dt}\right|_\gamma \frac{\partial}{\partial x^i}f_\alpha \qquad \forall\ \ f: M \to \mathbb{R} && (2.2)
\end{aligned}
$$

The (2.1) gives a manifestly coordinate independent definition while the subsequent equations gives expression involving local coordinates. Since the function is arbitrary, one can think of the $\frac{d}{dt}|_\gamma$ as an operator which takes function to numbers. There is one such operator for each curve γ and it is called a tangent vector to the manifold at the point $p = \gamma(0)$. One can collect all such tangent vectors at the same p and define a vector space in an obvious manner. This is called the *Tangent Space to M at p* and is denoted as $T_p(M)$. What is its dimension?

Consider eqn.(2.2). Stripping off the function, the tangent vectors are parametrized by the n numbers $\frac{dx^i}{dt}|_\gamma$ while $\frac{\partial}{\partial x^i}$ are linearly independent elements of the tangent space. This implies that the dimension of the tangent space is precisely n. The $\{\frac{\partial}{\partial x^i}\}$, form a basis, called a *coordinate basis*, for the Tangent Space. A general tangent vector is therefore expressible as $X := X^i\frac{\partial}{\partial x^i}$.

If we refer to another local coordinates y^i, then any given tangent vector is expressed as

$$
\begin{aligned}
\left\{\frac{dx^i}{dt}\right\}\frac{\partial}{\partial x^i} &= \left\{\frac{dx^i}{dt}\right\}\left\{\frac{\partial y^j}{\partial x^i}\right\}\frac{\partial}{\partial y^j} = \left\{\frac{dy^j}{dt}\right\}\frac{\partial}{\partial y^j} \qquad \text{or,} \\
X^i\frac{\partial}{\partial x^i} &= X^i\left\{\frac{\partial y^j}{\partial x^i}\right\}\frac{\partial}{\partial y^j} = Y^j\frac{\partial}{\partial y^j}
\end{aligned}
$$

We notice that if we have a set of quantities X^i which transform under coordinate transformation as $X^i \to Y^i = \frac{\partial y^i}{\partial x^j}X^j$, then the combination $X := X^i\frac{\partial}{\partial x^i}$ is *independent of the coordinates.*

Such quantities, X^i, are called *components of a contravariant vector* and are elements of the tangent space, which is a vector space of dimension n.

Now, it is a general construction that given a vector space V, one defines another vector space, called its *Dual*, V^* as the collection of linear functions on V. That is, consider $f : V \to \mathbb{R}$ such that $f(a\vec{u} + b\vec{v}) = af(\vec{u}) + bf(\vec{v})$. The set of all such linear functions can be given a vector space structure in an obvious manner: $(a \odot f_1 \oplus b \odot f_2)(\vec{x}) := af_1(\vec{x}) + bf_2(\vec{x}), \forall\, \vec{x} \in V$. If $\{\vec{e}_i\}$ is a basis for V so that $\vec{x} = x^i\vec{e}_i$, then $f(\vec{x}) = x^i f(\vec{e}_i) := x^i f_i$. All possible elements of V^* are obtained by varying the $\{f_i\}$ and thus dimension of V^* is the same as that of V. The tangent space is no exception and its dual is called the *Cotangent Space*, $T_p^*(M)$. A basis for $T_p^*(M)$ dual to a coordinate basis for $T_p(M)$ is denoted as $\{dx^i\}$ and is defined by

$dx^i(\frac{\partial}{\partial x^j}) := \delta^i_j$. A general element ω of the cotangent space, can be evaluated on a general element X of the tangent space as,

$$\omega(X) \ = \ \omega_i dx^i \left(X^j \frac{\partial}{\partial x^j} \right) \ = \ \omega_i X^i$$

Here, ω_i are called the components of a cotangent vector, relative to the basis $\{dx^i\}$. Referring to another coordinate system leads to,

$$\therefore \ \omega_i X^i \ = \ \omega'_i Y^i \ = \ \omega'_i \frac{\partial y^i}{\partial x^j} X^j \quad \Rightarrow \quad \omega'_i \ = \ \frac{\partial x^j}{\partial y^i} \omega_j \ .$$

Thus, we deduce that the components of a cotangent vector transform as: $\omega_i \to \omega'_i = \frac{\partial x^j}{\partial y^i}\omega_j$. The cotangent vectors themselves are *invariant under a coordinate transformation.*

There is another natural construction given two vector spaces, U, V, namely to construct another vector space called their *Tensor Product* and denoted as $U \otimes V$. Its dimension is the product of the dimensions of the two vector spaces. With the tangent and the cotangent spaces available, we can construct arbitrary tensor product spaces from $T_p(M)$ and $T^*_p(M)$ and then take their duals (linear functions). Elements of these duals are called *Tensors.* As it stands, these definitions are phrased independent of any reference to local coordinates. We will use an alternative but equivalent definition in terms of "components", as was illustrated for the tangent and cotangent spaces. The coordinate independent definitions are given in the appendix. Here is the definition we use.

A set of quantities, $T^{i_1 i_2 \dots i_p}{}_{j_1 j_2 \dots j_q}(x)$ that transform under a coordinate transformation $x^i \to y^i(x)$ as,

$$(T')^{i_1 i_2 \dots i_m}{}_{j_1 j_2 \dots j_n}(y(x)) \ = \ \left\{ \frac{\partial y^{i_1}}{\partial x^{m_1}} \frac{\partial y^{i_2}}{\partial x^{m_2}} \cdots \frac{\partial y^{i_p}}{\partial x^{m_p}} \right\} \left\{ \frac{\partial x^{n_1}}{\partial y^{j_q}} \frac{\partial x^{n_2}}{\partial y^{j_q}} \cdots \frac{\partial x^{n_q}}{\partial y^{j_q}} \right\} T^{m_1 m_2 \dots m_p}{}_{n_1 n_2 \dots n_q}(x)$$

are said to be components of a tensor of *contravariant rank p and covariant rank q*. The arguments y, x are two different local coordinates of the same point $p \in M$. These quantities are "born" with a manifold and represent quantities which have a coordinate invariant meaning.

Being elements of a vector space, tensors of the same rank at a given point, can be added and scalar multiplied. From tensors of different ranks, we can construct new tensors of higher ranks by multiplying the components. This is the operation of *tensor or outer product.* We can also equate one or more contravariant (upper) index pair-wise with covariant (lower) indices or the same or different tensors resulting in reduction in both the contravariant and the covariant ranks. This is called *contraction or interior products.* Elements of tangent space correspond to rank (1,0) tensors while those of the cotangent space correspond to rank (0,1). Functions are rank (0,0) tensors and also referred to as scalars.

Completely antisymmetric tensors of rank (0,k), $0 \leq k \leq n$, are called $k-$forms and for them another algebraic operations called *wedge product* is defined. We will not need it in this course, but the definitions are included in the appendix.

This concludes the discussion of algebraic operations that can be performed on tensors at each point of the manifold. We now proceed to tensor calculus, in particular, differentiation.

2.3 Affine Connection and Curvature

To discuss notions of differentiation, we must first introduce *Tensor Fields.* These are nothing but assignments of tensors of rank (p,q) to each point of the manifold. This assignment is such that the tensor components with respect to any coordinate basis are smooth i.e. partial derivatives of arbitrary order of the tensor components exist everywhere. *However, partial derivatives of tensor fields are not tensors themselves in general!*; the sole exception are the tensors of rank (0,0).

To see this, consider a rank (1,0) tensor $A^i(x)$. Consider its partial derivative, $\frac{\partial A^i}{\partial x^j}$. Under a coordinate transformation, we get,

$$\frac{\partial A'^{\,i}(y)}{\partial y^j} = \frac{\partial x^k}{\partial y^j}\frac{\partial}{\partial x^k}\left(\frac{\partial y^i}{\partial x^l}A^l(x)\right) = \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\frac{\partial A^l}{\partial x^k} + \frac{\partial x^k}{\partial y^j}\frac{\partial^2 y^i}{\partial x^k \partial x^l}A^l \tag{2.3}$$

The first term in the last equality has the correct form for a tensor component, the last term however is a spoiler. Had the transformations been at most linear, this term would have been absent. This is why while discussing derivatives of tensors with respect to Lorentz transformations, one does not face any issue. We need to consider some modification of derivative to construct a tensor. The reason is not hard to see. Taking derivatives involves taking difference of tensor components at two nearby points, but the tensor algebra holds only point-wise. This deficiency can be corrected by introducing an auxiliary quantity called an *Affine Connection*, $\Gamma^i{}_{jk}(x)$ whose transformation property is deduced as follows.

Define *a covariant derivative*, $\nabla_j A^i := \frac{\partial A^i}{\partial x^j} + \Gamma^i{}_{jk}A^k$ and demand that this quantity transforms as a tensor of rank (1,1). This fixes the transformation of the affine connection.

$$\begin{aligned}
\nabla'_j A'^{\,i} &:= \frac{\partial A'^{\,i}(y)}{\partial y^j} + \Gamma'^i{}_{jk}A'^k \\
&= \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\frac{\partial A^l}{\partial x^k} + \frac{\partial x^k}{\partial y^j}\frac{\partial^2 y^i}{\partial x^k \partial x^l}A^l \ + \ \Gamma'^i{}_{jk}\frac{\partial y^k}{\partial x^l}A^l \\
&= \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\left(\frac{\partial A^l}{\partial x^k} + \Gamma^l{}_{km}A^m\right) + \\
&\qquad \left[\left(\frac{\partial x^k}{\partial y^j}\frac{\partial^2 y^i}{\partial x^k \partial x^m} \ + \ \Gamma'^i{}_{jk}\frac{\partial y^k}{\partial x^m} - \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\Gamma^l{}_{km}\right)A^m\right] \\
&= \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\nabla_k A^l + 0
\end{aligned} \tag{2.4}$$

Thus we deduce that,

$$\Gamma'^i{}_{jk}(y(x)) := \frac{\partial y^i}{\partial x^l}\frac{\partial x^m}{\partial y^j}\frac{\partial x^n}{\partial y^k}\Gamma^l{}_{mn}(x) + \frac{\partial y^i}{\partial x^l}\frac{\partial^2 x^l}{\partial y^j \partial y^k} \tag{2.5}$$

The affine connection transformation has a tensor-like piece (the first term) which is homogeneous in the connection, but crucially has the *inhomogeneous* or connection independent piece as well (the second term). This piece is *symmetric* in the lower indices. It then follows that the antisymmetric combination, $T^i{}_{jk} := \Gamma^I{}_{jk} - \Gamma^I{}_{kj}$ actually transforms as a tensor of rank (1,2). This is known as the *Torsion tensor* of the affine connection. For our purposes, we will restrict to those affine connections which are symmetric in their lower indices i.e. the torsion tensor vanishes. In the appendix, the general affine connection is considered.

Now, unlike a tensor, *a (symmetric) connection can be made to vanish at any chosen point.* The proof is simple. Let x^i be local coordinates around a point p such that $x^i(p) = 0$ (this is only for convenience). Consider a coordinate transformation $y^i(x) := x^i + \frac{1}{2}a^i{}_{jk}x^jx^k + o(x^3)$. This implies that the inverse transformation is $x^i(y) = y^i - \frac{1}{2}a^i{}_{jk}y^jy^k + o(y^3)$. It follows,

$$\Gamma'^i{}_{jk}(y(0)) \ = \ \delta^i_l\delta^m_j\delta^n_k\Gamma^l{}_{mn}(0) + \delta^i_l(-a^l{}_{jk}).$$

By choosing the constants $a^i{}_{jk} = \Gamma^i{}_{jk}(0)$, the result follows.

Exercise: By exactly analogous reasoning show that partial derivatives of a scalar is a tensor of rank (0,1) without any affine connection modification. And for tensor of rank (0,1), affine connection term is needed and the definition: $\nabla_j B_i := \frac{\partial B_i}{\partial x^j} - \Gamma^k{}_{ji}B_k$ *constructs a tensor of rank (1,1).*

What about covariant derivatives of other tensor fields? Observe that partial derivatives of scalars are rank (0,1) tensors automatically. The affine connection is needed to cancel-off the double derivatives of the coordinate transformations, which appear index-by-index in a tensor transformation. Thus, we must define covariant derivatives on higher rank tensors by adding an affine connection term for each contravariant index and *subtracting* such a term for each covariant index.

It follows that like the usual partial derivatives, the covariant derivatives also act *linearly* and satisfy the Leibniz rule: $\nabla(AB) = A(\nabla B) + (\nabla A)B$. These basic properties are satisfied by all covariant derivatives i.e. for every choice of an affine connection and there are infinitely many affine connections, on a manifold.

There is one crucial property of partial derivatives which is *not* shared by a covariant derivative: *covariant derivatives do not commute in general.* Using the notation, $\frac{\partial}{\partial x^i} =: \partial_i$, consider,

$$\begin{aligned}
\nabla_l\nabla_k B_j &= \partial_l(\nabla_k B_j) - \Gamma^m_{lk}\nabla_m B_j - \Gamma^m_{lj}\nabla_k B_m \\
&= \left\{\partial_l\partial_k B_j - \Gamma^m{}_{kj}\partial_l B_m - \Gamma^m_{lk}\partial_m B_j - \Gamma^m_{lj}\partial_k B_m\right\} \\
&\qquad + \left\{-\partial_l\Gamma^n{}_{kj} + \Gamma^m_{lj}\Gamma^n{}_{km} + \underline{\Gamma^m_{lk}\Gamma^n{}_{mj}}\right\} B_n \qquad (2.6) \\
\therefore\ [\nabla_l, \nabla_k] B_j &= \left\{\partial_k\Gamma^i{}_{lj} - \partial_l\Gamma^i{}_{kj} + \Gamma^i{}_{km}\Gamma^m_{lj} - \Gamma^i{}_{lm}\Gamma^m_{kj}\right\} B_i \\
\text{or}\ [\nabla_l, \nabla_k] B_j &= R^i{}_{jkl}B_i \qquad \text{with} \qquad (2.7) \\
R^i{}_{jkl}(\Gamma) &:= \partial_k\Gamma^i{}_{lj} - \partial_l\Gamma^i{}_{kj} + \Gamma^i{}_{km}\Gamma^m_{lj} - \Gamma^i{}_{lm}\Gamma^m_{kj} \qquad (2.8)
\end{aligned}$$

The terms in the first braces, all involving derivatives of the tensor field, and the underlined term in the second braces in eqn(2.6) are *symmetric in* $k \leftrightarrow l$ and hence drop out in the commutator of the covariant derivatives in the next equation. Equation (2.7) is known as the "Ricci Identity" and equation (2.8) defines the *Riemann Curvature tensor.*

<u>Remarks:</u>

- There is an alternative notation to denote partial and covariant derivatives, namely, $\partial_j T \Leftrightarrow T_{\ ,j}$ and $\nabla_j T \Leftrightarrow T_{\ ;j}$. This will also be used when convenient.
- It is straight forward to verify that

  $$[\nabla_l, \nabla_k] A^i \ = \ - R^i{}_{jkl}A^j$$

  and the commutator on higher rank tensors goes index-by-index.

- From eq.(2.7), it is obvious that $R^i{}_{jkl}$ is a tensor of rank (1,3) because the left hand side is a tensor of rank (0,3) and B_i is also a tensor of rank (0,1). It is antisymmetric in the last two indices[2].

- The Riemann tensor depends only on the affine connection *and* its derivatives. While the Γ^2 terms can be made to vanish at any point, the derivatives cannot be. So no coordinate transformation can make the Riemann tensor vanish if it is non-zero to begin with. This also means that *vanishing of the Riemann tensor is a necessary condition for the affine connection to vanish in a neighbourhood.*

  Along with an affine connection are born the two tensors: the torsion tensor and the Riemann curvature tensor. We have chosen the torsion to be zero. The Ricci identity has an additional term for non-zero torsion.

- The Riemann tensor satisfies two important identities: the algebraic *cyclic identity* $\sum_{(jkl)} R^i{}_{jkl} = 0$ and the differential *Bianchi identity*, $\sum_{(klm)} \nabla_m R^i{}_{jkl} = 0$.

- From the Riemann tensor on defines the *Ricci Tensor*, $R_{ij} := R^k{}_{ikj}$ which will play a role later.

There are two important notions associated with an affine connection, that of parallel transport and that of an affine geodesic. Consider a vector field X^i and construct the differential operator $X \cdot \nabla := X^i \nabla_i$. Acting on an arbitrary tensor, it produces another tensor of the same rank,

$$X \cdot \nabla T = \frac{dx^i}{dt} \nabla_i T = \frac{dx^i}{dt} \left(\partial_i T \pm \text{connection terms} \right) = \frac{dT(x^i(t))}{dt} \pm \frac{dx^i}{dt} \times (\Gamma \cdot T).$$

So, $X \cdot \nabla T = 0$ is a first order differential equation which has a unique solution given an initial condition $T(x(0))$. Thus, given a tensor $T(p)$ at a point p and a vector field X^i, we can determine a tensor *along the integral curve of the vector field.*

Thus solution of $X \cdot \nabla T_{\parallel} = 0$ defines the notion of *parallel transport of* $T(p)$ *along the vector field* X.

Since tensor of any rank can be parallel transported along any vector field, we can construct parallel transport of the vector field along itself, $X \cdot \nabla X^i_{\parallel} = 0$. In general, $X_{\parallel} \neq X$. The vector fields which do satisfy the equality define integral curves which are called *Affine Geodesics*[3]. The explicit and perhaps a bit familiar form of the equation for geodesic curves is:

$$X \cdot \nabla X^i = X^j \partial_j X^i + \Gamma^i{}_{jk} X^j X^k = \frac{d^2 x^i}{dt^2} + \Gamma^i{}_{jk} \frac{dx^j}{dt} \frac{dx^k}{dt} = 0.$$

We have used $X^i = \frac{dx^i}{dt}$ which defines integral curves of a vector field. The curve is uniquely determined by giving the initial point $p = x(0)$ and an initial tangent 'velocity' $\frac{dx^i}{dt}|_0 = X^i(0)$.

The geodesics generalize the notion of 'straight paths' of the familiar Euclidean geometry. Note that whether a given curve is a geodesic or not depends on the affine connection used in the definition of the covariant derivative.

[2]There are different routes to defining the curvature and there are many conventions!

[3]There is a slightly general definition of affine geodesics, namely that $X_{\parallel} \propto X$ which implies $X \cdot \nabla X^i = \xi X^i$. However by reparametrizing the integral curves, this can be reduced to the equation $X \cdot \nabla X^i = 0$. Our geodesics are strictly speaking *affinely parametrized affine geodesics.*

To summarise: In order to generalize the notion of differentiation to topological spaces, we need to introduce a differential structure on the topological space which turns it into a manifold. A manifold naturally leads to invariant quantities called tensors of ranks (p,q). In order to have derivatives of tensor fields to be tensors, we needed to equip the manifold with an affine connection which immediately lead to the notions of torsion, Riemann Tensor, Ricci tensor and affine geodesics[4].

In the next section we introduce the metric tensor and make contact with general relativity.

2.4 Metric tensor and Pseudo-Riemannian geometry

Consider a symmetric, rank (0,2) tensor field, $g_{ij}(x)$ on a manifold M. At any given point, it is a real symmetric matrix and so can be diagonalised. By making scaling coordinate transformations, the diagonal elements can be made ± 1 i.e. by coordinate transformations we can always arrange to have, at one point, $g'_{ij} = \eta_{ij} := \eta_i \delta_{ij},\ \eta_i = \pm 1, 0$. Let n_+, n_-, n_0 be the number of positive, negative and zero values of the η_i. These numbers are characteristic of the matrix g_{ij} and do not change with coordinate transformations. If the tensor is smooth (and hence continuous), then on any connected piece of the manifold, these numbers cannot vary from point to point and hence are characteristic of the tensor field itself.

If $n_0 = 0$, the matrix g_{ij} is invertible (or *non-degenerate* and its inverse is denoted by $g^{ij}, g^{ij}g_{jk} = \delta^i_k$. We will refer to a non-degenerate symmetric tensor of rank (0,2) as a *metric tensor*. Its inverse is a tensor of rank (2,0) and is called the inverse metric. The n_- is called the *index of the metric*, $(n_+ - \ n_-)$ is called the *signature of the metric*. A manifold with a metric is called a *(Pseudo-)Riemannian manifold*.

The metric tensors with index(g) = 0 are called *Riemannian Metrics* and the others are generically called *Pseudo-Riemannian*. Signature $\pm(n-2)$ metrics are called *Lorentzian*. We deal with Lorentzian metrics only and choose our conventions so that $n_+ = 1, n_- = n - 1$. Not all manifold admit Lorentzian metrics, the appendix gives basic existence results.

Availability of a metric (and its inverse) allows us to convert contravariant tensors to covariant ones and vice-a-versa - in short it allows *raising and lowering of indices*[5]. For instance, we can define $R_{ijkl} := g_{im}R^m{}_{jkl}$ and also the *Ricci Scalar*, $R := g^{ij}R_{ij}$. More important for us is the next property:

There is unique symmetric affine connection such that covariant derivative of the metric vanishes. This unique connection is called the *Riemann-Christoffel connection*. It is given explicitly by,

$$\Gamma^i{}_{jk}(g) \ := \ \frac{1}{2}g^{il}\left(g_{lj,k} + g_{lk,j} - g_{jk,l}\right).$$

To obtain this, write the defining equation $\nabla_k g_{ij} = 0$ three times by cyclically permuting the indices; add two of these equations and subtract the third one. Remember to use the

[4]There are other notions of derivatives producing tensors eg the Lie derivative which uses mappings of manifold but no other structure. Consequently it does not lead to new geometrical structures over and above what is provided by a manifold. Likewise, for $k-$forms, there is the notion of exterior derivative. Again, while very useful, it does not lead to new structures.

[5]A similar property is shared by symplectic manifolds - phases spaces of classical mechanics - which have a non-degenerate *antisymmetric* rank (0,2) tensor field.

property that the affine connection is symmetric. The more general case of non-zero torsion is given in the appendix.

The Riemann tensor of the Riemann-Christoffel connection has further additional properties, (a) R_{ijkl} is also anti-symmetric in the first two indices; (b) R_{ijkl} is symmetric under exchange of the first pair of indices with the second pair; (c) the Ricci tensor is symmetric and (iv) the *Einstein Tensor*, $G_{ij} := R_{ij} - \frac{1}{2}Rg_{ij}$ satisfies $\nabla_j G^{ij} = 0$, by virtue of the Bianchi identity. The symmetry properties also allow us to determine the *independent* components of Riemann tensor (for $n-$dimensional manifolds) as, $n^2(n^2-1)/12$. These properties are summarised in the appendix.

There are a couple of important points to note for future interpretation.

Apart from raising and lowering indices, the metric tensor also allows us to define a notion of "length" for tensors. For examples we can define the "norm" of a vector field X^i by $||X||^2 := g_{ij}X^iX^j$ and similarly for higher rank tensors with one factor of the metric for each index. For Riemannian metrics, these are really norms - are positive semi-definite. For Lorentzian metrics, these could be positive, negative or even null. The corresponding vector field is then called *Time-like, Space-like and Light-like (or null)* respectively. The covariant constancy of the metric (also called the metric compatibility condition on the affine connection), implies that the norm of a geodesic tangent vector is preserved and more generally, "inner products" of parallelly transported tensors are preserved along the vector field.

The result that a symmetric affine connection can be made to vanish at a point also applies to the Riemann-Christoffel connection and now it implies that the *first derivatives of the metric can be made to vanish* at a point. Since we can always choose coordinates so that a metric can be taken to be the Minkowski metric, diag(1, -1, -1, . . . , -1), it follows that *in a sufficiently small neighbourhood of any point, there exist coordinates such that the metric is the Minkowski metric up-to first order coordinate variations.* Notice that the 'size' of this neighbourhood is controlled by the curvature tensor.

Chapter 3

Relativistic Formulations and Einstein Equation

Let us recall points of Einstein's arguments. Special relativity already tells us to regard space + time as space-time i.e. permit coordinate transformations which mix the spatial coordinates and time. Equality of inertial and gravitational mass tells us that uniform gravity is same as uniform acceleration. So gravitational phenomena can be incorporated by *permitting non-inertial observers and hence arbitrary coordinate transformations.* The rotating platform argument tells us that an observer accelerated with respect to an inertial one will infer a non-Euclidean geometry. Hence gravitational field can be expected to affect determination of geometry. Since matter produces and is affected by gravity, matter must influence and be influenced by geometry. *So gravitational phenomena must be related to space-time geometry.* Since freely falling lift nullifies effects of gravity, it must permit use of at least locally Minkowskian geometry. The phenomena of tides, which occur on larger scales must be the real manifestation of gravity.

With the hind sight of Riemannian geometry we have learnt, it is pretty natural to think of a four dimensional, Lorentzian manifold as a model for physical space-time. It permits arbitrary coordinate transformations, it has the potential for non-fixed (and infinitely many) metrics to generalize the Minkowskian metric, the possibility of locally Minkowskian metric to incorporate freely falling lifts and it has the curvature, determined by the metric, to exhibit (large scale) tidal effects through the geodesic deviation equation i.e. gravity.

We have three different aspects to address: (a) How does the Riemannian geometry framework relate to physical measurements of lengths and time intervals? (b) How are the equations of mechanics, electrodynamics etc are to be modified to incorporate gravity and (c) what is the law that determines the metric in any given physical situation.

3.1 Metric and measurements of intervals

Our physical notion of assigning/measuring lengths is by laying down, practically rigid, unit rods. Similarly, the elapsed time intervals are determined by some clock. We can use these to assign coordinates (both space and time) to events. However, assigning coordinates by using unit rods is ambiguous in a general non-Euclidean geometry as is seen easily by considering the 2-sphere. For example, going x-units east and then y-units north reaches a different point than going y-units north and then x-units east i.e. the most obvious procedure fails to assign *unique* set of numbers to points. An alternative is to arbitrarily assign coordinate labels in a one-to-one manner eg sticking pieces of graph papers. This is exactly how coordinates are assigned to points on a manifold. But now we need an *interpretation* of the coordinate intervals in terms of physical measurements of lengths. This is provided by the metric tensor in the following manner.

We revert to the more common notation and index the local coordinates by *Greek letters,* $\mu, \nu, \ldots$, index the *spatial coordinates* by *Roman letters,* $i, j, \ldots$. The Greek indices take values 0, 1, 2, 3 while the Roman indices take values 1, 2, 3.

Consider small coordinate differences Δx^μ between two points in a local neighbourhood. Let $g_{\mu\nu}$ be the metric at one of the points. Then the quantity: $(\Delta s)^2 := g_{\mu\nu}\Delta x^\mu \Delta x^\nu$ is invariant under coordinate transformations. This is because although x^μ do *not* transform as tensors, small coordinate differences do - $\Delta y^\mu \approx \frac{\partial y^\mu}{\partial x^\nu}\Delta x^\nu$. Δs^2 is the candidate for representing physical lengths, and it is referred to as the *invariant interval* associated with the *coordinate interval* Δx^μ.

In a Lorentzian manifold, the invariant interval could be positive (time-like), negative (space-like) or null (light-like). The time-like intervals are given by elapsed times on a physical clock. Which clock? From special relativity we know that clocks in different states of motion relative to a given inertial observer, tick at different rates. So we need to specify which clock will measure the invariant interval.

If a clock is at rest relative to the local coordinates, then for any points on its world line $\Delta x^i = 0$. The corresponding invariant interval will then be $\Delta\tau^2 = g_{00}\Delta t^2$. If the coordinate system we are using happens to be locally Minkowskian, then $g_{00} = 1$ and the coordinate time interval coincides with the invariant interval. But Minkowskian system is the one used by freely falling observer, hence *the invariant time-like interval denotes the time recorded by a freely falling clock momentarily coinciding with an observer stationary with respect to the local coordinates ($\Delta x^i = 0$).* The relation of the reading on this clock with the local time coordinate is just $\Delta\tau^2 = g_{00}\Delta t^2$.

We see immediately that two observers, both stationary with respect to the given local coordinates but located at two different points P, Q and measuring invariant time-like intervals for the same Δt, will see different elapsed times on the corresponding freely falling clocks. Specifically, $\Delta\tau|_P = \sqrt{\frac{g_{00}(P)}{g_{00}(Q)}}\Delta\tau|_Q$. This is the *gravitational time dilation effect*, first verified by Pound-Rebka experiment in 1959. This plays a role in the cosmological red-shifts.

Similar considerations apply to spatial invariant intervals ($\Delta s^2 < 0$).

3.2 Principle of Covariance and Principle of Equivalence

Now we turn to the second aspect namely how to adapt equations of physics to a general Riemannian manifold. Some guiding principles are needed.

These are (a) the principle of general covariance and (b) the principle of equivalence. The former is stated as laws of physics be covariant under general coordinate transformations. The latter is stated with various versions: (i) equality of gravitational and inertial mass, (ii) laws of physics assume a form dictated by special relativity in the locally inertial frames. See the discussion given in Weinberg's book.

With the knowledge of differential geometry we have, it is clear that general covariance is the stipulation that laws of physics be expressed as tensor equations. This is eminently reasonable because these are the only types of equations which retain their form in arbitrary coordinate systems (alternatively these are the *only* ones that have an observer independent, intrinsic meaning). Recall that when one went from Galilean relativity to special relativity one had to modify the expressions for energy and momenta so as to identify them as components of 4-vector. The laws of mechanics and electrodynamics including the Lorentz force were expressed as tensorial expressions where tensors were understood to be Lorentz tensors. In analogy, the transition to general relativity stipulates use of general tensors. The promotion of Lorentz tensors to general tensors still leaves wide open the possibilities for modification in the expressions for the laws of physics. This is sought to be limited by the "medium version" of principle of equivalence. Whatever tensor equations that we propose should be such as to reduce to the expression given by special relativity when referred to locally inertial coordinates (freely falling observers). This suggests that we begin with the Lorentz covariant (special relativistic) form of the laws of physics and replace the partial derivatives by covariant derivatives ('comma → semicolon' rule). this is not free of ambiguities when higher order derivatives are involved. Furthermore, suppose a tensor expression involved the curvature. By specializing to locally inertial systems one can make the Riemann-Christoffel connection vanish at a point, but certainly not the curvature. Thus covariant derivatives will reduce to ordinary derivatives (as for special relativity) but curvature terms will still be present and these have no place in special relativity whose space time is Riemann flat. An example will illustrate the point.

Consider Maxwell equations in the special relativistic case:

$$\sum_{\text{cyclic } \mu\nu\lambda} \partial_\mu F_{\nu\lambda} = 0 \tag{3.1}$$

$$\partial^\mu F_{\mu\nu} = j_\nu \tag{3.2}$$

The first first set of equations allow us to define $F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu$. Choosing the Lorentz gauge, $\partial^\mu A_\mu = 0$, one can write the second set of equations as,

$$\partial^\mu \partial_\mu A_\nu = j_\nu \tag{3.3}$$

One can make these equations generally covariant quite simply by replacing the derivatives by covariant derivatives and declaring the vector potential, field strengths etc as tensors. The Bianchi identity still allows vector potential to be introduced. The equation (3.2) when expressed in terms of the vector potential in the Lorentz gauge takes the form:

$$\nabla^\mu \nabla_\mu A_\nu - R^\mu_\nu A_\mu = j_\nu \tag{3.4}$$

If however we covariantized the equation (3.3), we will get the same equation as above but without the Ricci tensor term. Thus we see that there are more than one ways of generating general tensor equations.

If we go to locally inertial frame (so that the Γ connection is zero at the origin of the inertial frame), then *neither* of these equations go over to the special relativistic equation. Thus neither the principle of covariance nor the principle of equivalence is useful here to select one or the other equation. What does select between these two candidate equations is the *conservation of* j_μ. The covariant divergence of the left hand side of (3.4) is identically zero.

We got this ambiguity in covariantizing the equations because while ordinary double derivatives commute, covariant double derivatives do *not* commute (except when acting on scalars). Their commutator contains curvature components. There is no ambiguity in the equations expressed in terms of the field strengths since only their single derivatives appear.

When covariantizing the Klein-Gordon equation, one does not generate curvature terms since on scalars the covariant derivatives commute. However, we can add a so called *non-minimal* coupling term of the form $\alpha R\phi$ which is consistent with both the principles.

As a quick application of principle of covariance and principle of equivalence let us deduce the equation for the freely falling point particle. In an inertial frame, a free particle obeys the equation,

$$\begin{aligned}
\frac{d^2x^\mu}{d\tau^2} &= 0 \quad \leftrightarrow \\
\frac{dx^\nu}{d\tau}\frac{\partial}{\partial x^\nu}\frac{dx^\mu}{d\tau} &= 0 \quad \leftrightarrow \\
v^\nu\partial_\nu v^\mu &= 0 \quad \text{whose covariant version is,} \\
v^\nu\nabla_\nu v^\mu &= 0 \quad \leftrightarrow \\
\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} &= 0 \quad \text{The geodesic equation.}
\end{aligned} \tag{3.5}$$

As a by product, We thus deduce that in the geometrical set up of Pseudo- Riemannian geometry, the trajectories of freely falling test (point) particles are given by the geodesics.

We now come to the final ingredient: What is the law that determines the space- time metric in a given physical context. In the next lecture we will use the knowledge of Newtonian gravity combined with principle of covariance to arrive at the Einstein equations.

3.3 'Derivation' of Einstein Equations

Although there is need to modify Newton's gravity, the modification has to be such as to make small refinements in the predictions since Newton's theory has been enormously successful. So we have to be able to reproduce the equations,

$$\frac{d^2x^i}{dt^2} = -\frac{\partial}{\partial x^i}\Phi \tag{3.6}$$

$$\nabla^2\Phi = 4\pi G\rho \tag{3.7}$$

when a suitable 'limit' is taken. Suitable limit means when we identify a space- time appropriate for describing motion of a non-relativistically moving test particle in the gravitational field of an essentially static body. Since this situation corresponds to the Galilean picture of space and time, we may expect that the geometry be time independent and very close to the Minkowskian geometry, i.e. $g_{\mu\nu} \approx \eta_{\mu\nu} + h_{\mu\nu}$.

Let us then imagine a large body producing Newtonian gravitational potential in which a test particle is 'freely falling' (recall that motion under the influence of only gravitational force is called a free fall). Let (t, x^i) denote a coordinate system in the vicinity of the large body which is at rest. Let $x^\mu(\lambda)$ denote the trajectory of the freely falling particle. Clearly it satisfies the geodesic equation. Now,

$$\begin{aligned}
\text{Non-relativistic test particle} &\Rightarrow \left|\frac{dx^i}{d\lambda}\right| << \left|\frac{dt}{d\lambda}\right| \qquad \Rightarrow \\
\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} &\approx \Gamma^\mu_{00}\left(\frac{dt}{d\lambda}\right)^2 \\
\text{time independence of geometry} &\Rightarrow \\
\Gamma^\mu_{00} &= -\frac{1}{2}g^{\mu\rho}\partial_\rho g_{00} \\
\text{Close to Minkowskian geometry} &\Rightarrow g^{\mu\nu} \approx \eta^{\mu\nu} - h^{\mu\nu} \qquad \Rightarrow \\
\Gamma^\mu_{00} \approx -\frac{1}{2}\eta^{\mu\rho}\partial_\rho h_{00} &
\end{aligned} \tag{3.8}$$

The $\mu = 0$ geodesic equation then implies that $t = a\lambda + b$ and by eliminating λ in favour of t the remaining equations become,

$$\frac{d^2x^i}{dt^2} = \frac{1}{2}\eta^{ij}\partial_j h_{00} = -\delta^{ij}\partial_j\left(\frac{1}{2}h_{00}\right) \tag{3.9}$$

Comparing with the Newtonian equation (3.6), we see that the metric component g_{00} gets identified with $1 + 2\Phi$. Thus we obtain a relation between metric and Newtonian potential. Newton's theory determines the potential given a mass density ρ via the Poisson equation. ρc^2 is then an energy density (using special relativity) which we know, again using special relativity, to be the 00 component of the energy-momentum tensor $T_{\mu\nu}$. Thus the Newtonian equation can be expressed as,

$$\nabla^2 g_{00} = \frac{8\pi G}{c^2}T_{00} \tag{3.10}$$

This is a highly suggestive form and appealing to covariance one can expect an equation relating matter distribution and geometry to be of the form,

$$\mathcal{F}_{\mu\nu}(g) = \frac{8\pi G}{c^2}T_{\mu\nu} \tag{3.11}$$

where, $\mathcal{F}_{\mu\nu}$ is a tensor constructed from the metric and should satisfy the following properties:

1. $\mathcal{F}_{\mu\nu}$ is a symmetric tensor built from the metric and its derivatives and is covariantly conserved, $\mathcal{F}^{\mu\nu}{}_{;\nu} = 0$;

2. It has at the most second derivative of the metric and is linear in the second derivative;

3. For $g_{\mu\nu} \approx \eta_{\mu\nu} + h_{\mu\nu}$ the equation should match with the Newtonian form of the equation (3.7).

These are very natural and reasonable demands. The first one is just consistency with the known general properties of the energy-momentum tensor (appeal to special relativity and principle of general covariance). The last one is where we expect Newtonian gravity to be recovered. The second one is a technical demand that could be justified on the basis of simplicity and the Newtonian form of the equation.

Recall that the Riemann-Christoffel connection is defined via the equations $g_{\mu\nu;\lambda} = 0$. This allows us to express first (ordinary) derivatives of the metric in terms of the connection and metric. Likewise, the second derivatives of the metric can be expressed in terms of the first derivatives of the connection, the connection and the metric. We need not go beyond due to the second requirement. The linearity in the second derivative of the metric implies that $\mathcal{F}$ should be built out of a 4th rank tensor involving first derivatives of the connection and products of connections. But, mathematically, the only such tensor is the Riemann curvature tensor! From this we also have the Ricci tensor and the Ricci scalar. This leads to the form, $\mathcal{F}_{\mu\nu} = aR_{\mu\nu} + bRg_{\mu\nu} + \Lambda g_{\mu\nu}$.

Now we impose the conservation requirement. Blissfully, the Riemann tensor already satisfies the differential Bianchi identities:

$$\begin{aligned} R^{\rho}_{\sigma\mu\nu;\lambda} + R^{\rho}_{\sigma\nu\lambda;\mu} + R^{\rho}_{\sigma\lambda\mu;\nu} &= 0 \quad \Rightarrow \\ R_{\mu}{}^{\nu}{}_{;\nu} &= \frac{1}{2} R_{;\mu} \end{aligned} \tag{3.12}$$

Conservation condition thus implies $(a/2 + b)R_{;\mu} = 0$. If we take gradient of the Ricci scalar to be zero, then the proposed equation will imply gradient of the trace of the energy-momentum tensor to be zero. This is not generally true and so would be an undue restriction on the matter properties. So we must have $b = -a/2$. This leads to the proposed equation of the form,

$$a(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}) + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^2} T_{\mu\nu} \tag{3.13}$$

We have yet to use the third requirement. For metric close to the Minkowskian metric, the curvature terms are all order h while the Λ term is order h^0 and so will dominate. For large static body (or non-relativistic matter) the spatial components of $T_{\mu\nu}$ are much much smaller than the time-time component. This is inconsistent with dominating Λ term. So if we are to recover the Newtonian limit, $\Lambda = 0$ should hold (or it should be exceedingly small to have escaped detection in Newtonian gravity, in which case we may continue to neglect it.) All that remains now is to determine a. The spatial components of $T_{\mu\nu}$ being very small implies that $R_{ij} \approx \frac{1}{2}Rg_{ij}$. This implies $\sum R_{ii} = (R/2)\sum g_{ii} \approx (R/2)\sum \eta_{ii} = -(3/2)R$. Furthermore the Ricci scalar can be likewise simplified as $R \approx R_{00} - \sum R_{ii} \Rightarrow R \approx -2R_{00}$. The equation then approximates to $aR_{00} \approx \frac{4\pi G}{c^2}T_{00}$. By substituting the metric in the definitions, a straightforward calculation yields $R_{00} \approx -(1/2)\delta^{ij}\partial_i\partial_j h_{00} \approx (1/2)\nabla^2 h_{00}$. Comparison then gives $a = 1$. Thus we finally arrive at the Einstein field equations as:

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^2} T_{\mu\nu} \tag{3.14}$$

A number of remarks are in order.

(1) The coefficient in front of $T_{\mu\nu}$ is about $1.86 \times 10^{-27} cm.gm^{-1}$. From cosmology, the estimate of the possible cosmological constant, Λ, is about $10^{-56} cm^{-2}$. So although strict

Newtonian limit would rule out Λ, Newtonian gravity it self is not tested to the extent of detecting presence of Λ. Thus logically the Λ term is admissible. In fact exactly the same logic can be applied to seek more general field equations. Our second requirement was based on the form of the Newtonian limit and simplicity. Simplicity is a matter of taste and level of accuracy of Newtonian gravity could permit higher derivatives of the metric and hence more general equations that could nonetheless show the same Newtonian limit. In this sense, to propose the above equation as 'the' equation governing determination of space-time metric is a postulation and not a 'derivation'.

(2) There are other alternative heuristic derivations of the Einstein equations. One is based on the comparison of 'tidal forces' as understood in the context of geometry. In the Newtonian picture, tidal forces imply relative acceleration between two nearby bodies, both moving in the same *inhomogeneous* gravitational field. This is given by the gradient of the force or double derivatives of the potential. In the geometrical context, one represents the free fall of the nearby bodies by two neigh boring geodesics and obtains an expression for their relative motion in terms of the Riemann tensor. Identifying the two expressions and referring to the Poisson equation, leads one to try $R_{\mu\nu} = \frac{4\pi G}{c^2} T_{\mu\nu}$. This in fact was the equation first considered by Einstein. But contracted Bianchi identity then implies that trace of $T_{\mu\nu}$ must be constant which is an unphysical demand on matter. The correction is of course replacing the Ricci tensor by the Einstein tensor. This still retains the identification of the tidal accelerations with the geodesic deviation at least for non-relativistically moving sources of Newtonian gravity. Details may be seen in Wald's book. Weinberg also has yet another derivation allowing the $\mathcal{F}_{\mu\nu}$ to be not just dependent on metric and its derivatives. We will now accept the Einstein equations as a law of nature and turn to study its properties and implications.

(3) Mathematically, The Einstein tensor is an expression involving double derivatives of the metric. The equations are thus a system of 10 non-linear, partial differential equations for the 10 unknown functions of 4 coordinates, $g_{\mu\nu}(x^\alpha)$. However the equations are not independent. They satisfy 4 differential identities implied by contracted Bianchi identities. There is also the freedom to make arbitrary coordinate transformations. To specify a solution therefore one has to specify coordinates either by explicit choice/procedure or implicitly by some 'coordinate conditions'. In this regards, the equations are similar to the Maxwell equations for the gauge potential.

Being partial differential equations, these are necessarily *local* determinations. The solutions thus admit the notion of 'extension' as well as 'matching' solutions found in different local regions. We will see examples of this in the context of the Schwarzschild solution.

(4) The equations, on the gravitational side, involve only the Ricci tensor and the Ricci scalar and *not* the full Riemann tensor. Likewise, on the matter side, only $T_{\mu\nu}$ is involved and not always the *other details of the matter constituents*. For example, we may have a perfect fluid made up of whatever types of 'fluid particles' but the form of the stress-tensor is still the same – different fluids being distinguished by different 'equations of states'. When taking a gas of photons as a source, one needs only to use the $T_{\mu\nu}$ described in terms of pressure and density without any reference to the underlying electromagnetic fields satisfying Maxwell equations. In particular this means that even if the stress tensor is zero in a region, the geometry in the same region is only Ricci-flat but non necessarily Riemann-flat. Empty space-time does not necessarily mean Minkowski space-time (which is Riemann- flat). This is good because it permits non-flat space-times in the vicinity of a body even in the region

not occupied by the body. As an aside we note that the Riemann tensor for n dimensional geometry has $\frac{1}{12}n^2(n^2-1)$ independent components. For $n=2$ this equals 1 which can be taken to be the Ricci scalar. Indeed the Einstein tensor vanishes identically for $n=2$. For $n=3$ the independent components are 6 in number and can be conveniently taken to the components of the Ricci tensor. In this case, Ricci-flat implies Riemann-flat. For $n \geq 4$, Riemann tensor has more components than the Ricci tensor and hence Ricci-flat does not imply Riemann- flat (though the converse is of course true).

(5) Newtonian gravity was described in terms of a single function satisfying a time independent Poisson equation. Time dependent gravitational fields are thus possible only due to the time variation of the matter density. In Einstein's theory, gravity is much richer and equations are dynamical. Thus even in the absence of sources one can have *propagating* gravitational disturbances – the gravitational waves which have been inferred indirectly by observations of binary pulsars but direct detection is still awaited.

(6) There is another aspect of the equations related to the conservation property. Bianchi identities imply that covariant divergence of the Einstein tensor is zero that in turn implies that the covariant divergence of the stress tensor is zero. From our experience with flat space-time, we are used to inferring a conservation law from a divergence-free 'current' e.g. $\partial_\mu J^\mu = 0 \Rightarrow \int_{vol} \partial_\mu J^\mu = \int_{surf} J^\mu dS_\mu = 0$ where Gauss's theorem has been used. However, if one has a covariant divergence of a tensor to be zero, one does not get a corresponding (integrated) conservation law except in some special cases. This happens essentially because an integration on an n-dimensional manifold can be defined *only* for n-forms whenever arbitrary change of integration variables is permitted (as on a manifold). When a metric is available, one has a natural invariant volume element available and one can define integration of 0- forms (scalars) on an n-dimensional manifold. This fact underlies Stoke's theorem that implies the Gauss's theorem that is used in deducing a conservation law from a divergence equation. One can check easily that invariant volume times the covariant divergence of a contravariant vector can be expressed as ordinary divergence of a vector density and for this the Stoke's theorem can be applied. In equations:

$$
\begin{aligned}
\sqrt{g}\nabla_\mu J^\mu &= \sqrt{g}\partial_\mu J^\mu + \sqrt{g}\Gamma^\mu_{\mu\nu}J^\nu \\
&= \sqrt{g}\partial_\mu J^\mu + \sqrt{g}(\partial_\mu \ell n \sqrt{g})J^\nu \\
&= \partial_\mu(\sqrt{g}J^\mu) \\
&= \partial_\mu(\sqrt{g}\epsilon^{\mu\nu_1\cdots\nu_{n-1}}\omega_{\nu_1\cdots\nu_{n-1}}) \\
&= \mathcal{E}^{\nu_1\cdots\nu_n}\partial_{\nu_1}\omega_{\nu_2\cdots\nu_n} \\
&= d\omega
\end{aligned}
\tag{3.15}
$$

For the stress tensor, however, these manipulations do not go through and hence the divergence equation does not lead to a conservation law. How did one get the usual conservation laws for special relativity? Recall that in the special relativistic context, the stress tensor is a tensor *only* relative to Lorentz transformations. Hence the only changes of integration variables permitted are the (constant) Lorentz transformations. For these restricted change of variables, the integration *is* well defined. Furthermore the space-time is flat and so in the Minkowskian coordinates the connection is zero. Covariant divergence is then same as the ordinary divergence.

A physical way of stating this lack of conservation law is to note that the connection term is like a gravitational force (since metric is analogous to the gravitational potential). Presence of these terms implies that tidal forces can always do work on the matter and thus one cannot expect a separate conservation for matter.

There are cases where the divergence equation does lead to conservation equation. If we have a space-time with a symmetry i.e. transformations generated by a Killing vector which leave the metric invariant, then one can define conserved quantities. For instance, if ξ_μ is a Killing vector field i.e. satisfies $\xi_{\mu;\nu} + \xi_{\nu;\mu} = 0$, then one can define $J^\mu := T^{\mu\nu}\xi_\nu$. Its covariant divergence is zero and because of the argument presented above the quantity $Q := \int_{hypersurface} J^\mu \xi_\mu$ is conserved as one changes the hypersurface orthogonal to the Killing vector. However, generic space-times do not admit any Killing vectors. For further discussion I refer you to the books [4, 5].

Chapter 4

Spherically Symmetric Space-times

4.1 The Schwarzschild (exterior) Solution

To get glimpses of the refined theory of gravity one should now obtain some solutions of the field equation and compare its properties with the Newtonian gravity. A simplest situation to consider is the geometry in the presence of a massive, spherically symmetric, non-rotating body. We know the Newtonian gravitational field out side the body, $\Phi(r) = -\frac{GM}{r}$. We would like to know the geometry i.e. the appropriate metric tensor. To obtain this we must first choose suitable coordinates. Most natural choice, also close to the Newtonian picture, is to imagine concentric spheres surrounding the body. The sphere's themselves are labelled by a label r while the points on each sphere is labelled by the usual spherical polar angles, θ, ϕ. We also choose some time label t.

Since the body is non-rotating (and not moving i.e. t is such that the body does not move) we expect the geometry to be time independent. Further spherical symmetry implies that the metric should not depend on the angles except for the 'metric' on the spheres. We therefore make the ansatz,

$$ds^2 = f(r)dt^2 - g(r)dr^2 - r^2(d\theta^2 + sin^2\theta d\phi^2) \tag{4.1}$$

Remarks:

(1) One can show that this ansatz can always be chosen for spherically symmetric, static space-times. To show this though would need the machinery of Killing vectors etc.

(2) The two dimensional surfaces defined by t = constant, r = constant, have the *induced metric* which is the standard ,metric on a sphere. The area of such a sphere is given by,

$$Area \ = \ \int \sqrt{g_{ind}}d\theta d\phi \ = \ \int \sqrt{r^4 sin^2\theta} d\theta d\phi \ = \ 4\pi r^2 \tag{4.2}$$

The label r can thus be defined as: $r := \sqrt{\frac{area}{4\pi}}$. r is consequently called the 'areal radial coordinate'.

(3) The three dimensional space defined by t = constant, has a metric similar to the standard Euclidean metric expressed in the spherical polar coordinates. It would be exactly that if

$g(r) = 1$. One will then have r also as the radius of the sphere. However $g(r)$ is yet to be determined, so we cannot interpret r as the usual radius.

(4) The metric is independent of t. By inspection we see that we can scale t by a constant factor and absorb it by redefining $f(r)$. This freedom will be fixed shortly.

At this stage we have made a judicious choice of coordinates and parameterized the metric in terms of only two functions of a single variable. We thus expect Einstein equations to reduce to ordinary differential equations that can always be solved. The procedure is to compute the connection and then the Ricci tensor components as expressions involving f, g, r. Since we are looking for the geometry outside of the body, we take $T_{\mu\nu} = 0$ and then it follows that the Ricci tensor must be zero.

Straight forward application of the definitions leads to ($\prime$ denotes $\frac{d}{dr}$) :

$\Gamma^{\alpha}_{\beta\gamma}$	t	r	θ	ϕ
tt	0	$\frac{1}{2}g^{-1}f'$	0	0
tr	$\frac{1}{2}f^{-1}f'$	0	0	0
$t\theta$	0	0	0	0
$t\phi$	0	0	0	0
rr	0	$\frac{1}{2}g^{-1}g'$	0	0
$r\theta$	0	0	r^{-1}	0
$r\phi$	0	0	0	r^{-1}
$\theta\theta$	0	$-rg^{-1}$	0	0
$\theta\phi$	0	0	0	$cot\theta$
$\phi\phi$	0	$-g^{-1}rsin^2\theta$	$-sin\theta cos\theta$	0

$$\begin{aligned}
-R_{tt} &= -\frac{f''}{2g} + \frac{1}{4}\left(\frac{f'}{g}\right)\left(\frac{g'}{g} + \frac{f'}{f}\right) - \frac{f'}{rg} \ ; \\
-R_{rr} &= \frac{f''}{2f} - \frac{1}{4}\left(\frac{f'}{f}\right)\left(\frac{g'}{g} + \frac{f'}{f}\right) - \frac{g'}{rg} \ ; \\
-R_{\theta\theta} &= -1 + \frac{r}{2g}\left(-\frac{g'}{g} + \frac{f'}{f}\right) + g^{-1} \ ; \\
R_{\phi\phi} &= sin^2\theta \ R_{\theta\theta}; \qquad \text{all other components are zero.}
\end{aligned} \tag{4.3}$$

Clearly, $g^{-1}R_{rr} + f^{-1}R_{tt} = 0$ implies $fg =$ constant. In view of the scaling freedom in the definition of t we can take this constant to be equal to 1 [1]. The $R_{\theta\theta} = 0$ implies $rf' = 1 - f$ which can be immediately integrated to give $f(r) = 1 - R_S r^{-1}$ where R_S is an integration constant. If we appeal to the Newtonian limit for large r, we see that $f(r) = g_{00} = 1 + 2\Phi(r)$ which gives the identification, $R_S = 2GM$. Thus we have the famous Schwarzschild solution

[1] This has the effect of making the metric approach the standard Minkowski metric for r tending to infinity.

(1916):

$$ds^2 = \left(1 - \frac{2GM}{r}\right) dt^2 - \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + sin^2\theta d\phi^2) \tag{4.4}$$

Note that this describes the space-time outside the body (i.e. $r >$ the physical radius of the body) and is called the exterior Schwarzschild space- time. A natural length scale has crept in via the constant of integration, R_S which in the usual units is given by $R_S = \frac{2GM}{c^2}$ and is known as the Schwarzschild radius of the body of mass M. Evidently, for $r >> R_S$, the metric can be expressed as,

$$\begin{aligned}
ds^2 &= \left(1 - \frac{R_S}{r}\right) dt^2 - \left(1 + \frac{R_S}{r} + \frac{R_S^2}{r^2} + \cdots\right) dr^2 - r^2 d\Omega^2 \\
&= \left[dt^2 - dr^2 - r^2 d\Omega^2\right] \ + \ \left[-\frac{R_s}{r}\left(dt^2 + dr^2\right) + o\left(\left(\frac{R_S}{r}\right)^2\right)\right] \qquad (4.5) \\
&= \text{Minkowski metric} \ + \ \text{deviations}
\end{aligned}$$

To get a feel, let us put in some numbers. For our Sun:

$$R_S \approx \frac{2 \times (6.67 \times 10^{-8}) \times (2 \times 10^{33})}{(3 \times 10^{10})^2} \approx 3 \text{ km} \tag{4.6}$$

For contrast, the physical radius of the Sun is about 6,00,000 km. Thus already just outside the Sun, the deviation from Minkowskian geometry is of the order of 1 part in 10^5. For earth the deviation is about 1 part in 10^9. General relativistic corrections are thus very small. No wonder Newtonian gravity worked so well. For more compact objects such as white dwarfs and neutron stars the deviation factors are about 10^{-3} and 0.5.

This simple solution is useful for practical matters such as solar system tests of general relativity as well as for hints at the exotic aspects of GR such as black holes. We will first study the non-exotic aspects. We will take $r >> R_S$ and study the small corrections implied by GR.

4.1.1 Geodesics

The first aspects to study are the geodesics. Let $(t(\lambda), r(\lambda), \theta(\lambda), \phi(\lambda))$ denote a geodesic. Using over-dot to denote derivative with respect to λ and $'$ to denote derivative w.r.t. r and using the table of Γ's, we see that,

$$0 = \ddot{t} + \frac{f'}{f}\dot{r}\,\dot{t} \tag{4.7}$$

$$0 = \ddot{r} + \frac{f'}{2g}\dot{t}^2 + \frac{g'}{2g}\dot{r}^2 - \frac{r}{g}\dot{\theta}^2 - \frac{r sin^2\theta}{g}\dot{\phi}^2 \tag{4.8}$$

$$0 = \ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} - sin\theta cos\theta\dot{\phi}^2 \tag{4.9}$$

$$0 = \ddot{\phi} + \frac{2}{r}\dot{r}\dot{\phi} + 2cot\theta\dot{\theta}\dot{\phi} \tag{4.10}$$

It is clear that $\theta =$ constant is possible only for $\theta = \pi/2$. These are the equatorial geodesics. The equations simplify to:

$$0 = \ddot{t} + \frac{f'}{f}\dot{r}\,\dot{t} \qquad \Rightarrow \qquad \dot{t}f \equiv E \quad (\text{ a } \textit{positive} \text{ constant }) \tag{4.11}$$

$$0 = \ddot{r} + \frac{f'}{2g}\dot{t}^2 + \frac{g'}{2g}\dot{r}^2 - \frac{r}{g}\dot{\phi}^2 \tag{4.12}$$

$$0 = \ddot{\phi} + \frac{2}{r}\dot{r}\dot{\phi} \qquad \Rightarrow \qquad r^2\dot{\phi} \equiv EL \quad (\text{ a constant. }) \tag{4.13}$$

The radial equation can be integrated once to yield,

$$g\dot{r}^2 + E^2\left(\frac{L^2}{r^2} - \frac{1}{f}\right) \equiv -E^2\kappa \quad (\ \kappa \text{ is a constant. }) \tag{4.14}$$

It is easy to see by substitution that,

$$\left(\frac{ds}{d\lambda}\right)^2 = E^2\kappa \quad (\ \geq 0\ ) \tag{4.15}$$

κ is positive for time-like geodesics (material test bodies such as planets) and is zero for light-like geodesics. One can eliminate λ in favor of t by using $d\lambda = f dt/E$ to get,

$$r^2\frac{d\phi}{dt} = Lf \tag{4.16}$$

$$\frac{g}{f^2}\left(\frac{dr}{dt}\right)^2 - \frac{1}{f} + \frac{L^2}{r^2} = -\kappa \tag{4.17}$$

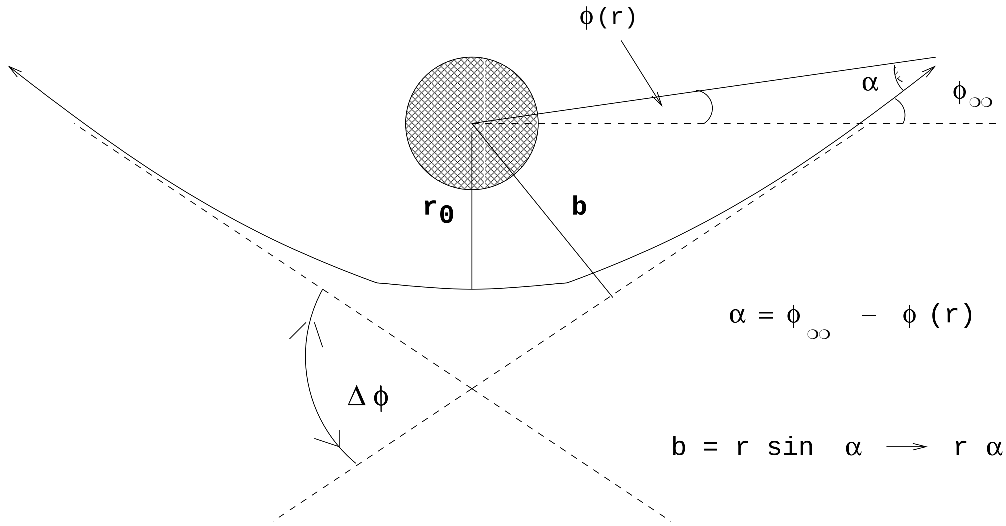
$$\left(\frac{ds}{dt}\right)^2 = \kappa f^2 \tag{4.18}$$

Notice that these equations are independent of E. The relevant constants of integration are κ and L. To get the orbit equation, we eliminate t in favor of ϕ using $dt = \frac{r^2}{Lf}d\phi$ to get,

$$0 = \frac{g}{r^4}\left(\frac{dr}{d\phi}\right)^2 + \frac{1}{r^2} + \frac{1}{L^2}\left(\kappa - \frac{1}{f}\right) \qquad \text{or} \tag{4.19}$$

$$\phi(r) = \pm\int dr\frac{\sqrt{g}}{r^2}\left(\frac{1}{L^2}(f^{-1} - \kappa) - \frac{1}{r^2}\right)^{-\frac{1}{2}} \tag{4.20}$$

These are the general set of equations for geodesics. These are essentially characterized by two constants, κ, L. We can now distinguished two types of orbits, bounded and unbounded (scattering). The relevant orbit parameters for bounded orbits are the maximum and the minimum values, $r_\pm$ and relevant question is whether the orbit precesses or not. For unbounded orbits the relevant parameters are asymptotic speed (or energy) and the impact parameter or the distance of closest approach and the important question is to obtain the scattering angle.



4.1.2 Deflection of light

Let us consider the scattering problem first. The geometry is shown in the figure. Asymptotically r is very large and thus $f, g \approx 1$. The incoming radial speed v, defined as $v := -\frac{drcos\alpha}{dt}$ is given by $v \approx -\frac{dr}{dt}$. Radial equation then implies $\kappa = 1 - v^2$. Likewise the impact parameter $b := rsin\alpha \approx r\alpha$. Differentiating w.r.t. t and using the angular equation one finds $L = bv$. It is convenient to further eliminate L in favor the distance of closest approach, r_0, defined by $\frac{dr}{d\phi} = 0$. This yields,

$$|L| = r_0\sqrt{f(r_0)^{-1} - 1 + v^2} \qquad \text{and the } \phi \text{ integral becomes,} \tag{4.21}$$

$$\phi(r) = \phi_\infty + \int_r^\infty dr \frac{\sqrt{g}}{r^2}\left[\frac{1}{r_0^2}\frac{f(r)^{-1} - 1 + v^2}{f(r_0)^{-1} - 1 + v^2} - \frac{1}{r^2}\right]^{-\frac{1}{2}} \tag{4.22}$$

We have obtained the expression in terms of directly observable parameters, v and r_0. The scattering or deflection angle is defined as $\Delta\phi := 2|\phi(r_0) - \phi_\infty| - \pi$.

For scattering of light, we have to take $v^2 = 1$ (recall that we are using units in which $c = 1$). The integral still needs to be done numerically.

Observe that so far we have used only the spherical symmetry and staticity of the metric and *not* the particular f, g of the Schwarzschild solution. If we only use the qualitative fact that the Schwarzschild solution is asymptotically flat i.e. approaches the Minkowskian metric for $r >> R_S$, then we can use a general form for f, g as an expansion in terms of the ratio r/R_S. We can now use the fact that for solar system objects $\frac{R_S}{r} << 1$ even for grazing scattering and can thus evaluate the integral to first order in $\frac{R_S}{r}$. It is convenient to use the so-called Robertson expansion for the f, g function instead of the exact expression. This is parameterized as:

$$f(r) = \left(1 - \frac{R_S}{r} + \cdots\right)$$

$$g(r) = \left(1 + \gamma\frac{R_S}{r} + \cdots\right) \tag{4.23}$$

For Schwarzschild solution, i.e. for GR, $\gamma = 1$. Then to first order one computes,

$$\Delta\phi = \frac{2R_S}{r_0}\left(\frac{1+\gamma}{2}\right) = \left(\frac{R_\odot}{r_0}\right)\left(\frac{2R_S}{R_\odot}\right)\left(\frac{1+\gamma}{2}\right) \tag{4.24}$$

Putting in the values for the solar radius, $R_\odot \approx 7 \times 10^5$ km and $R_S \approx 3$ km one gets,

$$\Delta\phi_\odot \approx 1.75'' \left(\frac{1+\gamma}{2}\right)\left(\frac{R_\odot}{r_0}\right) \tag{4.25}$$

This prediction was first confirmed by Eddington during the total solar eclipse in 1919. It has since been tested many times with improved accuracies. Current limits on γ put $\gamma = 1$ to within 10^{-4} [7].

4.1.3 Precession of perihelia

Now let us consider bounded orbits. Clearly any such orbit will have some maximum and minimum values of r, possibly equal in case of a circular orbit. These are easily determined from the orbit equation by setting $\frac{dr}{d\phi} = 0$. This is a cubic equation in r and so has either 1 or 3 real roots. The case where there is only one root corresponds to an unbounded orbit with a single r_{min}. The case of three roots is the one that admits bounded orbits. The maximum (r_+) and the minimum (r_-) are determined by,

$$0 = \frac{1}{r_\pm^2} - \frac{1}{L^2 f_\pm} + \frac{\kappa}{L^2}, \qquad f_\pm := f(r_\pm), \quad \Rightarrow \tag{4.26}$$

$$\kappa = \frac{\frac{r_+^2}{f_+} - \frac{r_-^2}{f_-}}{r_+^2 - r_-^2}; \tag{4.27}$$

$$L^2 = \frac{\frac{1}{f_+} - \frac{1}{f_-}}{\frac{1}{r_+^2} - \frac{1}{r_-^2}}; \qquad \text{also,} \tag{4.28}$$

$$\phi(r) = \phi(r_-) + \int_{r_-}^{r} \frac{dr}{r^2}\sqrt{g}\left\{\frac{1}{L^2 f} - \frac{\kappa}{L^2} - \frac{1}{r^2}\right\}^{-\frac{1}{2}} \tag{4.29}$$

The orbit is said to be non-precessing if the accumulated change in ϕ as one makes one traversal $r_- \to r_+ \to r_-$ equals 2π. Otherwise the orbit is said to be precessing with a rate,

$$\text{Precession per revolution} \equiv \Delta\phi := 2|\phi(r_+) - \phi(r_-)| - 2\pi. \tag{4.30}$$

Now one substitutes for κ, L^2 in terms of the orbit characteristics, $r_\pm$ and evaluates the integrals. This again has to be done numerically. However again for solar system objects, one can compute the precession to first order in R_S. For this one again uses the Robertson parameterization ($\gamma = 1, \beta = 1$ for Schwarzschild) ,

$$\begin{aligned} g(r) &= 1 + \gamma\frac{R_S}{r} + \cdots \\ f(r) &= 1 - \frac{R_S}{r} + \frac{(\beta-\gamma)}{2}\left(\frac{R_S}{r}\right)^2 + \cdots \qquad \Rightarrow \\ f^{-1}(r) &= 1 + \frac{R_S}{r} + \frac{(2-\beta+\gamma)}{2}\left(\frac{R_S}{r}\right)^2 + \cdots \end{aligned} \tag{4.31}$$

Now a bit of mathematical jugglery leads to the formula [4],

$$\Delta\phi = (2 + 2\gamma - \beta)\pi R_S \left[\frac{1}{2}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)\right] \tag{4.32}$$

The quantity in the square brackets is called the semi-latus-rectum. Usually astronomers specify an orbit in terms of the semi-major axis a, and the eccentricity e, defined by $r_\pm = (1 \pm e)a$. The semi-latus rectum, ℓ, is then obtained as $\ell = a(1 - e^2)$. The precession per revolution is then given by,

$$\Delta\phi = 3\pi\frac{2GM}{c^2}\frac{1}{\ell} \tag{4.33}$$

The precession will be largest for largest R_S and smallest ℓ and in our solar system the obvious candidates are Sun and Mercury. For Mercury $\ell \approx 5.53 \times 10^7$ km while R_S for the Sun is about 3 km. Mercury makes about 415 revolutions per century. These lead to *general relativistic* precession of Mercury per century to be about $43''$. This has also been confirmed. Observationally, determining the precession is tricky since many effects such as perturbation due to other planets, non-sphericity (quadrupole moment) of Sun also cause precession. For a more detail discussion of these, please see Weinberg's book.

4.2 Interiors of Stars

4.2.1 General Equations and Elementary Analysis

Let us now turn attention from vacuum solutions to non-vacuum solutions still continuing with compact bodies with spherical symmetry and staticity. What do we take for the stress tensor?

The most general stress tensor consistent with spherical symmetry and staticity can be constructed as follows. Given the metric ansatz, we can define 4 orthonormal vectors as:

$$\begin{array}{lllllll} e_0^\mu & := & \frac{1}{\sqrt{f}}(1,0,0,0) & , & e_1^\mu & := & \frac{1}{\sqrt{g}}(0,1,0,0) \\ e_2^\mu & := & \frac{1}{\sqrt{r}}(0,0,1,0) & , & e_3^\mu & := & \frac{1}{\sqrt{rsin\theta}}(0,0,0,1) \end{array} \tag{4.34}$$

Any stress tensor can then be written as $T^{\mu\nu} := \rho_{ab} e_a^\mu e_b^\nu$ with ρ_{ab} symmetric. Spherical symmetry and staticity implies $\rho_{ab} = diag(\rho_0, \rho_1, \rho_2, \rho_3)$ with $\rho_2 = \rho_3$. All these are functions only of r.

The Einstein equations can now be written down. Previously, the vacuum case we could just use Ricci tensor equal to zero. Now we must use the Einstein tensor. One gets only three non-trivial equations coming from G_{00}, G_{11} and G_{22}. The third one is a second order equation and can be traded for the conservation equation that is first order. Thus we can arrange our equations as 3 first order equations [8]:

$$r\frac{dg}{dr} = -g(g-1) + (8\pi\rho_0 r^2)g^2 \qquad (G_{00} = 8\pi T_{00}) \tag{4.35}$$

$$r\frac{df}{dr} = f(g-1) + (8\pi\rho_1 r^2)fg \qquad (G_{11} = 8\pi T_{11}) \tag{4.36}$$

$$r\frac{d\rho_1}{dr} = 2(\rho_2 - \rho_1) - \frac{\rho_0 + \rho_1}{2}r\frac{d\ell nf}{dr} \qquad \text{(Conservation equation)} \tag{4.37}$$

The (00) equation can be solved for $g(r)$ in terms of $\rho_0(r)$ as:

$$m(r) - m(r_1) := 4\pi \int_{r_1}^{r} \rho_0(r')r'^2 dr' \quad , \quad g(r) := \left(1 - \frac{2m(r)}{r}\right)^{-1} \tag{4.38}$$

Substituting the (11) equation in the conservation equation will give an equation involving only the $\rho's$. Once these are solved we can determine $f(r)$ from the (11) equation. We already see that we have to provide further information in order the equations can be solved. This involves specification of the stress tensor. If stress tensor is that of electromagnetism (spherically symmetric and static of course) then $\rho_2 = -\rho_1 = \rho_0 = Q^2/r^4$. Using this leads to the Reissner-Nordstrom solution. For the case of perfect fluid we have $\rho_0 \equiv \rho, \rho_1 = \rho_2 \equiv P$ together with an equation of state, $P(r) = P(\rho(r))$. Now our equation system is determined.

For the interior solution we take $r_1 = 0$ and $m(r_1) = 0$ to avoid getting a "conical singularity" at $r = 0$. There is supposed to be a maximum value R at which the density and the pressure is expected to drop to zero. This R is of course the radius of our static body.

(If ρ_0 is not integrable at $r = 0$, as for the Reissner-Nordstrom case, then the solution should be understood as an exterior solution. In such a case we can take r_1 to be ∞ and $m(r_1) \equiv M$. You can construct the solution easily. It is also a black hole solution.)

With these we can write the final equations as:

$$m(r) \;:=\; 4\pi \int_0^r \rho(r')r'^2 dr' \quad , \quad g(r) := \left(1 - \frac{2m(r)}{r}\right)^{-1} \tag{4.39}$$

$$r\frac{dP(\rho(r))}{dr} \;=\; -\left(\rho + P(\rho)\right)\frac{m(r) + 4\pi P(\rho)r^3}{r - 2m(r)} \qquad \text{(T-O-V eqn.)} \tag{4.40}$$

$$r\frac{d\ell n f}{dr} \;=\; 2\frac{m(r) + 4\pi P(\rho)r^3}{r - 2m(r)} \tag{4.41}$$

The 'T-O-V' equation stands for Tolman-Oppenheimer-Volkoff equation of hydrostatic equilibrium. The corresponding Newtonian hydrostatic equilibrium equation is obtained by taking $P << \rho, m(r) << r$. In practice, these equations are solved by starting with some arbitrary central density and corresponding pressure, $\rho(0), P(0) = P(\rho(0))$ and integrating the T-O-V equation together with the $m(r)$. One continues integration till a value $r = R$ at which the density and pressure vanish. Once $\rho, m(r)$ are known the last equation can be integrated. Its boundary condition is chosen so that the interior solution matches with the exterior Schwarzschild solution. Clearly, mass of such a body is just $M = m(R)$ while its surface is at $r = R$.

Note that $\rho(0)$ and the equation of state are inputs while R and M are the outputs. Since the equations are non-linear in ρ, we may *not* find a 'surface of body' for *all choices of the central density and/or for all possible equations of states*. If we do, then R, M have a complicated dependence on the central density. There is then an implicit relation between the mass and radius of a star. The possibility of non-finite size solution makes the question of *stability* of star quite non-trivial.

An instructive example which can be done exactly is the so-called incompressible fluid defined as P is independent of ρ and $\rho = \hat{\rho}$, a constant, for $r \leq R$ and zero otherwise. Then,

$m(r) = (4\pi\hat{\rho}r^3)/3$ and,

$$P(r) = \hat{\rho}\left[\frac{(1-2M/R)^{1/2} - (1-2Mr^2/R^3)^{1/2}}{(1-2Mr^2/R^3)^{1/2} - 3(1-2M/R)^{1/2}}\right] \quad (4.42)$$

$$P(0) = \hat{\rho}\left[\frac{(1-2M/R)^{1/2} - 1}{1 - 3(1-2M/R)^{1/2}}\right] \quad (4.43)$$

The central pressure thus blows up for $R = 9M/4$! There can be no body with uniform density and $M > 4R/9$. A corresponding calculation with Newtonian gravity has no such limit. Einstein's gravity has drastic consequences for stellar equilibria. It turns out that assuming only that the density is a non- negative monotonically decreasing function of r, the maximum mass possible for any given radius must be less than $4R/9$. That there must be such a limit follows by noting the $g(r)$ must be positive to maintain the Riemannian nature of the spatial metric. This already implies $M < R/2$. Further requiring $f(r)$ remain positive so as to maintain staticity sharpens this limit [5].

Real stars are of course not static. There are a variety of complicated processes going on in a star. Over a certain period however a star can be assumed to approximately in equilibrium. If it is also close to being spherical and at most slowly rotating then such a star can be well modelled by an interior Schwarzschild solution. These solutions are thus useful for identifying approximate *equilibrium* states of stars.

However, various possible equilibria may not be *stable*, a small perturbation in the central density parameter $\rho(0)$ may result in a solution without a finite size (a non-star solution). To appreciate this issue, These instabilities contribute to the evolution of stars [4].

Our current understanding of stages in stellar evolution is as follows. An ordinary body supports itself against gravitational collapse by simple mechanical forces. If it is massive so that gravitational forces are significant to over come the mechanical forces, then collapse proceeds, contracting and heating the body. Up to a certain size, thermal pressure is enough to balance gravity. For a still heavier body, nuclear fusion starts and a star is born.

The subsequent evolution depends on the range of masses of the star. The mass controls what happens after the hydrogen is mostly used up. If the mass is at most few times the solar mass then the star passes through a so called red giant phase at the core of which is a *white dwarf*. If the mass is high, then subsequent contraction reaches higher temperatures to ignite further nuclear fusions eventually leading up to Iron, Nickel. At this stage the core collapses producing a shock wave which throws off the mantle in a *supernova* explosion. Its remnant is either a *neutron star* or a *black hole*.

In the case of compact left over core which is *not* a black hole, the core is supported by what is called a *degeneracy pressure*. This arises from the quantum mechanical behavior of fermions (electrons, neutrons). The Pauli exclusion principle prevents fermions to occupy the same quantum state effectively resulting in a pressure. For the white dwarfs this pressure is provided by the electrons while for the neutron stars it is provided by the neutrons. The central densities are about $10^7 gm/cc$ and $10^{15} gm/cc$ respectively.

These two possible equilibrium states however are stable only up to an upper mass limit, the *Chandrasekhar limit*. For the white dwarfs it is about $1.4M_\odot$ while for the neutron stars it is about $2.5 - 3M_\odot$. The uncertainties are due to lack of knowledge about the equation of state for nuclear matter at high densities.

If a core is more massive than these limits, then presently there is *no* known mechanism for gravity to be resisted. Such a core must undergo a *complete gravitational collapse* to become a black hole (or a naked singularity?).

While these are details proper to astrophysics, suffice it to say that observationally one knows white dwarfs, neutron stars and believes that black holes exist too. Equally well, one does not yet have a good solution describing a rapidly rotating star matched with a suitable exterior solution (the Kerr solution is not adequate).

Chapter 5

Black Holes

5.1 The Static Black Holes

5.1.1 The Schwarzschild Black Hole

Imagine now that the gravitational collapse has proceeded so far that candidate 'surface of a star' is inside the sphere of radius equal to the Schwarzschild radius. The exterior Schwarzschild solution is thus now valid also for $R_\odot \leq r \leq R_S$. Here we meet the famous Schwarzschild singularity that caused enormous confusion in the early history. Quit simply, for $r = R_S$, g_{tt} vanishes and g_{rr} blows up. However if one computes the Riemann curvature components, then they are perfectly well behaved at $r = R_S$. Hence physical effects of gravity such as tidal forces are all finite. The apparent singularity is thus a computational artifact, more precisely it signals breakdown of the coordinate system.

For instance if we consider the flat Euclidean plane and express the Euclidean metric of Cartesian system in terms of the (r, θ) coordinates, then $g_{rr} = 1, g_{\theta\theta} = r$. Now the inverse metric is singular at the origin, $r = 0$. We know this is artificial because we know that (r, θ) is not a good coordinate system at the origin. For every $r > 0, 0 \leq \theta < 2\pi$, one has a one-to-one correspondence with points in the plane, but as $r \to 0$ *no unique θ can be assigned to the origin in a continuous manner.* One has to take the precise definitions of coordinate systems (charts) seriously.

Let us recall that given a vector field one has its integral curves defined by $X^\mu = \frac{dx^\mu}{d\lambda}$. If it so happens that as we move along the integral curves, the metric does not change, then the vector field is said to be a Killing vector and it satisfies the equation: $X_{\mu;\nu} + X_{\nu;\mu} = 0$. The parameter λ of the integral curves itself can be taken as one of the local coordinates and metric will be manifestly independent of this coordinate. Returning to our plane, we observe that $\xi^i \partial_i := \partial_\theta = -y\partial_x + x\partial_y$ is a Killing vector (expressing the rotational symmetry of the Euclidean metric). This is easiest to see in the Cartesian system where the connection is zero and $\xi_{i,j} + \xi_{j,i} = 0$ follows. Its $(\text{norm})^2$ is r^2 which vanishes at $r = 0$. The angular coordinate θ is the parameter of integral curves of the Killing vector. The vanishing of the norm means that the vector field vanishes there (we are in Euclidean geometry) and hence the angular coordinate cannot be defined. Some thing similar happens at $r = R_S$.

One of the Killing vector expressing stationarity of the metric is $\xi = \partial_t$ and its $(\text{norm})^2$ is just g_{tt} which vanishes at $r = R_S$. Since the metric is of Lorentzian signature, zero norm does not mean the vector vanishes. But it does mean that the vector ceases to be *time-like* which is needed to interpret t as time (as opposed to one of the spatial coordinate). In the case of the plane, the coordinate failure is cured by using the Cartesian coordinates which are perfectly well defined everywhere. Likewise one has to look for a different set of coordinates which are well behaved around $r = R_S$. These are usually (for effectively two dimensional space-time) discovered by looking at radial null geodesics crossing the $r = R_S$ sphere and choosing the affine parameters of these geodesics as new coordinates.

To arrive at these new coordinates, write the metric in the form,

$$\begin{aligned} ds^2 &= \left(1 - \frac{R_S}{r}\right)\left\{dt^2 - \left(1 - \frac{R_S}{r}\right)^{-2} dr^2\right\} - r^2 d\Omega^2 \\ &:= \left(1 - \frac{R_S}{r}\right)\left\{dt^2 - dr_*^2\right\} - r^2 d\Omega^2 \end{aligned} \tag{5.1}$$

Solving for $r_*(r)$ and choosing $r_*(0) = 0$ without loss of generality gives,

$$r_*(r) = r + R_S \ell n \left|\frac{r - R_S}{R_S}\right| \tag{5.2}$$

Notice that r_* ranges monotonically from $-\infty$ to ∞ as r ranges from R_S to ∞. This new radial coordinate r_* is called the *tortoise coordinate*. The (t, r_*) part of the metric is clearly conformal to the Minkowskian metric whose null geodesics are along the light cone $t = \pm r_*$. Introducing new coordinates (u, v) via

$$\begin{aligned} t &:= \frac{1}{2}(\epsilon_u u + \epsilon_v v) \quad , \quad r_* := \frac{1}{2}(-\epsilon_u u + \epsilon_v v) \quad , \quad \epsilon_u, \epsilon_v = \pm 1, \\ u &= \epsilon_u (t - r_*) \qquad , \quad v = \epsilon_v (t + r_*) \end{aligned} \tag{5.3}$$

implies $dt^2 - dr_*^2 = \epsilon_u \epsilon_v du dv$ and $ds^2 = (1 - R_S/r)\epsilon_u \epsilon_v du dv - r^2 d\Omega^2$. So retain the signature of the metric and noting that the pre-factor is *positive* for $r > R_S$ requires $\epsilon_u = \epsilon_v = \pm 1$.

As r_* varies from $-\infty$ to ∞ ($r \in (R_S, \infty)$), $u \in (\infty, -\infty), v \in (-\infty, \infty)$ for $\epsilon_u = +1$ (and oppositely for $\epsilon = -1$). Taking $\epsilon_u = 1$ for definiteness and substituting for r_* one sees that,

$$\left(1 - \frac{R_S}{r}\right) = \frac{R_S}{r} e^{-r/R_S} e^{(v-u)/(2R_S)} \tag{5.4}$$

$$\begin{aligned} ds^2 &= \frac{R_S}{r} e^{-r/R_S} \left(e^{-u/(2R_S)} du\right)\left(e^{v/(2R_S)} dv\right) - r^2 d\Omega^2 \\ &= \frac{4R_S^3}{r} e^{-r/R_S} dU dV - r^2 d\Omega^2 \quad , \quad \text{with} \end{aligned} \tag{5.5}$$

$$\begin{aligned} U &:= -e^{-u/(2R_S)} &&:= T - X \\ V &:= e^{v/(2R_S)} &&:= T + X \end{aligned} \tag{5.6}$$

$$-UV = \left(\frac{r}{R_S} - 1\right) e^{r/R_S} = X^2 - T^2 \tag{5.7}$$

The coordinates T, X defined in (5.6) are known as the Kruskal coordinates. Their relation to the Schwarzschild coordinates (t, r) is summarised below.

$$F(r) = X^2 - T^2 := \left(\frac{r}{R_S} - 1\right) e^{r/R_S}$$

$$
\begin{aligned}
\frac{t}{R_S} &= 2\ tanh^{-1}\left(\frac{T}{X}\right) && (5.8)\\
X &= \pm\sqrt{|F(r)|}\ cosh\left(\frac{t}{R_S}\right) \\
T &= \pm\sqrt{|F(r)|}\ sinh\left(\frac{t}{R_S}\right) && (5.9)\\
ds^2 &= \frac{4\ R_S^3\ e^{-r/R_S}}{r}\left(dT^2 - dX^2\right) - r^2(T,X)\ d\Omega^2 && (5.10)
\end{aligned}
$$

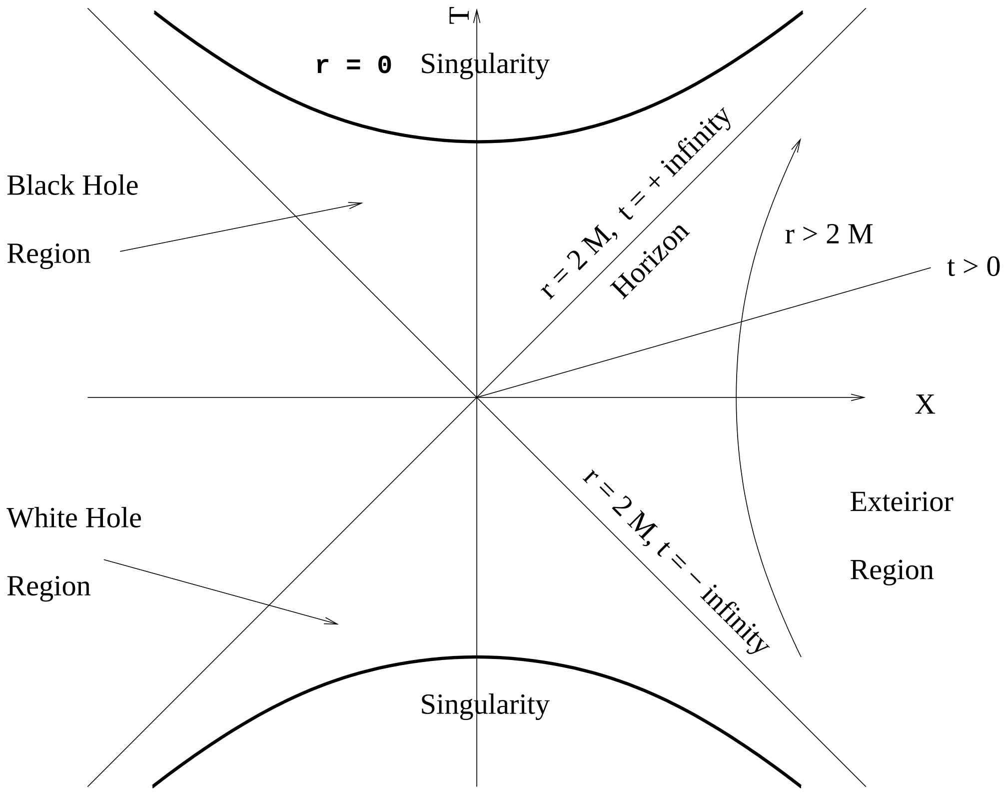


Figure 5.1: Kruskal Diagram for the Schwarzschild space-time

Looking at the figure representing the space-time ("extended") we can understand the $r = R_S$ singularity. The Schwarzschild time is ill defined at R_s since the stationary Killing vector becomes null. The full line segments at 45^0 are labeled by $r = R_S, t = \pm\infty$. The Schwarzschild coordinates provide a chart only for the right (and the left) wedge. To 'see' the top and the bottom wedges one has to use the Kruskal coordinates. Since the form of the $T - X$ metric is conformal to the Minkowski metric, the light cones are the familiar ones. one can see immediately that while we can have time-like and null trajectories *entering* the top wedge, we can't have any *leaving* it. Likewise we can have such 'causal' trajectories *leaving* the bottom wedge, we can't have any *entering* it. We have here examples of *one-way surfaces*. The top wedge is called the *black hole region* while the bottom wedge is called the *white hole region*. The line $r = R_S$ $(\times S^2)$, separating the top and the right wedges is called the *event horizon*. In fact existence of an event horizon is the distinguishing and (defining) property of a black hole. For the corresponding *Penrose Diagram*, see the appendix.

Incidentally, what would be the gravitational red shift for light emitted from the horizon? Well, the observed frequency at infinity would be zero but any way *no* light will be received at infinity! For light source very, very close to the horizon (but on the out side), the red shift factor will be extremely large. Consequently the horizon is also a surface of infinite red shift (strictly true for static black hole horizons). Imagine the converse now. Place an observer very near the horizon and shine light of some frequency at him/her from far away. The frequency he/she will see will be $\omega_\infty(1 - \frac{R_S}{r_{obs}})^{-1/2}$. If the light shining is the cosmic microwave background radiation with frequency of about 4×10^{11}, to see it as yellow color light of frequency of about 3×10^{15} the observer must be within a fraction of 10^{-8} from the horizon. For a Solar mass black hole this is about a hundredth of a millimeter from the horizon! At such locations the tidal forces will tear apart the observer before he/she can see any light.

The first, simplest solution of Einstein's theory shows a crazy space-time! How much of this should be taken seriously?

What we have above is an 'eternal black hole', which is nothing but the (mathematical) maximally extended spherically symmetric vacuum solution. From astrophysics of stars and study of the interior solutions it appears that if a star with mass in excess of about 3 solar masses undergoes a complete gravitational collapse, then a black hole will be formed (i.e. radius of the collapsing star will be less that the R_S. The space-time describing such a situation is not the eternal black hole but will have the analogues of the right and the top wedges. It will have event horizon and black hole regions. Are there other solutions that exhibit similar properties? The answer is yes but again mathematically peculiar. We will see these in the next lecture.

5.1.2 The Reissner-Nordstrom Black Hole

These space-times are solutions of Einstein-Maxwell field equations. Like the Schwarzschild solution, these are also spherically symmetric and static. Consequently, the ansatz for the metric remains the same as in (4.1). In addition, we need an ansatz for the electromagnetic field. It is straight forward to show that spherical symmetry and staticity implies that the only non-vanishing components of $F_{\mu\nu}$ are,

$$F_{tr} = \xi(r) \quad , \quad F_{\theta\phi} = \eta(r)\sin\theta \ . \tag{5.11}$$

The $dF = 0$ ('Bianchi identity') Maxwell equations then imply that $\eta(r) = Q_m$ is a constant while the remaining Maxwell equations imply that $\xi(r) = \frac{Q_e}{r^2}\sqrt{f(r)g(r)}$ where Q_e is a constant. The Q's correspond to electric and magnetic charges. There is no evidence for magnetic monopoles yet, so we could take $Q_m = 0$. However we will continue to assume it to be non-zero in this section.

The stress tensor for Maxwell field is defined as,

$$T_{\mu\nu} \ = \ \frac{1}{4\pi}\left[\frac{1}{4}g_{\mu\nu}\left(F_{\alpha\rho}F_{\beta\sigma}g^{\alpha\beta}g^{\rho\sigma}\right) - F_{\mu\alpha}F_{\nu\beta}g^{\alpha\beta}\right] \ . \tag{5.12}$$

Note: This can be derived starting from the usual Maxwell action ($\sim -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$) with the Minkowski metric replaced by a general metric. The stress tensor is then defined as the coefficient of $\delta g^{\mu\nu}$ in the variation of the action. The sign of the action is determined by the positivity of the Kinetic term (F_{0i}^2). The factor in front is determined by **Make precise**.

The overall sign can be deduced by checking that the energy density T_{tt} is positive while the factor can be deduced by matching with the special relativity result.

It follows that the non-zero components of $T_{\mu\nu}$ are given by,

$$
\begin{aligned}
T_{tt} &= \frac{1}{8\pi}\frac{Q^2}{r^4}f(r) \quad , \quad Q^2 \ := \ Q_e^2 + Q_m^2 \\
T_{rr} &= -\frac{1}{8\pi}\frac{Q^2}{r^4}g(r) \\
T_{\theta\theta} &= \frac{1}{8\pi}\frac{Q^2}{r^2} \quad , \quad T_{\phi\phi} \ = \ \sin^2\theta \ T_{\theta\theta}
\end{aligned} \tag{5.13}
$$

Due to the tracelessness of the stress tensor of electromagnetism, the Einstein equation to be solved becomes $R_{\mu\nu} = 8\pi T_{\mu\nu}$. Using the expressions given in (4.3, 5.13), it is straight forward to obtain the Reissner-Nordstrom solution:

$$
\begin{aligned}
f(r) &= \frac{\Delta(r)}{r^2} \quad , \quad g(r) \ = \ f^{-1}(r) \\
F_{tr} &= \frac{Q_e}{r} \quad , \quad F_{\theta\phi} \ = \ Q_g \ \sin\theta \\
\Delta(r) &:= r^2 - 2Mr + Q^2 \ , \quad M, \ Q \text{ are constants,}
\end{aligned} \tag{5.14}
$$

Evidently, for $Q = 0$ we recover the Schwarzschild solution with the identification $R_S = 2M$.

As before, the metric component g_{tt} vanishes when $\Delta = 0$ i.e. for $r = r_\pm := M \pm \sqrt{M^2 - Q^2}$. For $M^2 \geq Q^2$ we have thus *two* values of r at which $g_{tt} = 0$. For this range of values, we have a *Reissner-Nordstrom Black Hole*. For $M^2 = Q^2$, it is known as an *extremal* black hole while for $M^2 < Q^2$ ($r_\pm$ is complex), one has what is known as a *naked singularity*. As before, the Riemann curvature components blow up *only* as $r \to 0$ and since there is no one way surface cutting it off from the region of large r, it is called a naked singularity. We will concentrate on the black hole case.

A kruskal like extension is carried out in a similar manner. The tortoise coordinate r_* is now given by,

$$
r_*(r) \ = \ r + \frac{r_+^2}{r_+ - r_-}\ell n\left|\frac{r - r_+}{r_+}\right| - \frac{r_-^2}{r_+ - r_-}\ell n\left|\frac{r - r_-}{r_-}\right| \ . \tag{5.15}
$$

There are now *three* regions to be considered:

$$
\begin{aligned}
&\text{A} \ : \ 0 < r < r_- \ \leftrightarrow \ 0 < r_* < \infty \quad \text{(Stationary)} \\
&\text{B} \ : \ r_- < r < r_+ \ \leftrightarrow \ -\infty < r_* < \infty \quad \text{(Homogeneous)} \\
&\text{C} \ : \ r_+ < r < \infty \ \leftrightarrow \ -\infty < r_* < \infty \quad \text{(Stationary)}
\end{aligned}
$$

The Kruskal-like coordinates, U, V are to be defined in each of these regions such that the metric has the same form and then "join" then at the chart boundaries $r_\pm$. The corresponding *Penrose diagram* is show in the appendix.

5.2 The Stationary (non-static) Black Holes

5.2.1 The Kerr-Newman Black Holes

It turns out that for the Einstein-Maxwell system, the most general stationary black hole solution – the Kerr-Newman family – is characterized by just *three* parameters: mass, M, angular momentum, J and charge, Q. For $J = 0$ one has spherically symmetric (static) two parameter family of solutions known as the *Reissner-Nordstrom* solution. The $J \neq 0$ solution is axisymmetric and non-static. This result goes under the 'uniqueness theorems' and is also referred to as *black holes have no hair*. The significance of this result is that even if a black hole is produced by any complicated, non- symmetric collapse it settles to one of these solutions. All memory of the collapse is radiated away. This happens *only* for black holes!

The black hole Kerr-Newman space-time can be expressed by the following line element:

$$ds^2 = \frac{\eta^2\Delta}{\Sigma^2}dt^2 - \frac{\Sigma^2 sin^2\theta}{\eta^2}(d\phi - \omega dt)^2 - \frac{\eta^2}{\Delta}dr^2 - \eta^2 d\theta^2 \qquad \text{where,} \tag{5.16}$$

$$\begin{array}{lllcll} \Delta & := & r^2 + a^2 - 2Mr + Q^2 & ; & \Sigma^2 := & (r^2+a^2)^2 - a^2 sin^2\theta\Delta \\ \omega & := & \frac{a(2Mr - Q^2)}{\Sigma^2} & ; & \eta^2 := & r^2 + a^2 cos^2\theta \end{array}$$

$$\begin{array}{llll} a = 0 & , \quad Q = 0 & : & \text{Schwarzschild solution} \\ a = 0 & , \quad Q \neq 0 & : & \text{Reissner-Nordstrom solution} \\ a \neq 0 & , \quad Q = 0 & : & \text{Kerr solution} \end{array}$$

These solutions have a true curvature singularity when $\eta^2 = 0$ while the coordinate singularities occur when $\Delta = 0$. This has in general two real roots, $r_\pm = M \pm \sqrt{M^2 - a^2 - Q^2}$, provided $M^2 - a^2 - Q^2 \geq 0$. The outer root, r_+ locates the event horizon while the inner root, r_- locates what is called the *Cauchy horizon*. When these two roots coincide, the solution is called an *extremal* black hole.

When $\Delta = 0$ has *no real root*, one has a *naked singularity* instead of a black hole. A simple example would be negative mass Schwarzschild solution. The name *naked* signifies that the true curvature singularity at $\eta^2 = 0$ can be seen from far away. While mathematically such solutions exist, it is generally believed, but not conclusively proved, that in any realistic collapse a physical singularity will always be covered by a horizon. This belief is formulated as the "cosmic censorship conjecture". There are examples of collapse models with both the possibilities. The more interesting and explored possibility is the black hole possibility that we continue to explore.

We can compute some quantities associated with an event horizon. For instance, its area is obtained as:

$$A_{r_+} := \int_{r_+} \sqrt{det(g_{ind})} d\theta d\phi = \sqrt{\Sigma^2} \int sin\theta d\theta d\phi = 4\pi(r_+^2 + a^2) \tag{5.17}$$

For Schwarzschild or Reissner-Nordstrom static space-time we can identify $(g_{tt} - 1)/2$ with the Newtonian gravitational potential and compute the 'acceleration due to gravity' at the horizon by taking its radial gradient. Thus,

$$\text{When } a = 0, \text{ surface gravity, } \kappa \ := \ \frac{1}{2}\frac{dg_{tt}}{dr}|_{r=r_+} \ = \ \frac{r_+ - M}{r_+^2} \ = \ \frac{r_+ - M}{2Mr_+ - Q^2} \tag{5.18}$$

Although for rotating black holes 'surface gravity' can not be defined so simply, it turns out that when appropriately defined it is still given by the same formula (i.e. the last equality above).

There is one more quantity associated with the event horizon of a rotating black hole – the angular velocity of the horizon, Ω. It is defined in a little complicated manner. For the rotating black holes we have two Killing vectors: $\xi := \partial_t$ (the Killing vector of stationarity) and $\psi := \partial_\phi$ (the Killing vector of axisymmetry. Their *norms*^{2} are given by $g_{tt}, g_{\phi\phi}$ respectively. Both are *space-like* at the horizon. However there is another Killing vector, $\chi := \xi + \Omega\psi$, which is null and hence similar to the stationary Killing vector of the static cases. This Ω is defined to be the angular velocity of the horizon. It turns out to be equal to the function ω evaluated at $r = r_+$). From the definition given above it follows that,

$$\Omega := \frac{a}{r_+^2 + a^2}. \tag{5.19}$$

For charged black holes one also defines a surface electrostatic potential as,

$$\Phi := \frac{Qr_+}{r_+^2 + a^2} \tag{5.20}$$

Thus we have defined:

$$\begin{array}{lllllll} M & = & M & ; & r_+ & = & M + \sqrt{M - a^2 - Q^2} \\ A & = & 4\pi(r_+^2 \ + \ a^2) & ; & \kappa & = & \frac{r_+ \ - \ M}{2Mr_+ \ - \ Q^2} \\ J & = & Ma & ; & \Omega & = & \frac{a}{r_+^2 \ + \ a^2} \\ Q & = & Q & ; & \Phi & = & \frac{Qr+}{r_+^2 \ + \ a^2} \end{array} \tag{5.21}$$

Now one can verify explicitly that,

$$\delta M \ = \ \frac{\kappa}{8\pi}\delta A \ + \ \Omega\delta J \ + \ \Phi\delta Q \tag{5.22}$$

This completes our survey of examples of black hole solutions and some of their properties. All these are *stationary* solutions of Einstein-Maxwell field equations. In the next section we will consider more general black holes.

5.3 General Black Holes

One can very well imagine physical processes wherein a star collapses to form a black hole that settles in to a stationary black hole. However somewhat later another star or other body is captured by the black hole that eventually falls in to the black hole changing its parameters. This process can repeat. Such processes cannot be modeled by stationary space-times so one needs a general characterization of space-times that can be said to contain black hole(s).

One always imagines such space-times to be representing *compact* bodies i.e. sufficiently far away the space-time is essentially Minkowskian. Now the notion of a black hole is that there is a region within the space-time from which *nothing* can escape to *"infinity", ever*. 'Nothing' can be understood as causal curves (curves whose tangent vectors are either time-like or null) reaching out to farther distances. 'Infinity' and 'ever' needs to be defined more sharply in order to provide a precise enough definition of a black hole. The 'infinity' is specified to be the 'infinity' of an *asymptotically flat space-time*. This has a region identified as "future null infinity", $\mathcal{J}^+$. Consider now the set of all the points of space-time that can send signals to this $\mathcal{J}^+$. Call this the *past* of the future null infinity. If the space-time still has some points left out, then it is said to contain a black hole region (a four dimensional sub-manifold). The boundary of this region, (a three dimensional hypersurface) is called the event horizon. One can look at the intersection of the black hole region with a "$t = constant$" slice (technically a Cauchy surface) and identify each connected component as a black hole at the instant, t.

In a general space-time containing black holes various things can happen: new black holes may form, some may merge, some will grow bigger etc. However some things *cannot* happen.

For instance, once a black hole is formed, it can never disappear. A black hole may also never split in to more black holes (no bifurcation theorem). This result depends only on the definition of black holes and topology. It stipulates that while black holes can merge and/or grow, they can not split.

The 'evolution' of such black holes is tracked by a family of Cauchy surfaces. One can thus obtain the areas of the intersection of the horizon and the Cauchy slices. Extremely interestingly, the *area of a black hole may never decrease* (the Hawking's area theorem). This result prompted Bekenstein to think of black hole area as its entropy.

Note that the *no-bifurcation theorem* put some conditions on possible evolution of black holes. The area of a black hole may change due to accretion from other objects or merging of black holes. The Hawking theorem stipulates that in either of these processes, the area must *not* decrease. This is a stronger statement.

Indeed one can imagine processes involving black holes wherein a black hole does change its properties (eg. area) consistent with the above theorems. However the accretion/merger processes may be separated by long periods of 'inactivity'. During these periods, the black hole may be well approximated by *stationary black hole solutions*. For these a lot is known. Some of these results are summarized in the appendix [5, 10].

5.4 Black Hole Thermodynamics

In light of these results, the variational equation (5.7) we had above looks very much like the first law of thermodynamics. Indeed one can define general *stationary* black holes and obtain expressions for the area (A), surface gravity (κ), mass (M), angular momentum (J), angular velocity (Ω), charge (Q), surface potential (Φ) etc. Three of these, M, J, Q are defined with reference to infinity while the remaining four are defined at the horizon and are constant over the horizon. One can then consider variations of these quantities and prove that the first law expression seen explicitly actually holds much more generally. One thus has *laws of black hole mechanics* which are completely analogous to the usual *laws of thermodynamics.* Here is a table of analogies:

Laws of	Black Hole Mechanics	Thermodynamics
Zeroth law	κ is constant	T is constant
First law	$\delta M = \frac{\kappa}{8\pi}\delta A + \Omega\delta J + \Phi\delta Q$	$\delta U = T\delta S + P\delta V + \cdots$
Second law	$\delta A \geq 0$	$\delta S \geq 0$
Third law	Impossible to achieve $\kappa = 0$	Impossible to achieve $T = 0$

The analogy is very tempting, in particular, $\kappa \sim T, A \sim S$ is very striking. Like a thermodynamical system, black hole space-times are characterized by a few parameters. Just as for thermodynamical systems at equilibrium, all memory of the history of attaining the equilibrium is lost, so it is for the stationary black holes thanks to the uniqueness theorems. A typical thermodynamical system has a total energy content, U and a volume, V which are fixed *externally.* In *equilibrium* the system exhibits further *response* parameters such as temperature, T and pressure, P which are uniform through out the system. In going from one equilibrium state to another one the system ensures that its *entropy,* S has not decreased and of course the energy conservation is not violated. It is also important to note that the thermodynamic quantities $T, P, ...$ are functions *only* of 'conjugate' quantities $S, V, ...$. Black holes also have parameters, referring to the global space-time, such as M, J, Q and also 'response' parameters, referring to the horizon, such as κ, A, Ω, Φ and these must also be functions only of the previous set of parameters. This of course is true for the explicit stationary black hole solutions. A natural and some what confusing question is: what is the thermodynamic system here - the entire black hole space-time or only the horizon? If it is the former then equilibrium situation should correspond to stationary space-times. If it is the latter it is enough that the geometry of the horizon alone is suitably 'stationary'. The latter is physically more appealing while historically black hole thermodynamics was established using the global definitions of black holes. Only over the past few years the more local view is being developed using generalization of stationary black holes called "isolated horizons". For these also the mechanics-thermodynamics analogy is established [11].

However if taken literally one immediately has a problem. If a black hole has a non-zero temperature, it must radiate. But by definition nothing can come out of a black hole (since the surface gravity is defined for the horizon, we expect horizon to radiate). So how can we reconcile these? Here Hawking became famous once more. He observed that so far quantum theory has been ignored. There are always quantum fluctuations. It is conceivable then that positive and negative energy particles that pop out of the vacuum (and usually disappear again) can get separated by the horizon and thus cannot recombine. The left over particle can be thought of as constituting black hole radiation. He in fact demonstrated that a black hole indeed radiates with the radiation having a black body distribution at a temperature given by $k_B T = \frac{\hbar\kappa}{2\pi}$! This provides the proportionality factor between surface gravity and temperature. Consequently, the entropy is identified as $S = \frac{k_B}{\hbar}\frac{A}{4}$. How much is this temperature? Restoring all dimensional constants the expression is [5]:

$$\begin{aligned} T &= \frac{\hbar c^3}{8\pi G k_B M_\odot}\left(\frac{M_\odot}{M}\right) \quad {}^0K \\ &= 6 \times 10^{-8}\left(\frac{M_\odot}{M}\right) \end{aligned} \tag{5.23}$$

Notice that heavier black hole is cooler, so as it radiates it gets hotter and radiates stronger in a run-away process. A rough estimate of total evaporation time is about $10^{71}(M/M_\odot)^3$. The end point of evaporation is however controversial because the semi-classical method used in computations cannot be trusted in that regime.

Another fertile area for research has been the *microscopic* i.e. statistical mechanical understanding of black hole entropy. For normal systems, the entropy being an extensive quantity goes as *volume* while for a black hole it goes as area of the horizon (this may be thought of as another argument for thinking of horizon as the thermodynamic system). A simple way to see that entropy *can be* proportional to the area is to use the Wheeler's 'it from bit' picture. Divide up the area in small area elements of size about the Planck area ($\ell_p^2 \sim 10^{-66} cm^2$). The number of such cells is $n \sim A/(\ell_p^2)$. Assume there is spin-like variable in each cell that can exist in two states. The total number of possible such states on the horizon is then 2^n. So its logarithm, which is just the entropy, is clearly proportional to the area. Of course same calculation can be done for volume as well to get entropy proportional to volume. What the picture shows is that one can associate *finitely many* states to an elementary area.

There are very many ways in which one obtains the Bekenstein entropy formula. Needless to say, it requires making theories about quantum states of a black hole (horizon). Consequently everybody attempting any theory of quantum gravity wants to verify the formula. Indeed in the non-perturbative quantum geometry approach the Bekenstein formula has been derived using the 'isolated horizon' framework, for the so-called non-rotating horizons. String theorists have also reproduced the formula although only for black holes near extremality.

Recall that extremal solutions are those which have $r_+ = r_-$ which implies that the surface gravity vanishes. For more general black holes this is taken to be the definition of extremality. For un-charged, rotating extremal black holes $M = |a|$ while for charged, non-rotating ones $M = |Q|$. Since vanishing surface gravity corresponds to vanishing temperature one looks for the third law analogy. It has been shown that the version of third law, which asserts that it is impossible to reach zero temperature in finitely many steps, is verified for the black holes - it is impossible to push a black hole to extremality (say by throwing suitably charged particles) in finitely many steps. There is however another version of the third law that asserts that the entropy vanishes as temperature vanishes. This version is *not* valid for black holes since extremal black holes have zero temperature but finite area.

What began as a peculiar solution of Einstein equations has evolved in to fertile research area particularly offering testing ground for glimpses at the quantum version of GR. Black holes is an arena where GR, statistical mechanics and quantum theory are all called in for an understanding.

Chapter 6

Cosmology

6.1 Robertson-Walker Space-times

Let us now leave the context of compact, isolated bodies and the space-times in their vicinity and turn our attention to the space-time appropriate to the whole universe. We can make no progress by piecing together space-times of individual compact objects such as stars, galaxies etc, since we will have to know all of them! Instead we want to look at the universe at the largest scale. Since our observations are necessarily finite (that there are other galaxies was discovered only about 80 years ago!), we have to make certain assumptions and explore their implications. These assumptions go under the lofty names of 'cosmological principles'.

One fact that we do know with reasonable assurance is that the universe is 'isotropic on a large scale'. What this means is the following. If we observe our solar system from any planet, then we do notice its structure, namely other planets. If we observe the same from the nearest star (alpha centauri, about 4 light years), we will just notice the Sun. Likewise is we observe distant galaxies, they appear as structure less point sources (which is why it took so long to discover them). If we look still farther away then even clusters of galaxies appear as points. We can plot such sources at distances in excess of about a couple of hundred mega-parsecs on the celestial sphere. What one observes is that the sources are to a great extent distributed uniformly in all directions. We summarize this by saying that the universe on the large scale is isotropic about us. We appear to occupy a special vantage point! One may accept this as a fact and ponder about why we occupy such a special position. However since Copernicus we have learnt that it is theoretically more profitable to systematically deny such privileged positions. The alternative is then to propose that universe must look isotropic from all locations (clusters of galaxies). Since universe appears isotropic to us at present, we assume that the same must be true for other observers else where i.e. there is a common 'present' at which isotropic picture hold for all observers. Denial of privileged position also amounts to assuming that the universe is *spatially* homogeneous i.e. at each instant there is a spatial hypersurface (space at time t) on which all points are equivalent. Isotropy about each point means that there must be observers (time-like vector field) who will not be able to detect any distinguished direction. It follows then that these observers must be orthogonal to the spatial slices. The statement that on large scale the universe is spatially homogeneous and isotropic is called the 'cosmological principle'. There is a stronger version, the so-called 'perfect cosmological principle' that asserts that not only we do not have special position, we are also not in any special epoch. Universe is homogeneous in time as well. It is eternal and unchanging. This principle leads to the 'steady state cosmologies'. The so-called standard cosmology is based on spatial homogeneity and isotropy and this is what is discussed below. Weinberg presents discussions on alternative cosmologies.

A spatially homogeneous space-time can be viewed as a stack of three dimensional spatial slices. Spatially homogeneity (and indeed isotropy about each point of a slice) also implies that these spatial slices must have a "constant curvature" i.e. the curvature tensors must have a specified form involving a constant, in particular the Ricci scalar is a constant. Such three dimensional Riemannian spaces are completely classified and come in three varieties depending on the sign of the curvature. Labeling each of the slices by a time coordinate, τ, and denoting the normalized constant curvature by k, one can write the form of the metric for the universe as:

$$ds^2 = d\tau^2 - a^2(\tau)\left\{\begin{array}{ll} d\psi^2 + sin^2\psi d\Omega^2 & \text{Spherical, } k=1 \\ d\psi^2 + \psi^2 d\Omega^2 & \text{Euclidean, } k=0 \\ d\psi^2 + sinh^2\psi d\Omega^2 & \text{Hyperbolic, } k=-1 \end{array}\right\} \text{where,} \tag{6.1}$$

$$d\Omega^2 := d\theta^2 + sin^2\theta d\phi^2$$

$$ds^2 = d\tau^2 - a^2(\tau)\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right) \qquad \text{(Alternative form)} \tag{6.2}$$

$$ds^2 := d\tau^2 - a^2(\tau) ds_3^2 \tag{6.3}$$

The $a^2(\tau)$ determines the value of the constant spatial curvature and is accordingly called the scale factor. It is allowed to depend on τ. The space-times with the above form for the metric are called Robertson-Walker geometries. Most of modern cosmography – mapping of the cosmos – is based on these geometries.

Thus if $R(\tau)$ is a distance between two galaxies at the same τ, then its change with τ can be obtained as:

$$v := \frac{dR(\tau)}{d\tau} = \frac{R(\tau_0)}{a(\tau_0)}\frac{da(\tau)}{d\tau} = \frac{R(\tau)}{a(\tau)}\dot{a} \equiv H(\tau)R(\tau) \tag{6.4}$$

Hence, for example, speed of recession of galaxies is proportional to their separation. This is the famous conclusion drawn by Hubble. He actually observed the relation between the red-shift factor and separation. Let us obtain the red shift factor by methods discussed before.

Let k^μ denote the null geodesic of the light emitted from a source P_1 in the slice at τ_1 being received at P_2 in the slice at τ_2. Assume for the moment that one can always find a Killing vector, ξ^μ, such that it coincides with the component of k^μ along Σ_i at points P_i. Such a Killing vector is necessarily spatial and orthogonal to u^μ. Since k is null, it follows that $\omega := k \cdot u = \pm k \cdot \xi/(||\xi||)$. Now applying our previous result that $k \cdot \xi$ is constant along the null geodesic, it follows that $\omega_2/\omega_1 = ||\xi||_1/||\xi||_2 = a(\tau_1)/a(\tau_2)$. Therefore $z := \omega_1/\omega_2 - 1 = (a(\tau_2) - a(\tau_1))/a(\tau_1)$. For nearby galaxies we may approximate $\tau_2 - \tau_1 \approx R$ ($c = 1$ units) and $a(\tau_2) \approx a(\tau_1) + \dot{a}(\tau_2 - \tau_1)$ to get $z \approx H(\tau)R(\tau)$. This was the relation observed by Hubble and is known as the Hubble law. It was the red shift ($H > 0$) that was observed so one inferred that the universe is actually expanding.

Universe is of course not empty. The stress tensor must also be consistent with the assumptions of homogeneity and isotropy. This turns out to be of the form of perfect fluid:

$$T_{\mu\nu} = \rho(\tau)u_\mu u_\nu + P(\tau)(u_\mu u_\nu - g_{\mu\nu}), \tag{6.5}$$

where P is the pressure, ρ is the energy density and u_μ is the normalized velocity of the observers, orthogonal to the spatial slices. Our system of equations now have 3 unknown functions, a, ρ, P of a single variable τ for each choice of the spatial curvature, k. Turning the crank, the Einstein equations reduce to:

$$3\frac{\ddot{a}}{a} = -4\pi(\rho + 3P) \qquad \text{(Raychoudhuri equation)} \qquad (6.6)$$

$$3\frac{\dot{a}^2}{a^2} = 8\pi\rho - \frac{3}{a^2}k \quad k = \pm 1, 0 \; ; \qquad \text{(Friedmann equation)} \qquad (6.7)$$

$$\dot{\rho} = -3(\rho + P)\frac{\dot{a}}{a} \qquad \text{(Conservation equation)} \qquad (6.8)$$

The first striking inference is that if ρ, P are both positive, as they are for normal matter, then we can *not* have a static universe, $a =$ constant, for any choice of k[1]. Further, $\ddot{a} < 0$ implies $\dot{a}$ must be monotonically decreasing implies that it can not change sign. Hence the universe is always expanding or always contracting except possibly when there is a change over from expanding to contracting phase. Note that the scale factor affects *all length measurements in a given slice* in the same manner.

This observed fact of expanding universe immediately implies that the universe must have been extremely small a finite time ago. If H is assumed to be constant then the age of the universe must be about H^{-1}! Calling $\tau = 0$ when $a = 0$ held, one says that the universe began in a "big bang", from a highly singular geometry. All these are consequences of the Robertson-Walker geometry and qualitative properties of the pressure and density. This is a very striking prediction of GR, which is consistent with observation. Let us return to the equations again.

Our equations are still under-determined. One can verify that the first order equations (6.7) and (6.8) imply the second order equation (6.6 Thus we have two equations for three unknown functions. We need a relation between the density and the pressure. Such a relation is usually postulated in the form $P = P(\rho)$ and is called an equation of state for the matter represented by the stress tensor. At a phenomenological level it characterizes internal dynamical properties of matter. There are two popular and well-motivated choices, namely, $P = 0$ (dust) and $P = \frac{1}{3}\rho$ (radiation). Once this additional input is specified, one can solve the conservation equation to obtain a as a function of ρ (or vice a versa). Plugging this in the 2nd equation gives a differential equation for $\rho(\tau)$. This way one can determine both the scale factor and the matter evolutions. Here is a table of solutions from Wald's book. These are referred to as Friedmann-Robertson-Walker (FRW) cosmologies.

[1]Strictly, this conclusion is not valid if a non-zero cosmological constant is present. A cosmological constant is incorporated in the above equations by the replacements: $\rho \to \rho + \Lambda, P \to P - \Lambda$. Now the right hand side of the Raychoudhuri equation needed not be negative and the singular state in the past can be avoided.

6.2 Friedmann-Robertson-Walker Cosmologies

	Dust, $P = 0$	Radiation, $P = \frac{1}{3}\rho$
$k = 1$	$a = (C/2)(1 - cos\eta)$ $\tau = (C/2)(\eta - sin\eta)$	$a = \sqrt{C'}\left\{1 - \left(1 - \frac{\tau}{\sqrt{C'}}\right)^2\right\}^{1/2}$
$k = 0$	$a = (\frac{9C}{4})^{1/3}\tau^{2/3}$	$a = (4C')^{1/4}\sqrt{\tau}$
$k = -1$	$a = (C/2)(cosh\eta - 1)$ $\tau = (C/2)(sinh\eta - \eta)$	$a = \sqrt{C'}\left\{\left(1 + \frac{\tau}{\sqrt{C'}}\right)^2 - 1\right\}^{1/2}$
	$\rho a^3 = \text{constant}$	$\rho a^4 = \text{constant}$

One can get this far with just cosmological principle, GR and some assumptions about the matter. There are two separate issues to be addressed now. Firstly, we need to make some physically well motivated assumptions regarding the composition of the universe i.e. components of density and pressure together with their equations of state, to obtain a sufficiently general solution of the equations. Secondly, we need to identify suitable parameters which can be determined from observations. Let us consider the first aspect.

We can divide up the matter contents in to three classes: (i) non-relativistic matter (dust) which is characterized by constituents such as galaxies moving with non-relativistic speeds there by exerting negligible pressure ($P_{NR} = 0$) and energy density ρ_{NR}, (ii) relativistic matter such as radiation (photons, neutrinos and other highly relativistic particles) with equation of state $P_R = \frac{1}{3}\rho_R$ and (iii) a possible cosmological constant term with equation of state $P_\Lambda = -\rho_\Lambda = \Lambda$. The total energy density and pressure is the sum of these three types. On physical grounds (since earlier universe was smaller it must have been hotter), we expect the non-relativist matter to be dominant during the epoch through which the galaxies etc have existed and the relativistic matter to be dominant in the earlier phase of evolution. The Λ component would have to be estimated in comparison with the others. In reality of course all three components are present all through but during specific era we can concentrate only one component and neglect the others. Our model now consists of $P = \frac{1}{3}\rho_R - \Lambda, \ \rho = \rho_{NR} + \rho_R + \Lambda$.

As noted above, the conservation equation immediately gives $\rho_R \propto a^{-4}, \ \rho_{NR} \propto a^{-3}$ and of course $\rho_\Lambda = \text{constant}$. It is now assumed that these behaviours of the energy densities continues to hold generally to a very good approximation. Now only the Friedmann equation remains to be solved.

Since there are unknown constants of proportionality in the behaviour of the densities, we still cannot obtain a general solution for the scale factor evolution. Also, more than explicit form of $a(\tau)$, we are interested in obtaining evolutions of *observable* quantities.

There are three convenient quantities chosen for this purpose: (1) the Hubble parameter $H(\tau) := \frac{\dot{a}}{a}$, (2) the deceleration parameter $q(\tau) := -\frac{a\ddot{a}}{\dot{a}^2}$ and (3) the critical density $\rho_c(\tau) :=$

$\frac{3H^2}{8\pi G}$. The densities are traded for in terms of $\Omega_i := \frac{\rho_i}{\rho_c}$, $i = R, NR, \Lambda$. The same quantities evaluated at present epoch are suffixed by 0.

The Friedmann and the Raychoudhuri equations can be rewritten as

$$\Omega = 1 + \frac{\kappa}{a^2 H^2} \tag{6.9}$$

$$P = \frac{H^2}{8\pi G}(2q - \Omega) = -\frac{1}{8\pi G}\left(\frac{\kappa}{a^2} + (1-2q)H^2\right) \tag{6.10}$$

We can eliminate the constants of proportionality in the densities by taking ratios with their present epoch values, eg $\rho_{NR} = \rho_{NR,0}\frac{a_0^3}{a^3}$ etc. Dividing the Friedmann equation by a_0^2 leads to

$$\left(\frac{\dot{a}}{a_0}\right)^2 = \frac{8\pi G}{3}\frac{a^2}{a_0^2}\left(\rho_{NR,0}\left(\frac{a_0}{a}\right)^3 + \rho_{R,0}\left(\frac{a_0}{a}\right)^4 + \Lambda\right) - \frac{\kappa}{a_0^2} \tag{6.11}$$

$$= H_0^2\Bigg[\ \Omega_{NR,0}\left(\frac{a_0}{a}\right) + \Omega_{R,0}\left(\frac{a_0}{a}\right)^2 + \Omega_{\Lambda,0}\left(\frac{a}{a_0}\right)^2 - \Omega_{NR,0} - \Omega_{R,0} - \Omega_{\Lambda,0} + 1\ \Bigg] \tag{6.12}$$

Here we have used eqn (6.9) at present epoch, in going to the second equation above. Putting $a = a_0 x$ gives,

$$\dot{x}^2 = H_0^2\left[\Omega_{NR,0}x^{-1} + \Omega_{R,0}x^{-2} + \Omega_{\Lambda,0}x^2 + 1 - \Omega_0\right] \tag{6.13}$$

The right hand side of this equation involves the present values of the density parameters. If these are determined observationally, the the equation can be integrated to obtain evolution of the scale factor normalized to $a_0 = 1$ (say).

Noting that $\dot{x} = Hx$, we can eliminate the $\dot{x}^2$ to *directly* obtain $H(x)$ as,

$$H^2(x) = H_0^2\left[\Omega_{NR,0}x^{-3} + \Omega_{R,0}x^{-4} + \Omega_{\Lambda,0} + (1 - \Omega_0)x^{-2}\right] \tag{6.14}$$

Using equation (6.9) and its present epoch version (for $\kappa \neq 0$), we can eliminate the Hubble parameter in favour the total energy density to get,

$$\Omega(x) - 1 = \frac{\Omega_0 - 1}{\left[\Omega_{NR,0}x^{-1} + \Omega_{R,0}x^{-2} + \Omega_{\Lambda,0}x^2 - (\Omega_0 - 1)\right]} \tag{6.15}$$

These expressions can also be written in terms of the red shift by noting that $x = \frac{a}{a_0} = (1+z)^{-1}$. Notice that $x \to 0$ as $z \to \infty$ and $x = 1$ for $z = 0$.

If the total energy density is exactly equal to 1 at any epoch (which implies spatially flat universe), then it must remain so for all epochs. In the early epochs corresponding to $x \to 0$, we see that $\Omega(x) \to 1$, independent of the contribution of cosmological constant. The Hubble parameter is also independent of Λ in the early universe.

We have succeeded in determining the general evolutions of the scale factor, Hubble parameter and the density parameter *in terms of* observationally determinable present epoch values. The crucial assumption has been the dependence of various density components on the scale factor. This may be construed as characterizing the FRW cosmologies.

One can obtain relations among the present epoch density parameters, Hubble parameter and the deceleration parameter by simply writing the equations (6.9, 6.10) at the present epoch. Since the present day universe is matter dominated, we may neglect the contributions of radiation and write $\Omega_0 \approx \Omega_{NR,0} + \Omega_{\Lambda,0}$ and $P_0 \approx -\Omega_{\Lambda,0}\rho_{c,0}$. The equations then imply,

$$\Omega_{NR,0} + \Omega_{\Lambda,0} \ = \ 1 + \frac{\kappa}{a_0^2 H_0^2} \quad , \quad \Omega_{NR,0} - 2\Omega_{\Lambda,0} \ = \ 2q_0 \ , \tag{6.16}$$

Thus, if we could determine q_0 and $a_0 H_0$ by some means, we could also determine the two density parameters. The radiation density parameter is to be determined separately.

The two parameters, H_0, q_0 are determined from a *distance-red shift plot* for various sources[2]. The distance typically involves the *comoving path length*, ℓ, which is defined as,

$$\ell \ := \ \int_{\tau_e}^{\tau_o} \frac{d\tau}{a(\tau)} \ = \ \int \frac{da}{\dot{a}a} \ = \ \frac{1}{a_0}\int \frac{dx}{x\dot{x}} \ , \tag{6.17}$$

and we have the equation (6.13) for $\dot{x}$, we can directly obtain the luminosity distance as a function of the red shift and of course the density parameters. Thus, in an FRW model, we do have a distance-red shift relation expressed in terms of the density parameters. Making observational determination of such a relation therefore determines the density parameters. Further using $x = (1+z)^{-1}$ one obtains,

$$D_L(z) \ = \ \frac{(1+z)}{H_0}\int_0^z dz \left[\Omega_{NR,0}(1+z)^3 + \Omega_{R,0}(1+z)^4 + \Omega_{\Lambda,0} + (1-\Omega_0)(1+z)^2\right]^{-\frac{1}{2}} \tag{6.18}$$

Until recently, the observations of sources was limited to small red shifts (< 1) and one can evaluate the integral approximately to read off the parameters from a plot. With supernovae observations (SNa), one has measured red-shifts of about 5.

[2]There are various measures of distances in cosmology - the *luminosity distance* inferred from the apparent brighrtness of sources, the *angular diameter distance* inferred from the apparent sizes of sources etc. They all involve ℓ multiplied by suitable scale factor and various powers of the red-shift factor $(1+z)$ [4].

Chapter 7

The Big Bang Model

So far we have sketched qualitative evolution of the gross features of the universe according to the FRW models, notably the finiteness of the age of the universe and attendant divergence of densities at the beginning of the universe. We also note another observed fact: the matter in the universe is built up from more elementary constituents. The stars are made up of atoms which are made up of protons, neutrons and electrons. The nucleons have a further substructure in terms of quarks, gluons etc. This substructures get revealed as the structures collide and are broken. The standard model of particle physics summarizes the most elementary constituents (as of today) and their elementary interactions.

Now it stands to reason that as we go back in time, the universe contracts and heats up. The current constituents then increase their average motion, begin to collide and break apart. Further back in time it is conceivable that we will end up with a hot soup of elementary particles of the standard model. Conversely, we can imagine the universe to have begun as a hot soup of elementary particles which kept combining in to larger structures thanks to the universal expansion. The current constituents (at least some of these) notably various nuclei and atoms can thus be thought of as being cooked up during the evolution of the universe. The idea is also attractive from another point of view. If we assume the components of the soup to be in thermal equilibrium, then we can understand how the matter distribution came to be largely homogeneous and isotropic. Can we build a detail picture of this cooking process? Amazingly, the answer turns out to be YES and we obtain a *Thermal History* of the universe.

7.1 Thermal History of the Universe

We begin by assuming that at some early epoch, the universe consisted of (anti) nucleons, (anti) leptons and photons at some high temperature T. We know that current universe has atoms of various elements and photons. We are assuming that these (at least some of lighter elements) are cooked during the expansion of the universe. The question we want to understand is: What determines the products and their abundances during the cooking process? For this we must note a few points.

The abundances of various products will be correlated and possibly fixed if the products were in thermal equilibrium at some epoch. To realize and maintain an equilibrium, there must be processes (interactions) among these different species of matter. These have some *reaction rates* typically proportional to the average speed, the total cross-section and the number densities (just from the definition of a cross-section). These quantities are also functions of the equilibrium temperature and the chemical potentials of the species. However, the universe expands at a certain rate making the temperature fall at some rate making the reaction rates also to fall at some rate. As long as the reaction rate is higher that the expansion rate, thermal contact will be present and equilibrium will be maintained. If however reaction rate falls below the expansion rate, the reaction effectively ceases breaking the thermal contact between species. It is conceivable (and true) that the reaction rate falls *faster* than the expansion rate (which also falls generically as noted earlier), then at some stage the rates will equal and thermal contact between two species will be broken. Their number densities (or abundances) will thus be frozen at the values at this cross over time (temperature). Starting with certain number of species with mutual interactions of different types, it is possible that different species will freeze out at different epochs generating different abundances. We thus see a mechanism of generating different products as well as their relative abundances. The task is to determine the details!

Before describing the steps in the computation, let us note another qualitative feature expected. At sufficiently high temperature, we can expect a hot gas of charged (+ neutral) particles. The charged particles will be interacting violently producing photons. Let us momentarily call all particles other that photons as 'matter' and imagine an epoch wherein there was an equilibrium between matter and photons. As the universe cooled, at some stage the charged particles would combine to form neutral atoms. From this stage onwards the photons would be mostly decoupled and stream freely and we can expect to see a left over distribution of photons. What sort of spectrum (energy as a function of frequency) can be expected for such radiation today? It is *not* automatic that the spectrum will be black body spectrum, it depends on duration over which decoupling takes place and the matter temperature variation with time. *If* the decoupling is fast (opacity drops sharply), *then* one obtains the black body spectrum with a temperature $T_{\gamma,0} := T_{matter}(t_*)a(t_*)/a_0$. Here t_* is the time at which decoupling takes place sharply. Note that the time of decoupling is determined by details of the thermal history of matter prior to decoupling and hence the photon temperature today is dependent on this thermal history.

If however we *assume* that thermal history of matter was such that its temperature during the thermal contact period relaxed as $T_{matter}(t) = A/a(t)$ where A is some constant, *then* the photon distribution through out thermal contact and decoupling will be the black body distribution at temperature $T_\gamma(t) = A/a(t)$. As noted earlier, black body distribution form is preserved during expansion as a consequence of RW geometry independent of dynamics, we can determine the constant A as $T_{\gamma,0}a_0$. Notice that this fixes the assumed relaxation of matter temperature as well. A determination of the present temperature of the photons would thus give information not only about the decoupling era but even before that.

But can such an assumption be true? For this we have to look at matter-photon equilibrium a little more closely. We know density and pressures of matter and photon at temperature T as,

$$\rho(T) = bT^4 + mn + \frac{nk_BT}{\gamma - 1} \ , \ P(T) = \frac{1}{3}bT^4 + nk_BT \ , \ b := \frac{8\pi^5 k_B^4}{15h^3c^3} \sim 7.6 \times 10^{-15}(\text{cgs}) \ , \tag{7.1}$$

where, m is the mass of matter particles (assumed to be a single species), n is their number density and γ is the specific heat ratio. The particle number conservation implies $n(\tau)a^3(\tau) =$

$n_0 a_0^3$ while conservation equation leads to,

$$\frac{a}{T}\frac{dT}{da} = -\left[\frac{\sigma + 1}{\sigma + \frac{1}{3(\gamma-1)}}\right] \quad , \quad \sigma := \frac{4bT^3}{3nk_B} \tag{7.2}$$

If $\sigma \ll 1$ then $T \propto a^{-3(\gamma-1)}$ while if $\sigma \gg 1$ then $T \propto a^{-1}$. Furthermore, for very large σ the scale factor dependence cancels between that of temperature and the number density making σ a constant preserving its large value. This case is said to define a *hot big bang*. In a hot big bang soup, one can also relate the large constant value of σ to the ratio of present photon density and matter density. So in a hot big bang our assumption would hold. But is our universe a hot big bang universe?

For this we appeal to Gamow's theory of assuming that various nuclei *are* cooked from the soup. If so, there should have been a production of deuterium. This can take place if the temperature is of the order of 10^9 (dissociation temperature of deuterium, from nuclear physics) and the density of nucleons of the order of 10^{18} per cm^3 so that about 50 per cent of nucleons can fuse to form deuterium. This immediately gives $\sigma \sim 10^{11} \gg 1$! So we see that we do live in a hot big bang universe. Since the present density of baryons, estimated from visible matter, is of the order of 10^{-6} per cm^3, we also obtain the ratio of the two scale factors to be about 10^8 and so the present temperature of the photons should be about the same factor dividing 10^9, i.e. about 10. More detailed numbers give $T_{\gamma,0} \sim 5^0 K$. This was the prediction of Gamow in the late forties! We will discuss the story of this microwave background radiation later.

Let us return to an illustration of how thermal history is constructed. The logical steps in the calculations are the following.

1. For a thermodynamic system consisting of several interacting species of constituents in equilibrium, the number densities n_i are determined by the temperature T and the chemical potentials μ_i (from grand canonical ensemble of statistical mechanics). The chemical potentials are constrained by conservation laws obeyed by the interactions. If the chemical potentials are assumed to be zero to begin with, the number densities, pressures etc depend only on the temperature and of course intrinsic properties such as masses, couplings etc. This is commonly assumed.

2. The conservation equation (4.37) can be written as,

$$\frac{d\{a^3(P+\rho)\}}{d\tau} = a^3\frac{dP}{d\tau} = a^3\frac{dT}{d\tau}\frac{dP}{dT} \tag{7.3}$$

3. The equilibrium condition implies existence of the entropy function and the first law of thermodynamics gives integrability conditions among densities and pressures, namely,

$$\frac{dP}{dT} = \frac{P+\rho}{T} \tag{7.4}$$

   These two equations together give the conserved quantity,

$$\frac{d}{d\tau}\left[\frac{a^3}{T}(P(T)+\rho(T))\right] = 0 \tag{7.5}$$

   which is just the entropy in a volume $a^3(\tau)$.

4. Vanishing chemical potentials also imply that at any given temperature, only those species whose mass-energy is less that $K_B T$ will participate in the equilibrium system. This allows us to combine particle physics knowledge to determine the species present significantly at any temperature. Particle physics also tells us possible and dominant interactions and their cross-sections thereby determining the reaction rates. The Friedmann equation on the other hand gives the expansion rate $H(\tau)$.

5. Equating these two rates determines the temperature at which the reaction under consideration terminate. If this the only reaction responsible for coupling between two species, then termination of the reaction implies decoupling of the species. Their number densities are then fixed by this decoupling temperature and the densities subsequently fall as a^{-3}. This determines the relative abundances.

6. The conserved quantity together with the decoupling temperature allows us to determine the ratios of the scale factor which in turn determines the ratios of the times. This gives us the time scales of various epochs.

The steps 4 and 5 above are the non-trivial ones. For example consider the epoch wherein the temperature has dropped so much that we have only protons (left over from earlier epochs) and electrons together with photons in equilibrium. With further decrease of temperature, it becomes possible for protons and electrons to form neutral atoms. These however can again be broken apart by the photons. In general then we expect to have some atoms as well. Now we have two reactions to consider: $p^+ + e^- \to H + \gamma$ and $H + \gamma \to p^+ + e^-$. The dissociation energy of the hydrogen atom is 13.6 eV corresponding to an equivalent temperature of about $10^5 K$. The photons must have this much energy to cause dissociation. The rate of dissociation reaction however also depends on the densities of photons (with energy higher than 13.6 eV) and the hydrogen atoms. For the forward reaction, the rate depends on the densities of protons and electrons. At equilibrium, the rates of both the reactions are exactly matched with certain equilibrium densities of all the four species. The relevant photon density falls faster than the other densities thereby lowering the dissociation rate eventually switching off this reaction. This temperature is about $4000^0 K$, roughly a tenth of the temperature equivalent of the dissociation energy. Thus a switch-off temperature is *less* that the temperature indicated by energy considerations. These calculations are non-trivial.

Here is a summary of the thermal history of the universe. The first group of epochs, some times called the *Hadron Era* is quite uncertain and is where theorists can have a field day. The second group, the *Lepton Era* is some what better since standard model of particle physics is better understood here. The third group of *Plasma Era* is perhaps best understood and is where maximum support for Big Bang Model comes from. The last group of *Post-Recombination Era* is more in the domain of observational cosmology and where theories of structure formations (galaxies and their distributions) are to be confronted with observations. The table uses $1^0 K \sim 10^{-4} eV$ and is taken from Padmanabhan's book.

In the next lecture we will take a closer look at the microwave background radiation.

Time	Energy Scale	Temperature	Main Events
10^{-42} seconds	10^{18} GeV	10^{31} K	Quantum Gravity, Begin Inflation?
$10^{-32\pm6}$ seconds	$10^{13\pm3}$ GeV	$10^{26\pm3}$ K	End Inflation, Begin Re-heating ?
$10^{-18\pm6}$ seconds	$10^{6\pm3}$ GeV	$10^{19\pm3}$ K	End Re-heating, Begin Hot Big Bang ?
10^{-10} seconds	10^{2} GeV	10^{15} K	Electro-Weak Phase Transition?
10^{-4} seconds	10^{2} MeV	10^{12} K	Quark-Hadron Phase Transition?
10^{-2} seconds	10^{1} MeV	10^{11} K	Baryons-Leptons-Photons in equilibrium
10^{0} seconds	10^{0} MeV	10^{10} K	Neutrino Decoupling, $e^{\pm}$ annihilation
10^{2} seconds	10^{-1} MeV	10^{9} K	Primordial Nucleo-Synthesis
10^{4} years	10^{0} eV	10^{4} K	Matter-Radiation Equality
10^{5} years	10^{-1} eV	10^{3} K	Photon Decoupling
10^{9} years	10^{-3} eV	10^{1} K	First Bound Structure
Now	$\sim 10^{-3}$ eV	2.73 K	The Present

7.2 The Cosmic Microwave Background Radiation

In building up our knowledge of the universe, we uses several different kinds of observations in conjunction with certain theoretical models. One class of observations is the observations of structures in the universe i.e. (super) clusters of galaxies, voids, filaments etc and their statistics. From the table discussed earlier, all these correspond to matter dominated era and our current observations go back to about $z \sim 7$ (300,000 years after the big bang). Clearly, there is a huge range (essentially infinite) of red shift values that are still to be subjected to observations. One of the crucial tool for these observations is the *Cosmic Microwave Background Radiation* (CMBR) alluded to earlier.

According to the Big Bang model, universe would have gone through an epoch where protons, electrons, photons and neutral atoms (hydrogen) would have been in equilibrium. After a drop of temperature to about 4000^0K, the photons would decouple and stream freely carrying with them the information at the decoupling epoch. These photons constitute the CMBR. Observe that we *cannot* get a direct snapshot of period prior to decoupling by electromagnetic observations since during the plasma epoch all prior information would have been washed out. If we could observe the analogously predicted *neutrino background*, then we could have a similar snapshot of a much earlier epoch. But this is beyond our means. It turns out that the *angular distribution* of CMBR photons contains a wealth of information allowing us to constrain models of much earlier era. This is what we will discuss briefly.

The CMBR was first predicted by George Gamow and his collaborators in the late 40's when they were trying to obtain abundance of chemical elements via the hot big bang. Their prediction remained unnoticed since their main goal of chemical abundances did not work out. It could not have worked out since we now know that except the very light nuclei, all others are cooked in the interiors of stars where not only are the temperatures high but also the densities. The prediction of CMBR was effectively forgotten until it was discovered accidentally by Wilson and Penzias in 1965. Penzias and Wilson in fact were testing an antenna built to observe echo satellite and they observe a background 'hiss' not attributable to any particular direction in the sky. They reported an equivalent temperature (at wave length of 7.35 cm) of $3.5 \pm 1^0 K$. Its theoretical significance (identification with CMBR) was provided by Decke, Peebles, Roll and Wilkinson. This was of course observation at one frequency. Since then CMBR has been observed at wave lengths ranging from about 100 cm down to about 0.05 cm. Observations for larger wave lengths are limited by VHF radiation from our own galaxy while for shorter wavelengths ($<$ 3 cm) emission from our atmosphere interferes. The lower waveleghths are observed from balloon, rocket borne instruments and finally from the **CO**smic **B**ackground **E**xperiment satellite. These ranges cover both sides of the Planck distribution curve and the current value of the photon temperature is $2.725 \pm 0.001^0 K$.

Another striking feature of CMBR is its *isotropy*. Only at a level of about 1 part in 10^5 there are deviations from isotropy. Both the observed black body spectrum and isotropy are very strong corroboration of both the cosmological principle and the hot big bang model. For any other cosmological theory, these gross features of CMBR put stringent restrictions.

However, although small, *there are anisotropies!* Establishing their reality also took almost 25 years. Note isotropy of a distribution is an observer dependent statement. Suppose in one frame we find a distribution which is isotropic. The same distribution as seen by an observer moving relative to the first one, will have a 'dipole component'. The radiation received from the front direction will be blue shifted while that from the back direction will be red shifted. The radiation from other directions will also be shifted with shift determined by the component of the velocity in that direction. This will change the equivalent temperature thereby inducing an anisotropy in the angular distribution of temperature. This is expected and in fact will give our velocity relative to the isotropy frame. Conversely, if one observes only a dipole anisotropy, it implies that there exist a frame in which the distribution is isotropic. Additional anisotropies cannot be so removed by going to a different frame. The dipole anisotropy was found in the late seventies - early eighties. There were hints of 'quadrupole' anisotropies which were not conclusively. Finally by 1992, COBE established presence of anisotropies to $l = 30$ multipole. The recent **W**ilkinson **M**icrowave **A**nisotropy **P**robe (WMAP) made measurements to about $l \sim 500$.

There is a further bit to the story. The photon decoupling does not take place at the same time i.e. the *Last Scatter Surface* (LSS) is not a sharp surface but has a thickness. It is at a red shift of about 1100 with a thickness of about 80. This also has implications for the anisotropies.

The anisotropy data is presented in the following form. The basic observable quantity is the temperature in the direction $\hat{n}$: $T(\hat{n}) := \bar{T}(1 + \Delta T(\hat{n}))$. Here $\bar{T}$ is the temperature averaged over the directions and $\frac{\Delta T}{\bar{T}}(\hat{n})$ is taken as the definition of the measured anisotropy. This is expanded in the spherical harmonics as,

$$\Delta(\hat{n}) \ := \ \tfrac{\Delta T}{\bar{T}}(\hat{n}) \ = \ \sum_{\ell,m} a_{\ell,m} Y_{\ell,m}(\hat{n}) \tag{7.6}$$

The anisotropies are now encoded in the coefficients $a_{\ell,m}$. In principle, from the observed temperature distribution, one can infer the multipole coefficients $a_{\ell,m}$. In practice, however, a different quantity, *two point correlation function* is computed from the data because that is the quantity that can be theoretically computed from some proposed theory of the *reasons* for anisotropies. The two point correlation functions is defined as follows.

Consider two directions $\hat{n}, \hat{m}$ in the sky. Compute the average over the sky of the quantity $\Delta(\hat{n})\Delta(\hat{m})$ *keeping an angle α between the two directions to be fixed*, i.e.

$$< \Delta(\hat{n})\Delta(\hat{m}) > \ := \ \int d\Omega(\hat{n}) \int d\Omega(\hat{m}) \delta(\hat{n}\cdot\hat{m} - cos(\alpha))\Delta(\hat{n})\Delta(\hat{m}) \tag{7.7}$$

$$= \ \frac{1}{4\pi} \sum_{\ell} a_\ell^2 P_\ell(cos(\alpha)) \quad , \quad a_\ell^2 \ := \ \sum_{m=\ell}^{\ell} |a_{\ell,m}|^2 \tag{7.8}$$

A theory is supposed to give a prediction for the $a_{\ell,m}$.

Now, the anisotropies are supposed to be arising due to some *fluctuations* in the plasma. There are several possible sources for these eg (i) non-homogeneous gravitational potential which will red/blue shift the photons (leads to the so called "acoustic peaks"); (ii) the photons scatter off the electrons which introduces a Doppler shift and (iii) intrinsic fluctuations of the radiation field which could be present.

These fluctuations are generically assumed to be "Gaussian" which means that the two point correlation functions are characterized by a single number. Thus, theoretical models assume that the coefficients $a_{\ell,m}$, for each (ℓ, m), are distributed in a Gaussian form and what one expects to see in the data are their average values. This is stated in the form,

$$< a^*_{\ell,m} a_{\ell',m'} >_{\text{ensemble}} \ = \ C_\ell \delta_{\ell,\ell'} \delta_{m,m'} \qquad \Rightarrow C_\ell = \frac{a_\ell^2}{2\ell + 1} \tag{7.9}$$

The theoretical models compute C_ℓ while CMBR data gives the measured values. A representative plot of C_ℓ against ℓ is shown in the figure.

It turns out that the location of the peaks as well as their heights are sensitive to the parameters of the theoretical models and the data is able to constrain these severely. The theoretical models go way back before nucleo-synthesis and thus CMBR is able to indirectly give information about much earlier epochs. Furthermore, relating the CMBR fluctuations to matter fluctuations one is able to infer the possible seeds for subsequent structure formation. Measuring the *polarization* of the CMBR photons and analyzing their anisotropies gives further information including detection of the *first star formation*. The reason is that once stars are formed, the stellar processes contribute polarized photons.

In summary, CMBR is first a confirmation of the Big Bang model, its anisotropies contain on the one hand clues about earlier era and also a correlation with seeds for structure in much later era. As an observational tool, it is considered as heralding the age of precision (observational) cosmology [14].

In the next lecture we will consider some of the 'chinks in the armer' of the Big Bang Model!

Chapter 8

Appendix I

This is a summary of basic definitions which also serves to state some of the conventions.

1. A **Chart** (u_α, ϕ_α) around a point $p \in M$ means that $p \in u_\alpha$ and ϕ_α gives local coordinates around $p : \phi_\alpha(p) \leftrightarrow (x^1(p), x^2(p), \cdots, x^n(p))$.

2. An **Atlas** is a collection of *compatible* charts such that the u_α provide an open cover of underlying topological space and compatibility refers to *coordinate transformations* for overlapping u_α, u_β being differentiable (C^∞) with a differentiable inverse.

3. *Equivalence classes of Atlases* with respect to the compatibility relation defines **Differentiable Structures**.

4. By a **Manifold** we will always mean a connected, locally connected, Hausdorff topological space with a C^∞ structure of dimension n; typically denoted by M.

5. A **Differentiable Function** $f : M \to \mathbb{R}$ means that $f(x^i)$ is a differentiable (C^∞) function of the n variables which are the local coordinates.

6. A **Differentiable Curve** γ on M means a map $\gamma : (a, b) \to M \leftrightarrow (x^1(t), \cdots, x^n(t)) \in \gamma,\ t \in (a, b)$ and $x^i(t)$ are differentiable functions of the single variable t.

7. A **Tangent Vector to M at p** is an operator, $\frac{d}{dt}|_\gamma$ associated with every smooth curve γ through p, which maps smooth functions on M to real numbers by the expression:

$$\frac{d}{dt}f|_\gamma := \lim_{\epsilon \to 0} \frac{f(\gamma(\epsilon)) - f(\gamma(0) = p)}{\epsilon} \quad = \quad \frac{dx^i(t)}{dt}|_\gamma \frac{\partial}{\partial x^i} f \ .$$

   The set of all tangent vectors is naturally a vector space of dimension n and is called the **Tangent Space**. It is denoted by $T_p(M)$.

   Every chart (i.e. local coordinate system) around p gives a natural basis for $T_p(M)$, namely, $\{\frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^n}\}$ and is called a *coordinate basis.* A *generic basis* is denoted by $\{E_a, a = 1, \cdots, n\}$.

8. The vector space **Dual to** $T_p(M)$ is called the **Cotangent Space** and is denoted by $T^*_p(M)$. The basis *dual* to $\{\frac{\partial}{\partial x^i}\}$ is denoted as $\{dx^1, \cdots, dx^n\}$ and satisfies, $dx^i(\partial_j) = \delta^i_j$. Likewise, the basis dual to a generic basis $\{E_a\}$ is denoted by $\{E^a\}$ and satisfies, $E^a(E_b) = \delta^a_b$.

9. Given the tangent and the cotangent spaces at p, $T_p(M), T_p^*(M)$ one defines *tensor products* of these as:

$$(\Pi^s_r)_p \ := \ \underbrace{T_p^* \otimes \cdots \otimes T_p^*}_{\text{r-factors}} \ \otimes \ \underbrace{T_p \otimes \cdots \otimes T_p}_{\text{s-factors}}$$

This is a vector space of dimension $(n)^{r+s}$ and its elements are ordered $(r+s)-$tuples:

$$(\omega^1, \cdots, \omega^r, X_1, \cdots, X_s) \ \in \ (\Pi^s_r)_p \ \ \Leftrightarrow \ \ \omega^i \in T_p^* \ \text{ and } \ X_j \in T_p \ .$$

A **Tensor of rank (r, s) at $\boldsymbol{p \in M}$** is a real valued function $T : (\Pi^s_r)_p \to \mathbb{R}$ which is *linear* in each of its arguments. r is called the *contravariant rank* and s is called the *covariant rank*. Evidently, a tensor of rank (r, s) is an element of the vector space *dual* to $(\Pi^s_r)_p$. The dual vector space is denoted as $\mathbb{T}^r_s$.

Given a basis E_a of T_p and its dual basis E^a of T_p^*, one defines *basis tensors*,

$$E_{a_1\cdots a_r}{}^{b_1,\cdots,b_s} := E_{a_1} \otimes \cdots E_{a_r} \otimes E^{b_1} \otimes \cdots E^{b_s}$$

such that

$$E_{a_1\cdots a_r}{}^{b_1,\cdots,b_s}(E^{c_1}, \cdots, E^{c_r}, E_{d_1}, \cdots, E_{d_s}) := \delta^{c_1}_{a_1} \cdots \delta^{b_s}_{d_s}$$

A generic tensor is then expanded as:

$$T = \sum T^{a_1\cdots a_r}{}_{b_1\cdots b_s} E_{a_1\cdots a_r}{}^{b_1,\cdots,b_s} \quad \Leftrightarrow \quad T^{a_1\cdots a_r}{}_{b_1\cdots b_s} = T(E^{a_1}, \cdots, E^{a_r}, E_{b_1}, \cdots, E_{b_s})$$

The $T^{a_1\cdots a_r}{}_{b_1\cdots b_s}$ are the components of the tensor. When specialized to coordinate bases, they have the familiar transformation under a change of local coordinates:

$$(T')^{i_1\cdots i_r}{}_{j_1\cdots j_s}(x') = \frac{\partial (x')^{i_1}}{\partial x^{m_1}} \cdots \frac{\partial (x')^{i_r}}{\partial x^{m_r}} \frac{\partial x^{n_1}}{\partial (x')^{j_1}} \cdots \frac{\partial x^{n_s}}{\partial (x')^{j_s}} (T)^{m_1\cdots m_r}{}_{n_1\cdots n_s}(x)$$

The vector space structure takes care of the operations of addition of tensors and of scalar multiplication.

There are three more common operations: *tensor (or outer) product, interior product and contractions*. These are defined as,

Tensor Product (Outer Product):

$$(T_1 \times T_2)(\omega^1, \cdots, \omega^{r_1}, \omega^{r_1+1}, \cdots, \omega^{r_1+r_2}; X_1, \cdots, X_{s_1}, X_{s_1+1}, \cdots, X_{s_1+s_2}) :=$$

$$T_1(\omega^1, \cdots, \omega^{r_1}; X_1, \cdots, X_{s_1}) \ T_2(\omega^{r_1+1}, \cdots, \omega^{r_1+r_2}; X_{s_1+1}, \cdots, X_{s_1+s_2})$$

In terms of components:

$$(T_1 \times T_2)^{a_1\cdots a_{r_1} a_{r_1+1}\cdots a_{r_1+r_2}}{}_{b_1\cdots b_{s_1} b_{s_1+1}\cdots b_{s_1+s_2}} :=$$

$$(T_1)^{a_1\cdots a_{r_1}}{}_{b_1\cdots b_{s_1}} \ (T_2)^{a_{r_1+1}\cdots a_{r_1+r_2}}{}_{b_{s_1+1}\cdots b_{s_1+s_2}}$$

Interior Products: There are two of these, one with an element X of the tangent space and one with an element ω of the cotangent space.

$$(i_X T)(\omega_1, \cdots, \omega_r; X_1, \cdots, X_{s-1}) := T(\omega_1, \cdots, \omega_r; X, X_1, \cdots, X_{s-1}) \quad \Leftrightarrow$$

$$(i_X T)^{a_1,\cdots,a_r}{}_{b_1,\cdots,b_{s-1}} := X^b\ (T)^{a_1,\cdots,a_r}{}_{b,b_1,\cdots,b_{s-1}}$$

$$(i_\omega T)(\omega_1,\cdots,\omega_{r-1};X_1,\cdots,X_s) := T(\omega,\omega_1,\cdots,\omega_{r-1};X_1,\cdots,X_s) \quad \Leftrightarrow$$

$$(i_\omega T)^{a_1,\cdots,a_{r-1}}{}_{b_1,\cdots,b_s} := \omega_a\ (T)^{a,a_1,\cdots,a_{r-1}}{}_{b_1,\cdots,b_s}$$

Contraction:

$$T(\omega_1,\cdots,\omega_{r-1};X_1,\cdots,X_{s-1}) := T(\omega_1,\cdots,E^a,\cdots,\omega_{r-1};X_1,\cdots,E_a,\cdots,X_{s-1}) \quad \Leftrightarrow$$

$$T^{a_1,\cdots,a_{r-1}}{}_{b_1,\cdots,b_{s-1}} := T^{a_1,\cdots,c,\cdots,a_{r-1}}{}_{b_1,\cdots,c,\cdots,b_{s-1}}$$

10. A tensor of rank $(0,k)$ is called a **k-form** if it satisfies:

$$T(X_1,\cdots,x_i,\cdots,x_j,\cdots,x_k) = -T(X_1,\cdots,x_j,\cdots,x_i,\cdots,x_k) \ \ \forall \ \ i,j$$

These are *completely antisymmetric* covariant tensors of rank k. Evidently, $0 \le k \le n$ must hold.

Given any tensor of rank $(0,k)$ we can always construct a $k-form$ by the process of *antisymmetrization*:

$$(\text{anti } T)(X_1,\cdots,X_k) := \frac{1}{k!}\sum_{\sigma\in S_k}\text{sign}(\sigma)\ T(X_{\sigma(1)},\cdots,X_{\sigma(k)}) \quad \Leftrightarrow$$

$$(\text{anti } T)_{a_1,\cdots,a_k} := \frac{1}{k!}\sum_{\sigma\in S_k}\text{sign}(\sigma)\ T_{a_{\sigma(1)},\cdots,a_{\sigma(k)}} \quad := \quad T_{[a_1,\cdots,a_k]}$$

The space all $k-forms$ forms a vector space, denoted as Λ^k and has the dimension ${}^nC_k$.

Denote by Λ the *direct sum* of all these Λ^k : $\Lambda = \sum_{k=0}^n \ \oplus\Lambda^k$.

On Λ one defines the **Exterior (or Wedge) Product**. Let ω be a p-form and η be q-form such that $p+q \le n$. Then we define the *wedge product* of these to be the (p + q)-form, denoted as $\omega\wedge\eta$, by,

$$\omega\wedge\eta := \frac{(p+q)!}{p!q!}\text{anti }[\omega\otimes\eta]$$

In terms of components,

$$\begin{aligned}(\omega\wedge\eta)_{a_1,\cdots,a_{p+q}} &= \frac{(p+q)!}{p!q!}\omega_{[a_1,\cdots,a_p}\ \eta_{a_{p+1},\cdots,a_{p+q}]} \\ &= \frac{1}{p!q!}\sum_{\sigma\in S_{p+q}}\text{sign}(\sigma)\ \omega_{a_{\sigma(1)},\cdots,a_{\sigma(p)}}\ \eta_{a_{\sigma(p+1)},\cdots,a_{\sigma(p+q)}}\end{aligned}$$

These definitions, in particular the normalization factors, imply:

$$\begin{aligned}(\omega\wedge\eta)\wedge\zeta &= \omega\wedge(\eta\wedge\zeta) && \text{Associativity of wedge product} \\ \omega\wedge\eta &= (-1)^{pq}\eta\wedge\omega && \text{Commutation property}\end{aligned}$$

This takes care of the **Tensor Algebra** that we will need.

11. **Exterior Differentiation:** The *exterior differentiation* is defined for *k-forms* to produce a *(k + 1)-form*. It is defined as:

$d : \Lambda^k \to \Lambda^{k+1} \quad , k = 0, 1, \cdots, n$ such that

(i) for $f \in \Lambda^0$, $d(f) := df \in \Lambda^1$ is given by, $df(X) = X(f) \quad \forall\ X \in T_p(M)$. In local coordinates, $df = \frac{\partial f}{\partial x^i} dx^i$. This is called the *differential* of f.

(ii) For ω of higher ranks, express it in terms of its expansion in a coordinate basis,

$$\begin{aligned} \omega &= \omega_{[i_1,\cdots,i_k]} dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \qquad i_1 < i_2 < \cdots < i_k \\ &= \frac{1}{k!} \omega_{[i_1,\cdots,i_k]} dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \quad \text{unrestricted sum,} \end{aligned}$$

its exterior derivative is then defined by,

$$\begin{aligned} d\omega &= (d\omega_{[i_1,\cdots,i_k]}) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}, \qquad i_1 < i_2 < \cdots < i_k \quad \text{where} \\ d\omega_{[i_1,\cdots,i_k]} &= \sum_{i_{k+1}=1}^{n} \left(\frac{\partial \omega_{[i_1,\cdots,i_k]}}{\partial x^{i_{k+1}}} \right) dx^{i_{k+1}} \quad \in \Lambda^1. \end{aligned}$$

Alternatively, the components of $d\omega$ are also given by,

$$(d\omega)_{i_1 \cdots i_{k+1}} = (k+1) \partial_{[i_1} \omega_{i_2 \cdots i_{k+1}]}$$

Some of its basic properties are:

(a) The exterior differentiation is obviously a linear operation.

(b) $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^p \omega \wedge d\eta \quad \forall \quad \omega \in \Lambda^p,\ \eta \in \Lambda^q$.
Due to the presence of sign factor, this is called the *anti-derivation property.*

(c) $d^2\omega = 0 \quad \forall \quad \omega \in \Lambda$ (Nil-Potency property).

(d) If d' is any other map from $\Lambda^k \to \Lambda^{k+1}$ satisfying *linearity, anti-derivation, nil-potency and the action on functions producing their differential*, then such a map coincides with the exterior differentiation defined above. In other words, the four properties *uniquely characterize* exterior differentiation.

(e) $\omega \in \Lambda^k$ is called a **Closed Form** if $d\omega = 0$ and it is called an **Exact Form** if it can be expressed as $\omega = d\xi$, where $\xi \in \Lambda^{k-1}$. Clearly, every exact form is closed but the converse need not be true.
Denote: $Z^k :=$ the (vector) space of all closed k-forms ($d\omega = 0,\ \forall\ p \in M$) and $B^k :=$ the vector space of all exact k-forms, $B^k \subset Z^k$. Define $H^k := Z^k/B^k$, i.e. the space of all closed forms modulo exact forms. This vector space is called the $\boldsymbol{k^{th}}$ **Cohomology Class of** $\boldsymbol{M}$. For *compact manifolds*, its dimension is *finite*, $b^k := \dim H^k$, and is called the $\boldsymbol{k^{th}}$ **Betti Number** of the manifold. This number turns out to be a *Topological Invariant.*

(f) **Poincare Lemma:** Every closed form is *locally* (i.e. in a contractible neighborhood) is exact. In particular, $\mathbb{R}^n$ being contractible, *all* closed forms are exact and hence all its Betti numbers are zero.
Exercise: *For S^1 compute b^1.*

12. **Lie Differentiation:** This is defined by using diffeomorphisms generated by vector fields, $X^i\partial_i$ (locally: $x^i \to (x')^i := x^i + \epsilon X^i(x)$). Abstractly, for each smooth vector field X on M, it is defined as a map $\mathcal{L}_X : \mathbb{T}^r_s \to \mathbb{T}^r_s$ satisfying the following properties:

(a) It is *linear*;

(b) $\mathcal{L}_X f := X(f) \quad \forall \;\; f : M \to \mathbb{R}$;

(c) $\mathcal{L}_X \; Y := [X, Y] \quad \forall$ vector fields Y on M;

(d) $\mathcal{L}_X(S \otimes T) := (\mathcal{L}_X S) \otimes T + S \otimes \mathcal{L}_X T)$. In particular,
$\mathcal{L}_X(\langle \omega, Y \rangle) := \langle \mathcal{L}_X \omega, Y \rangle + \langle \omega, \mathcal{L}_X Y \rangle, \;\; \forall \; \omega$, 1-forms and $\forall \; Y$, vector fields, on M.

The corresponding *local expressions* are:

(a) $\mathcal{L}_X f = X^i \frac{\partial}{\partial x^i} f(x)$;

(b) $\mathcal{L}_X Y = \left[X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \right] \frac{\partial}{\partial x^i}$;

(c) $\mathcal{L}_X \omega = \left[\omega_j \frac{\partial X^j}{\partial x^i} + X^j \frac{\partial \omega_i}{\partial x^j} \right] dx^i$;

(d) More generally, one can show (*Prove this!*):
$\mathcal{L}_X \omega = i_X d\omega + d(i_X \omega), \;\; \forall \;\; \omega \in \Lambda^k, k = 0, \cdots, n$; It follows that $d\mathcal{L}_X \omega = \mathcal{L}_X d\omega$, i.e. the Lie-derivative and the exterior derivatives commute. (*prove this*).

13. **Covariant Differentiation:** Let $X, Y, \cdots$ denote smooth vector fields on M and let $S, T, \cdots$ denote tensor fields of rank (r, s). Let $\nabla_X : \mathbb{T}^r_s \to \mathbb{T}^r_s$ denote a *family* of maps, labelled by vector fields X, satisfying the following properties:

(a) ∇_X is linear;

(b) $\nabla_X(f) := X(f) \quad \forall \;\; f : M \to \mathbb{R}$;

(c) $\nabla_{fX+gY}(T) = f\nabla_X(T) + g\nabla_Y(T) \;\; \forall$ *functionsf, g* and vector fields X, Y ;

(d) $\nabla_X(S \otimes T) = (\nabla_X S) \otimes T + S \otimes (\nabla_X T)$ and in particular,
$\nabla_X \langle \omega, Y \rangle = \langle \nabla_X \omega, Y \rangle + \langle \omega, \nabla_X Y \rangle$;

Then $\nabla_X T$ is called *a* **Covariant derivative of T with respect to X**.

Note: This is similar to the definition of the Lie derivative. It *differs crucially* in the property (13c). Also, while Lie derivative of vector fields is specified as part of its definition, there is no such stipulation for covariant derivative. These differences allow *several different* covariant derivatives to be defined. Given a family ∇_X satisfying the above properties, one can define a map $\nabla : \mathbb{T}^r_s \to \mathbb{T}^r_{s+1}$ by,

$$(\nabla T)(\eta^1, \cdots, \eta^r; X, X_1, \cdots, X_s) := (\nabla_X T)(\eta^1, \cdots, \eta^r; X_1, \cdots, X_s)$$

This map ∇ is well defined *provided* ∇_X satisfies the property (13c).

The freedom in the possible maps ∇_X is parametrized (locally) by an **Affine Connection**, $\boldsymbol{\Gamma}$, introduced via the covariant derivatives of vector fields E_a:

$$\nabla_{E_b} E_c := \Gamma^a{}_{bc} E_a, \qquad \nabla_{\partial_j} \partial_k := \Gamma^i{}_{jk} \partial_i$$

Note that the right hand sides in the above equations being vector fields they are expressed as linear combinations of the basis vector fields and the expansion coefficients are the 'components' of the affine connection.

Exercise: Changing to a different coordinate basis and using the definition of the corresponding components, deduce the transformation law for the components of the affine connection and verify that the affine connection is *not* a tensor.

The familiar 'semicolon notation' for covariant derivatives is obtained as follows. For a (contravariant) vector field, $A := A^i\partial_i$ denote: $\nabla_{\partial_i} A := A^j{}_{;i}\partial_j$.

$$
\begin{array}{rcll}
\nabla_{\partial_i}(A^j\partial_j) & = & (\nabla_{\partial_i}A^j)\partial_j + A^j\nabla_{\partial_i}\partial_j & \Longrightarrow \\
A^k{}_{;i}\partial_k & = & (\partial_i A^j)\partial_j + A^j\Gamma^k{}_{ij}\partial_k & \Longrightarrow \\
A^k{}_{;i} & = & A^k{}_{,i} + \Gamma^k{}_{ij}A^j & \text{The usual definition.}
\end{array}
$$

Exercise: For a 1-form field $B := B_i dx^i$, denote $\nabla_{\partial_i} B := B_{j\,;i}dx^j$ and show that $B_{k\,;i} = B^k{}_{,i} - \Gamma^j{}_{ik}B_j$.

Watch out for the position of the lower indices since Γ *is not necessarily symmetric in these.*

14. **Parallel Transport and Affine Geodesics:** We have defined covariant derivative of a tensor field T, along a vector field X, as $\nabla_X T$. Let $X = X^i\partial_i$ in some coordinate neighborhood around a point p. Let γ be an *integral curve* of X through p, i.e. around p, $X^i(\gamma(t)) = \frac{dx^i(t)}{dt}$. Then,

$$
\begin{array}{rcl}
\nabla_X T & = & \nabla_{X^i\partial_i}T \ = \ X^i\nabla_{\partial_i}T, \qquad \text{Denote:} \quad \nabla_i := \nabla_{\partial_i} \\
& = & X^i\nabla_i T \quad := \quad X\cdot\nabla T \\
& = & \dfrac{dx^i}{dt}\nabla_i T \\
& = & \dfrac{dx^i}{dt}(\partial_i T \pm \quad \text{connection terms.}) \\
& = & \dfrac{dT(x^i(t))}{dt} \pm \dfrac{dx^i}{dt}\ \text{times connection terms.}
\end{array}
$$

Therefore, *if* $\nabla_X T$ ($= X\cdot\nabla T) = 0$, then we get a first order, ordinary differential equation for $T(x^i(t))$. This always has a solution in a sufficiently small neighborhood $t \in (-\epsilon, \epsilon)$ and the solution is uniquely determined by giving the initial value; $T(p)$. Therefore, given a tensor at p and a vector field X, we can determine a tensor *along an integral curve of* X *through* p. The tensor so determined is called a **Tensor parallelly transported along** $\boldsymbol{\gamma}$**.** Notice that this is determined by the connection.

What is parallel about it? If the connection vanished, then the parallelly transported tensor just equals the tensor at p i.e. is "parallel" in the intuitive sense.

Thus, by definition, *a tensor parallelly transported along* X satisfies: $X\cdot\nabla T_{||} = 0$. A non-zero covariant derivative thus measure the the deviation from "parallality".

Such parallelly transported tensors are defined for arbitrary rank. In particular, one can consider parallel transport of X along itself. In general, this will be non-zero. Equivalently, $X_{||} \not\sim X$. However, for special cases of vector fields we may actually find $X\cdot\nabla X = 0$. The *integral curves of such a vector field are called (Affinely parametrized)* **Affine Geodesics**. If we allow X to satisfy $X\cdot\nabla X \propto X$, then integral curves of such vector fields are called *non-affinely parametrized* affine geodesics.

Exercise: Derive the coordinate form of the geodesic equations $(X\cdot\nabla X)^i = 0$. Show that a non-affinely parametrized geodesic can always be re-parametrized to an affine parameterization.

Although an affine connection is not a tensor, one can construct two natural tensors from it and its derivatives.

15. **The Torsion Tensor:** Given an affine connection (or covariant derivative) via ∇_X (or ∇), one naturally defines the **Torsion Tensor** T as:

$$T(\omega, X, Y) := \langle\ \omega, \nabla_X Y - \nabla_Y X - [X, Y]\ \rangle \quad \forall\ \ \omega, X, Y.$$

Clearly, this is a tensor of rank $(1, 2)$ and is manifestly antisymmetric in its covariant rank arguments. To show that this is well defined (i.e. does define a tensor) one has to show: $T(f\omega, gX, hY) = fghT(\omega, X, Y)\ \ \forall\ \ functions f, g, h$. The stipulated properties of ∇_X are crucial for this proof.

Exercise: Show that $T^i{}_{jk} := T(dx^i, \partial_j, \partial_k) = \Gamma^i{}_{jk} - \Gamma^i{}_{kj}$.

An affine connection is said to **Symmetric** if its Torsion tensor is zero.

Exercise: For a symmetric connection, show that,

$$\mathcal{L}_X Y = \nabla_X Y - \nabla_Y X \quad \Leftrightarrow \quad (\mathcal{L}_X Y)^i = X^j Y^i{}_{;j} - Y^j X^i{}_{;j}\ .$$

16. **The Riemann Curvature Tensor and the Ricci Tensor:** Given an affine connection one naturally defines another tensor of rank $(1, 3)$, called the **Riemann Curvature Tensor** as:

$$R(\omega, Z, X, Y) := \langle\ \omega, \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]} Z\ \rangle\ \ \forall\ \ \omega, X, Y, Z.$$

Exercise: Show that $R^i{}_{jkl} := R(dx^i, \partial_j, \partial_k, \partial_l)$ are given by,

$$R^i{}_{jkl} \ = \ \partial_k \Gamma^i{}_{lj} - \partial_l \Gamma^i{}_{kj} + \Gamma^i{}_{km}\Gamma^m{}_{lj} - \Gamma^i{}_{lm}\Gamma^m{}_{kj}$$

The definition is independent of the torsion being zero or non-zero.

The **Ricci Tensor** is a tensor of rank (0,2) and is defined as:

$$R(X, Y) \ := \ R(E^a, X, E_a, Y) \quad \Leftrightarrow \quad R_{ij} \ := \ R^k{}_{ikj}\ .$$

17. **Cartan Structural Equations:** The definitions associated with an affine connection imply certain *identities* which can be interpreted as alternative *definitions* of the curvature and the torsion tensors. To see this recall (and define) for generic bases, E_a, E^a:

$$\nabla_{E_b} E_c := \Gamma^a{}_{bc} E_a \ \ ; \ \ [E_b, E_c] := C^a{}_{bc} E_a \ \ ; \ \ \mathcal{E}^a{}_b := \Gamma^a{}_{cb} E^c \ \ \text{(Connection 1-forms);}$$

$$\begin{aligned}
T^a{}_{bc} &:= T(E^1, E_b, E_c) \\
&= \Gamma^a{}_{bc} - \Gamma^a{}_{cb} - C^a{}_{bc}; \\
R^a{}_{bcd} &:= R(E^a, E_b, E_c, E_d) \\
&= E_c(\Gamma^a{}_{db}) - E_d(\Gamma^a{}_{cb}) + \Gamma^a{}_{cf}\Gamma^f{}_{db} - \Gamma^a{}_{df}\Gamma^f{}_{cb} - \Gamma^a{}_{fb}C^f{}_{cd}; \\
T^a &:= \frac{1}{2} T^a{}_{bc} E^b \wedge E^c; \qquad \text{Torsion 2-forms} \\
R^a{}_b &:= \frac{1}{2} R^a{}_{bcd} E^c \wedge E^d \qquad \text{Curvature 2-forms.}
\end{aligned}$$

These definitions imply the relations:

$$\begin{aligned} dE^a &= -\mathcal{E}^a{}_b \wedge E^b + \frac{1}{2}T^a{}_{bc}E^b \wedge E^c \\ d\mathcal{E}^a{}_b &= -\mathcal{E}^a{}_c \wedge \mathcal{E}^c{}_b + \frac{1}{2}R^a{}_{bcd}E^c \wedge E^d \end{aligned}$$

These are rewritten as (the **Cartan Structural Equations**) :

$$\begin{aligned} T^a &= dE^a + \mathcal{E}^a{}_b E^b; \\ R^a{}_b &= d\mathcal{E}^a{}_b + \mathcal{E}^a{}_c \wedge \mathcal{E}^c{}_b \end{aligned}$$

Note: The connection, the torsion and the Riemann curvature have been defined in a manifestly coordinate (or basis) independent manner. If an arbitrary basis is used and components relative to this are obtained, the these must satisfy the Cartan structural *equations.*

In practice, these are also used to compute the connection 1-forms and curvature 2-forms especially when the torsion vanishes. The structural equations immediately imply the two famous identities: the cyclic identity and the Bianchi identity by simply taking the exterior derivative of these equations.

These are *identities* in the sense that these are valid for all affine connections and for for all choices of bases.

18. **The Cyclic Identity:**

$$\begin{aligned} dT^a &= 0 + d\mathcal{E}^a{}_b \wedge E^b - \mathcal{E}^a{}_b \wedge dE^b \\ &= (R^a{}_b - \mathcal{E}^a{}_c \wedge \mathcal{E}^c{}_b) \wedge E^b - \mathcal{E}^a{}_b \wedge (T^b - \mathcal{E}^b{}_c \wedge E^c) \\ &= R^a{}_b \wedge E^b - \mathcal{E}^a{}_b \wedge T^b \end{aligned}$$

Exercise: Specializing to coordinate bases and using the explicit definitions of wedge products, covariant derivatives etc show that the above relation in terms of forms is equivalent to:

$$\sum_{(jkl)} R^i{}_{jkl} = \sum_{(jkl)} T^i{}_{jk;l} + \sum_{(jkl)} T^i{}_{mj} T^m{}_{kl}$$

The (jkl) denotes sum over cyclic permutations of the indices.

The right hand side is zero for a symmetric connection and is the more familiar form of the cyclic identity.

19. **The Bianchi Identity:**

$$\begin{aligned} dR^a{}_b &= 0 + d\mathcal{E}^a{}_c \wedge \mathcal{E}^c{}_b - \mathcal{E}^a{}_c \wedge d\mathcal{E}^c{}_b \\ &= (R^a{}_c - \mathcal{E}^a{}_d \wedge \mathcal{E}^d{}_c) \wedge \mathcal{E}^c{}_b - \mathcal{E}^a{}_c \wedge (R^c{}_b - \mathcal{E}^c{}_d \wedge \mathcal{E}^d{}_b) \\ &= R^a{}_c \wedge \mathcal{E}^c{}_b - \mathcal{E}^a{}_c \wedge R^c{}_b \end{aligned}$$

Exercise: Specializing to coordinate bases, show that this is equivalent to:

$$\sum_{(klm)} R^i{}_{jkl;m} = \sum_{(klm)} R^i{}_{jkn} T^n{}_{lm}$$

Again the right hand side vanishes for symmetric connection and is the more familiar form of the Bianchi identity.

Convince yourself that one can not obtain any more identities from the structural equations.

20. **The Ricci Identities:** There is another set of *identities* known as the *Ricci identities* which are usually given in component form relative to coordinate bases. In a local approach, these are also used to *define* the curvature tensor. These are obtained by evaluating double covariant derivatives on an arbitrary tensor and antisymmetrizing.

    Recall that covariant derivative of a tensor is tensor and so is its double covariant derivative. However, only for an *antisymmetric combination*, the result has a term independent of derivatives of the tensor and a term involving a covariant derivative of the tensor. The coefficients involve the curvature and the torsion tensors respectively.

    Exercise: Using the definitions: $\nabla_i B_j := \partial_i B_j - \Gamma^k{}_{ij} B_k$ and $\nabla_i A^j := \partial_i A^j + \Gamma^j{}_{ik} A^k$ show that,

$$\begin{aligned}(\nabla_l \nabla_k - \nabla_k \nabla_l) A^i &= -R^i{}_{jkl} A^j + T^j{}_{kl}\ \nabla_j A^i \\ (\nabla_l \nabla_k - \nabla_k \nabla_l) B_j &= \quad R^i{}_{jkl} B_i\ + T^i{}_{kl}\ \nabla_i B_j\ .\end{aligned}$$

    These extend to arbitrary rank tensors in an obvious manner (index-by-index).

21. **Implications of Curvature and Torsion:**

    (a) An infinitesimal parallelogram with all sides being geodesics exists *iff* the Torsion tensor vanishes.

    (b) A tensor field T satisfying $\nabla_X T = 0$ exists through out a neighborhood u_p *iff* the Riemann tensor vanishes in the neighborhood. Riemann $= 0$ is thus an *integrability* condition for a parallelly transported tensor field to be definable in a neighborhood.

    (c) A tensor field, parallelly transported along a closed (and contractible) loop equals the original tensor *iff* the Riemann tensor vanishes.

    Therefore, in general, geodesics which begin as parallel do not remain so subsequently. Curvature is thus a measure of *geodesic deviation*. See item (27).

Notice that we have got all the notions of geodesics, curvature etc *without* introducing any *metric tensor*.

22. **The Metric Tensor:** A *symmetric tensor field* $\mathbf{g}$ *of type* $(0, 2)$ is called a **Metric Tensor** field on the manifold. This is of course to be distinguished from the (metric =) distance function introduced while motivating the definition of topology.

    At any point p, we can define a *symmetric Matrix*, $g_{ab} := g(E_a, E_b)$ by choosing a basis for the tangent space. This can always be diagonalised by a real linear, orthogonal basis transformation and by scaling the basis vectors (or local coordinates in case of coordinate basis) can be further brought to a form:

$$g(e_i, e_j) = \eta_{ij} = \eta_i \delta_{ij},\ \ \eta_i = \pm 1, 0\ .$$

    Let $n_\pm, n_0$ be the number of *positive, negative and zero* values of $\eta_i, n = n_+ + n_- + n_0$. These numbers are characteristic of the matrix i.e. are *independent* of the initial basis chosen to obtain the matrix. Furthermore, on a connected manifold and smooth metric tensor, these numbers *cannot* change from point-to-point and are thus characteristic of the metric tensor itself.

The metric tensor g is said to be **Non-Degenerate** if $n_0 = 0$. In this case, one can define a smooth tensor field, g^{-1} of the rank $(2, 0)$ such that at every point, $g^{ab} := g^{-1}(E^a, E^b)$ satisfies, $g^{ab} = g^{ba}, g^{ac}g_{cb} = \delta^a_b$. g^{-1} is naturally called the **Inverse Metric Tensor**. In practice, one does not use a separate symbol for the inverse metric, it is inferred from the index positions.

n_- is called the **Index of g, ind(g)** while $n_+ - n_-$ is called the **Signature of g, sig(g)**.

For the case of $ind(g) = 0$, the metric is said to **Riemannian** (or *Euclidean*); otherwise it is generically called **Pseudo-Riemannian**. When $n_- = n - 1$, the metric is said to be **Lorentzian** (or *Minkowskian*).

Our Convention: $diag g \sim (+1, -1, \cdots, -1)$ and we will be considering only *non-degenerate, Lorentzian signature* metrics.

Such *Manifolds with metric* will be referred to as *Pseudo-Riemannian manifold or* **Space-Times**.

Basic *existence results: (See Hawking-Ellis)*

(a) Any *paracompact* manifold admits a *Riemannian* metric;

(b) Any *non-compact, paracompact* manifold admits a *Lorentzian* metric;

(c) A compact manifold admits a *Lorentzian* metric *iff* its *Euler character*, $\chi(M) := \sum_{k=0}^{n}(-1)^k b^k$, is zero.

23. **Weyl, Diffeomorphism and Conformal Equivalences and Isometries:** There are many different notions of *equivalence* in use. These are:

    (a) Two metrics g_1, g_2 are said to be **Weyl Equivalent** if $g_2 = e^{\Phi} g_1$ for some smooth $\Phi : M \to \mathbb{R}$.

    (b) Two metrics g_1, g_2 are said to be **Diffeomorphism Equivalent** if $g_2 = \phi^* g_1$ for some *diffeomorphism* $\phi : M \to M$ and ϕ^* denotes the corresponding *pull-back map.*

    (c) Two metrics g_1, g_2 are said to be **Conformally Equivalent** if there exists a diffeomorphism $\phi : M \to M$ such that $g_2 = e^{\Psi}(\phi^* g_1)$ for some smooth function $\Psi : M \to \mathbb{R}$.

    (d) A *diffeomorphism* $\phi : M \to M$ is said to be an **Isometry of a metric g**, if $\phi^* g = g$. Likewise, it said to be a **Conformal Isometry of g** if $\phi^* g = e^{\Psi} g$ for some smooth $\Psi : M \to \mathbb{R}$.

24. **Extra Operations available due to a Metric Tensor:** There are many additional features that a manifold with metric acquires. Since a non-degenerate metric gives us both g_{ab} and g^{ab}, it allows us to set up a *canonical* (standard/natural) isomorphism between the tangent and the cotangent spaces. In other words it allows us to **raise and lower indices** of tensors of rank (r, s). (This is a property of second rank, non-degenerate tensors. In Hamiltonian formulation one has the anti-symmetric non-degenerate $(0, 2)$ tensor – the symplectic 2-form – which also plays a similar role. It leads to symplectic geometry.)

(a) A metric defines a *unique* affine connection via the
Result: *Given a non-degenerate metric, there exists a unique affine connection,* Γ *satisfying,*
(i) $T^i{}_{jk}(\Gamma) = 0$;
(ii) $\nabla_k g_{ij} = 0 \ \ \forall \ \ i, j, k.$
The condition (ii) alone allows us to obtain the affine connection as:

$$\Gamma^k{}_{ij} = \left\{ \frac{1}{2} g^{kl} \left(g_{lj,i} + g_{li,j} - \ g_{ij,l} \right) \right\} - \frac{1}{2} \left\{ g_{im} T^m{}_{jn} g^{nk} + g_{jm} T^m{}_{in} g^{nk} \right\} + \frac{1}{2} T^k{}_{ij}$$

For the zero-torsion case, the connection is given only by the first term and is called the **Riemann-Christoffel Connection** of the *metric connection*. This is the connection used in general relativity.

All the definitions of curvature etc are immediately applicable for this special connection. However, in addition now one can also define the **Ricci Scalar** $R := g^{ij} R_{ij}$.

Because of the vanishing torsion and availability of raising and lowering of indices, the Riemann tensor has further properties under interchange of its indices. These are summarized in the item 25.

(b) A metric tensor also allows us to define an *invariant volume form*, the *Hodge Dual*, the *co-differential* and the Laplacian. These are seen as follows.

i. Recall that Λ^n is one dimensional. An n-form at p, $\omega \in \Lambda^n$ is said to be a **Volume Element** at p. Two volume elements are said to be *equivalent* if $\omega_2 = \lambda \omega_1, \lambda > 0$. This is an equivalence relation and has exactly *two* equivalence classes which are called **Orientations on $\boldsymbol{\Lambda^n}$**. The n-form $\omega := E^1 \wedge \cdots \wedge E^n$ always defines a *volume element.*

A basis $\{E_a\}$ for $T_p(M)$ is said to be **Positively Oriented with respect to $[\boldsymbol{\omega}]$** if $\omega(E_1, \cdots, E_n) > 0$.

An n-form field μ on M is said to be **Volume Form** on M if $\mu(p) \neq 0, \ \forall \, p \in M$.

M is said to be **Orientable** if it admits a volume form and is said to be **Oriented** if a particular choice of volume form is made. This definition of orientability turns out to be equivalent to the one given in terms of the sign of the Jacobian of coordinate transformations in the overlapping charts.

Locally,

$$\begin{aligned}
\mu &= \frac{1}{n!} \mu_{i_1 \cdots i_n} dx^{i_1} \wedge \cdots \wedge dx^{i_n} \ = \ \mu_{1 \cdots n} dx^1 \wedge \cdots \wedge dx^n \\
&= \mu'_{1 \cdots n} dx'^1 \wedge \cdots \wedge dx'^n \\
&= \mu'_{1 \cdots n} \frac{\partial x'^1}{\partial x^1} \cdots \frac{\partial x'^n}{\partial x^n} dx^1 \wedge \cdots \wedge dx^n \\
&= \frac{1}{n!} \mu_{i_1 \cdots i_n} \frac{\partial x'^{i_1}}{\partial x^{i_1}} \cdots \frac{\partial x'^{i_n}}{\partial x^{i_n}} dx^{i_1} \wedge \cdots \wedge dx^{i_n} \qquad \Rightarrow \\
\mu_{i_1 \cdots i_n} &= \mu'_{j_1 \cdots j_n} \frac{\partial x'^{j_1}}{\partial x^{j_1}} \cdots \frac{\partial x'^{i_n}}{\partial x^{i_n}} \\
&= \left(\det \frac{\partial x'}{\partial x} \right) \mu'_{j_1 \cdots j_n}
\end{aligned}$$

The components of an n-form thus transform by a determinant. Such a quantity is called a **Tensor Density**.
It follows that $\sqrt{|\det g_{ij}|}$ transforms as,

$$\sqrt{|\det g'_{ij}|} = |\left(\det \frac{\partial x}{\partial x'}\right)| \sqrt{|\det g_{ij}|}$$

It is apparent now that $\mu_g := \sqrt{|\det g_{ij}|} dx^1 \wedge \cdots \wedge dx^n$ defines a volume form (since the metric is non-degenerate) and is invariant under coordinate transformations. Notationally this **Invariant Volume Form** is also denoted as

$$\mu_g := \sqrt{|\det g_{ij}|} dx^1 \wedge \cdots \wedge dx^n := \sqrt{g}\ d^n x\ .$$

ii. **Levi-Civita Symbol $\boldsymbol{\mathcal{E}_{i_1 \cdots i_n}}$:**

$$\mathcal{E}_{i_1 \cdots i_n} \ := \ \begin{cases} 1 & \text{if } i_1 \cdots i_n \text{ is an even permutation of } (1 \cdots n) \\ -1 & \text{if } i_1 \cdots i_n \text{ is an odd permutation of } (1 \cdots n) \\ 0 & \text{otherwise.} \end{cases}$$

This allows us to write,

$$dx^1 \wedge \cdots \wedge dx^n = \frac{1}{n!} \mathcal{E}_{i_1 \cdots i_n} dx^{i_1} \wedge \cdots \wedge dx^{i_n} \qquad \text{etc.}$$

iii. On Λ^k, the space of k-forms, define an *inner product (or pairing)* as,

$$(\omega, \eta)|_p = \frac{1}{k!} \omega_{i_1 \cdots i_k} \eta^{i_1 \cdots i_k}|_p \quad , \quad \eta^{i_1 \cdots i_k} := g^{i_1 j_1} \cdots g^{i_k j_k} \eta_{i_1 \cdots i_k}\ .$$

It is obvious $(\omega, \eta) = (\eta, \omega)$ (symmetry) and $(\omega, \eta) = 0\ \forall\ \eta \Rightarrow \omega = 0$ (non-degeneracy).
Now the **Hodge Isomorphism (or Hodge * operator)** is defined as: $* : \Lambda^k \to \Lambda^{n-k}$ such that

$$\alpha \wedge (*\beta) := (\alpha, \beta) \mu_g \quad \forall \quad \alpha \in \Lambda^k \qquad \text{This defines} \quad * \beta.$$

Exercise: Show that

$$\begin{aligned} \alpha \wedge *\beta &= \beta \wedge *\alpha\ ; \\ *\beta &= (-1)^{\text{index(g)}}\ (-1)^{k(n-k)}\ \beta; \\ (*\alpha, *\beta) &= (-1)^{\text{index(g)}} (\alpha, \beta)\ . \end{aligned}$$

Exercise: Using these definitions obtain the local expressions for components of $*\beta$:

$$(*\beta)_{i_1 \cdots i_{n-k}} = \frac{1}{k!} (-1)^{k(n-k)} \epsilon_{i_1 \cdots i_{n-k} j_1 \cdots j_k} \beta^{j_1 \cdots j_k} \quad , \quad \epsilon_{i_1 \cdots i_n} := \mathcal{E}_{i_1 \cdots i_n} \sqrt{g}\ .$$

Note: The Levi-Civita symbol is just a numerical quantity and as such is *not* subject to coordinate transformations. The $\epsilon_{i_1 \cdots i_n}$ on the other hand transforms under coordinate transformations due to the explicit factor of $\sqrt{g}$. Indeed,

$$\begin{aligned} \epsilon'_{i_1\cdots i_n} &:= \mathcal{E}_{i_1\cdots i_n}\sqrt{g'} \\ &= |det\frac{\partial x}{\partial x'}|\epsilon_{i_1\cdots i_n} \\ &= \frac{\partial x^{j_1}}{\partial x^{i_1}}\cdots\frac{\partial x^{j_1}}{\partial x^{i_1}}\epsilon_{i_1\cdots i_n} \ . \end{aligned}$$

Thus $\epsilon_{i_1\cdots i_n}$ transforms as a *tensor density* of rank (0,n).

iv. On k-form fields we defined the *exterior differential* $d : \Lambda^k \to \Lambda^{k+1}$. With a non-degenerate metric tensor available, one also defines the **Co-Differential** $\boldsymbol{\delta}$ as: $\delta : \Lambda^k \to \Lambda^{k-1}$,

$$\delta\omega := (-1)^{\text{index(g)}}(-1)^{nk+n+1} * d * \omega \ .$$

Exercise: Show that $\delta^2\omega = 0 \ \forall \ \omega \in \Lambda$.

The nil-potency of δ allows us to define:
ω is said to be **Co-Closed** if $\delta\omega = 0$;
It is said to be **Co-Exact** if it can be written as $\omega = \delta\xi, \xi \in \Lambda^{k+1}$;
It is said to be **Harmonic** if it is both closed and co-closed, $d\omega = 0 = \delta\omega$.

Using the Exterior differential and the co-differential one defines the **Laplacian Operator** on k_forms as: $\boldsymbol{\Delta := d\delta + \delta d}$. Evidently it maps k-forms to k-forms.

v. On n-dimensional manifolds *only* integrals of n-forms are well defined. These are locally given by,

$$\int_M \omega := \int dx^1 \wedge \cdots \wedge dx^n \omega_{1\cdots n} := \int d^n x(\omega_{1\cdots n}) \ .$$

On the space of smooth $k - formfields$ one defines a bilinear, symmetric, non-degenerate quadratic form:

$$\langle\omega|\eta\rangle := \int_M \omega \wedge *\eta = \frac{1}{k!}\int_M \omega_{i_1\cdots i_k}\eta^{i_1\cdots i_k}\sqrt{g}d^n x \ .$$

Exercise: For Riemannian manifolds without boundary show that

$$\langle\omega|\delta\eta\rangle = \langle d\omega|\eta\rangle \ .$$

For the case of a Riemannian metric, index(g) = 0, the d and δ are *Adjoints* of each other and the Laplacian is "Self-Adjoint" (for suitable boundary conditions). One can then also write the **Hodge Decomposition** (which is an *orthogonal decomposition*) for any k-form as:

$$\omega = \alpha + d\beta + \gamma \ \ , \ \ d\alpha = 0 \ , \ d\gamma = 0 = \delta\gamma \ .$$

25. **Number of Independent Components of the Riemann Tensor for the Metric Connection (without Torsion):** Availability of metric tensor allows us to define

$R_{ijkl} := g_{im}R^m{}_{jkl}$. Use of the Riemann-Christoffel connection, which implies zero torsion, simplifies many expressions. These are summarized as:

$$
\begin{aligned}
R_{ijkl} &= -R_{ijlk} && \text{From definition ;} \\
\sum_{(jkl)} R_{ijkl} &= 0 && \text{Cyclic identity;} \\
\sum_{(klm)} R_{ijkl;m} &= 0 && \text{Bianchi identity;} \\
(\nabla_l\nabla_k - \nabla_k\nabla_l)T^{i_1\cdots i_m}{}_{j_1\cdots j_n} &= -\sum_{\sigma=1}^{m} R^{i_\sigma}{}_{jkl}T^{i_1\cdots j\cdots i_m}{}_{j_1\cdots j_n} && \text{Ricci Identities} \\
&\quad + \sum_{\sigma=1}^{n} R^{i}{}_{j_\sigma kl}T^{i_1\cdots i_m}{}_{j_1\cdots i\cdots j_n}
\end{aligned}
$$

Further Properties:

$$
\begin{aligned}
R_{ijkl} &= -R_{jikl} \\
R_{ijkl} &= R_{klij} \\
R_{ij} &= R_{ji} && \text{Symmetry of Ricci Tensor} \\
R &:= g^{ij}R_{ji} && \text{The Ricci Scalar} \\
G_{ij} &:= R_{ij} - \frac{1}{2}Rg_{ij} && \text{The Einstein Tensor} \\
\nabla_j G^{ij} &= 0 && \text{“Contracted Bianchi Identity”.}
\end{aligned}
$$

The calculation of the number independent components the Riemann tensor is slightly tricky due to the various symmetries and the cyclic identities.

Given (ijkl) consider sub-cases (i) two of the indices are equal, eg R_{ijil} with $i \neq j, i \neq l$, (ii) two pairs of indices are equal eg R_{ijij} and (iii) all indices are unequal. For the first two sub-cases, the cyclic identities give no conditions (are trivially satisfied). The number of components in case (i) is $\frac{n(n-1)}{2} \times (n-2)$. For the case (ii), the number is $\frac{n(n-1)}{2}$. For the case (iii) a priori we have $n(n-1)(n-2)(n-3)$. Since $i \leftrightarrow j, k \leftrightarrow l, (ij) \leftrightarrow (kl)$ are the same components we divide by $2 \cdot 2 \cdot 2 = 8$. The cyclic identity is non-trivial and allows one term to be eliminated in favor of the other two. This gives the number to be $\frac{2}{3}\frac{1}{8}n(n-1)(n-2)(n-3)$. Thus, the **total number of independent components** is given by,

$$
\frac{n(n-1)(n-2)}{2} + \frac{n(n-1)}{2} + \frac{1}{12}n(n-1)(n-2)(n-3) \quad = \quad \frac{n^2(n^2-1)}{12} \ .
$$

For $n = 2$, the number of independent components is just 1 and the Riemann tensor is explicitly expressible as:

$$
R_{ijkl} = \frac{R}{2}(g_{ik}g_{jl} - g_{jk}g_{il}) \ .
$$

For $n = 3$, the number of independent components is 6 and equals the number of independent components of the Ricci tensor. One can express,

$$
R_{ijkl} = (g_{ik}R_{jl} - g_{jk}R_{il} - g_{il}R_{jk} + g_{jl}R_{ik}) - \frac{1}{2}R(g_{ik}g_{jl} - g_{jk}R_{il}) \ .
$$

For $n \geq 4$, the number of independent components of the Riemann tensor is *larger* than those of the Ricci tensor plus the Ricci scalar. Hence in these cases, the Riemann tensor *cannot* be expressed in terms of R, R_{ij}, g_{ij} alone. We need the "fully traceless" Weyl or Conformal tensor.

26. **The Weyl Tensor:** This is a combination of the Riemann tensor, the Ricci tensor, the Ricci scalar and the metric tensor which vanishes when any pair of indices is 'traced' over by the metric (contracted by the metric). It is given by,

$$C_{ijkl} := R_{ijkl} - \frac{1}{n-2}(g_{ik}R_{jl} - g_{jk}R_{il} - g_{il}R_{jk} + g_{jl}R_{ik}) + \frac{1}{(n-1)(n-2)}(g_{ik}g_{jl} - g_{jk}R_{il}) \ .$$

27. **Geodesic Deviation – Relative Acceleration:** In the following the connection is a metric connection.

Consider a *smooth, 1-parameter family of affinely parametrized geodesics*, $\gamma(t,s)$ so that for each fixed $\hat{s}$ in some interval, $\gamma(t,\hat{s})$ is a geodesic. Smoothness of such a family means that there is a map from $(t,s) \in I_1 \times I_2$ into M and this map is smooth. Let this map be denoted locally as $x^i(t,s)$.

We naturally obtain two vector fields tangential to the embedded 2-surface: $u^i(s,t) := \frac{\partial x^i(s,t)}{\partial t}$ and $X^i(s,t) := \frac{\partial x^i(s,t)}{\partial s}$. The former is *tangent to a geodesic* and hence $u \cdot \nabla x^i = 0$. The latter is called a *generic deviation vector*. From the smoothness of the family (i.e. existence of 2-dimensional embedded surface) it follows that $[\partial_t, \partial_s] = 0$ and this translates into (for torsion free connection) $X \cdot \nabla u^i = u \cdot \nabla X^i$.

Claim: By an $s-$*dependent* affine transformation of t one can ensure that $X \cdot \nabla u^2 = 0$.

Corollary: u^2 is independent of t, s and $u \cdot X$ is a function of s alone.

Claim: For non-null geodesics $u^2 \neq 0$, it is possible to make a further affine transformation to arrange $u \cdot X = 0$.

In other words, for a family of time-like or space-like geodesics it is possible to arrange the parameterization such that the deviation vector is orthogonal to the geodesic tangents. One defines:

$X^i, \quad X \cdot u = 0$	the **Displacement vector**;
$v^i := u \cdot \nabla X^i$	the **Relative Velocity**;
$a^i := u \cdot \nabla v^i$	the **Relative Acceleration**.

By contrast, for any curve, $Y \cdot \nabla Y^i$ *is called the* **Absolute Acceleration**.

It follows:

$$\begin{aligned}
a^i &= u^j \nabla_j (u^k \nabla_k X^i) = u \cdot \nabla (X \cdot \nabla u^i) \quad ([X,u]=0) \\
&= X^j u \cdot \nabla(\nabla_j u^i) + (\nabla_j u^i) u \cdot \nabla X^j \\
&= X^j u^k \nabla_k \nabla_j u^i + (\nabla_j u^i) X \cdot \nabla u^j \\
&= X^j u^k \nabla_j \nabla_k u^i - R^i{}_{kjl} u^k X^j u^l + (X \cdot \nabla u^j) \nabla_j u^i \\
&= (X \cdot \nabla)(u \cdot \nabla u^i) - R^i{}_{kjl} u^k X^j u^l \quad \text{Or,} \\
\boldsymbol{a^i} &= \boldsymbol{-R^i{}_{jkl} u^j X^k u^l} \qquad \textbf{The Deviation Equation.}
\end{aligned}$$

Chapter 9

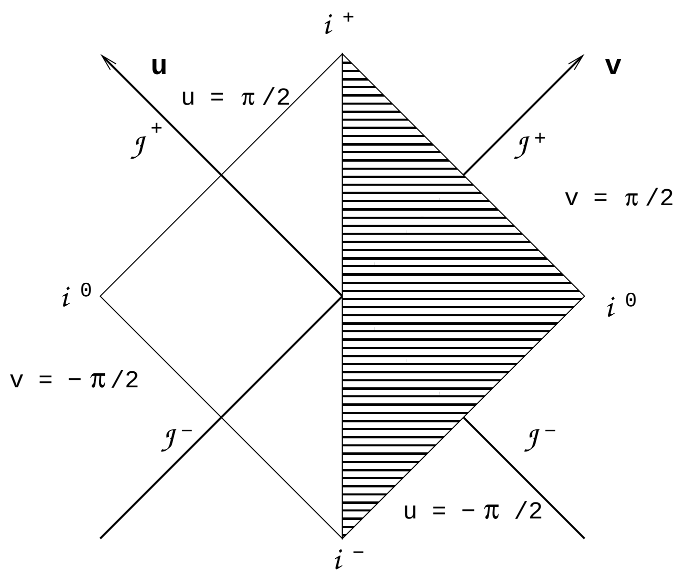
Appendix II

The following is some additional material which could be of use. Again only simplest of the cases are discussed and further details are to be found in the references included at the end.

Conformal Diagrams (Penrose Diagrams)

There are diagrams which enable one to represent the space-time as finite regions. These arise out of discussion of "Asymptotic Flatness". In the following only Minkowski and Schwarzschild space-times are discussed.

The Minkowski space-time:



The metric in the standard t, r, θ, ϕ coordinates is,

$$ds^2 = dt^2 - dr^2 - r^2 d\Omega^2.$$

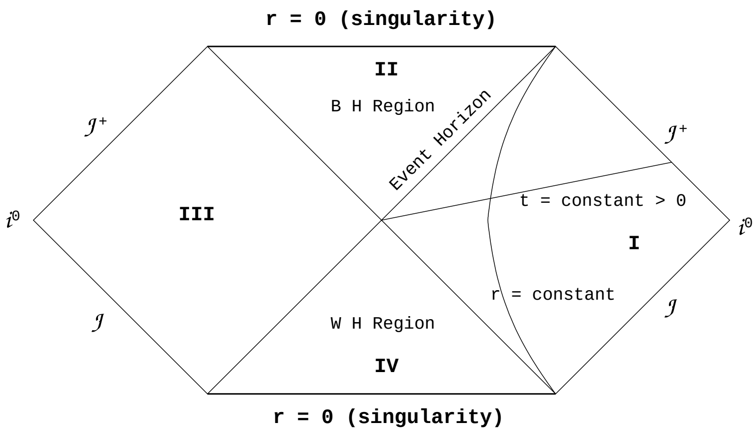
Define $U \equiv t - r$ and $V \equiv t + r$. Suppressing the angular part,

$$ds^2 \ = \ dUdV$$

Clearly $r \geq 0$ implies $V \geq U$ but otherwise both U, V range over the full real line. Now define $U \equiv tan(u)$ and $V \equiv tan(v)$. The u and v now range over $(-\pi/2, \pi/2)$ giving the diamond shape shown in the figure. The Minkowski space-time (θ, ϕ suppressed) is the shaded portion reflecting the restriction $v \geq u$. The terminology used for various points/segments of the diagram is shown in the figure.

<u>Problem:</u> *If one considers 2 dimensional Minkowski space-time how will the corresponding diagram look like?*

<u>The (Kruskal) Extended Schwarzschild space-time:</u>



The metric in the standard t, r, θ, ϕ coordinates is, (θ, ϕ part suppressed)

$$ds^2 \ = \ (1 - 2m/r) \ \{ \ dt^2 - \frac{1}{(1 - 2m/r)^2} dr^2 \ \}$$

In terms of the "tortoise coordinate", $r_\star$,

$$r_\star \ \equiv \ r + 2m \ \ell n \ (\ |r/2m - 1| \ ),$$

the metric is:

$$ds^2 \ = \ (1 - 2m/r) \ \{dt^2 - dr_\star^2\}$$

Define $U \equiv -e^{(r_\star - t)/4m}$ and $V \equiv e^{(r_\star + t)/4m}$. We see that $U \leq 0$ and $V \geq 0$ and that the metric is non-singular across $r = 2m$. The Kruskal extension now consists of keeping the same form of the metric but allowing U, V to range over full real line. Further defining $U \equiv T - X$ and $V \equiv T + X$ one has the familiar Kruskal form of the metric:

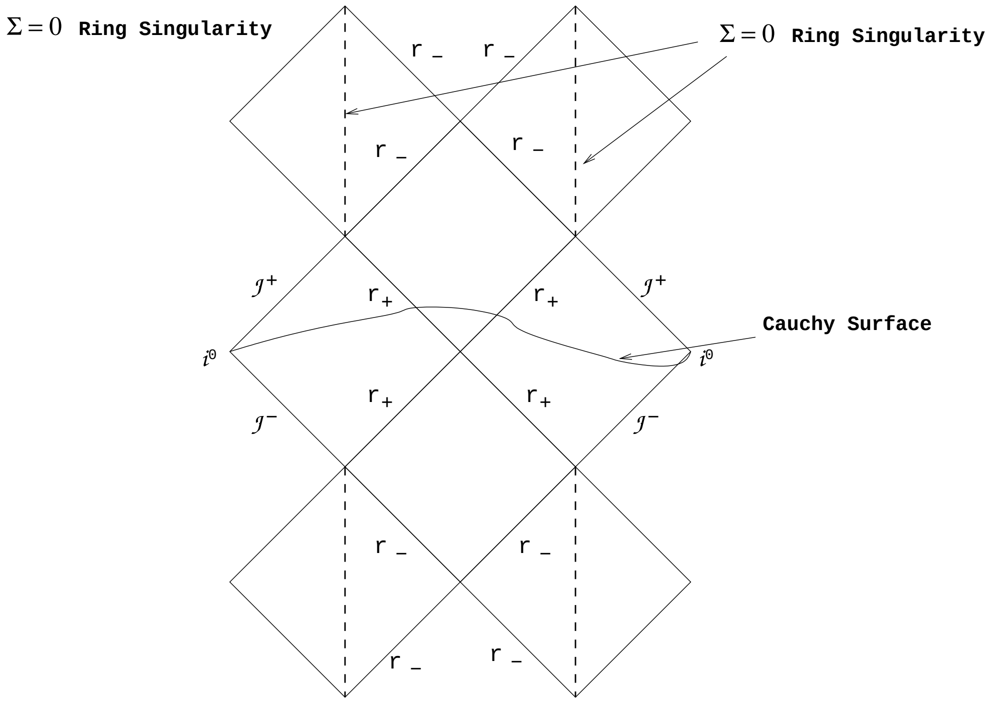
$$ds^2 \ = \ \frac{32m^3}{r} e^{-r/2m} \ (dT^2 - dX^2)$$

with r defined implicitly in terms of T, X by,

$$X^2 - T^2 \ = \ (\frac{r}{2m} - 1) \ e^{r/2m}$$

If one wants one can obtain t in terms of T, X but is not essential. The condition that $r > 0$ translates in to $T^2 - X^2 < 1$ (or $UV < 1$). As in the Minkowski case one can obtain a bounded picture by defining $U \equiv tan(u)$ and $V \equiv tan(v)$. The u, v range over $(-\pi/2, \pi/2)$ and the $r = 0$ $UV = 1$ translates in to $u + v = \pm\pi/2$. (Do you see this?). In terms of u, v coordinates the extended Schwarzschild space-time is shown in the figure above.

Similar analysis is done for the Kerr-Newman family of solutions. The resulting Penrose diagram is shown in the figure below.



The above are all examples of the so called Asymptotically flat space-times. The general definition in essence stipulates that an asymptotically flat space-time has as (conformal)"boundary" components three crucial segments: the *the future Null infinity ($\mathcal{J}^+$), the past Null infinity ($\mathcal{J}^-$) and the Spatial infinity (i^0)* .

<u>Black Holes and "Uniqueness" Results</u>

A general definition of a *Black Hole* space-time requires it to be *larger* than the set of points from which one can send physical signals (time-like or null curves) to the Future Null Infinity. The "extra" region is the Black Hole region, its *boundary* is the Event Horizon (3 dimensional) while the intersection of the event horizon with a "Cauchy surface" (eg constant t surface in the above examples) is the more familiar 2-dimensional event horizon. The event horizon is always a Null hyperface (3 dimensional surface whose normal is light-like)

As an exercise identify the black hole region and event horizon in the Kerr-Newman example. For precise definitions see the Wald's book for instance.

<u>Some general results about black holes.</u>

<u>Result a :</u> *A black hole at "instant" Σ (a Cauchy surface) may never bifurcate.*

This result which says that a black hole may never disappear (Classically of course) does not even use Einstein's equations, and follows purely from the definition of black hole and topological arguments.

The event horizon at instant Σ is a 2 dimensional surface and the induced metric on it gives us the definition of its area. This is defined as the AREA of an instantaneous black hole. For *stationary* black holes the 2 dimensional surface is compact and thus has finite area.

Problem: *Take the metric for the Kerr-Newman solution. From the Penrose diagram notice that $r = r_+, t =$ constant is the instantaneous event horizon. Find the induced metric on this 2 dimensional surface. Integrate over the surface the $\sqrt{det(\text{ induced metric })}$ and compute the area.*

Result b : *The area of a black hole never decreases.*

This result depends on the condition $R_{\mu\nu}k^\mu k^\nu \geq 0 \ \forall$ null k^μ being true. Via the Einstein equations, this is translated in to a condition on the stress-energy tensor $T_{\mu\nu}$. These "energy conditions" are listed separately.

Now a few results for stationary, black hole solutions of source free Einstein equations are collected.

Result c : *Stationary, vacuum, black holes are either static or axisymmetric.*

Result d : *For stationary vacuum black holes, the instantaneous event horizon is topologically the two sphere, S^2.*

Result e : *For stationary vacuum black holes, the Killing vector ξ corresponding to stationarity, is tangential to the event horizon. Thus it has to be either space-like or light-like.*

If ξ is every where non space-like outside the horizon (No ergosphere) then on the horizon it is light-like. The solution then must be static.

If ergosphere is present but intersects the event horizon, then ξ is space-like on a portion of the event horizon. In this case there exist another Killing vector χ which commutes with ξ and is light-like on the event horizon. A linear combination ψ, of ξ and χ can be constructed which is space-like and whose orbits are closed. In other words the space-time is stationary and axisymmetric.

This leads to two further definitions:

$$\chi \ \equiv \ \xi \ + \ \Omega_H \ \psi;$$

Ω_H is called the "angular velocity" of the event horizon.

$$\nabla^\mu \ \chi^2 \ \equiv \ - 2\kappa\chi^\mu \qquad \text{On the event Horizon}$$

where κ is called the "surface gravity".

Result f : *The surface gravity is constant over the horizon*

This result depends on the stress-tensor satisfying the so called "dominant energy condition". This result allows the interpretation of κ being the "temperature".

A useful equivalent expression for the surface gravity is: Define:

$$V \equiv \sqrt{|\chi^2|} \ ,$$

$$a^\mu \ \equiv \ \frac{\chi \cdot \nabla \ \chi^\mu}{V^2} \ , \qquad a \ \equiv \ a^2$$

Then,

$$\kappa = lim(Va) = lim(\sqrt{V^2 a^\mu a_\mu})$$

Here lim means that the quantity is to be evaluated in the limit of approaching the horizon.

Problem: *For the Kerr-Newman solution find the angular velocity of the horizon.*

Problem: *For the Kerr-Newman solution find the surface gravity*

The algebra simplifies considerably if you first express χ^2 (away from the horizon) in the form $(..)\Delta + (..)\Delta^2$. Here Δ is the usual expression for the Kerr-Newman solution. The lim of course is the limit $r \to r_+$ or $\Delta \to 0$ You should get the answer stated in the class.

It is instructive to compute κ for the Schwarzschild solution using the basic definition of κ. You will notice a problem in using the t, r, θ, ϕ coordinates. Use of Kruskal coordinates will remove the problem. (Of course, for Schwarzschild solution, $\chi = \xi$) Try it!

NOTE: In the class the laws of black hole thermodynamics were derived *using* the particular explicit Kerr-Newman solution. For a general stationary black hole, without knowing its explicit form, the derivation is more involved. One needs to also define the Mass, Angular Momentum, Charge of such a black hole which is done in terms of the so called "Komar" integrals (expressions). For these details you have to see Wald for instance.

The Energy Conditions

The energy conditions, conditions that any stress-energy tensor $T_{\mu\nu}$ representing "physical" matter (sources of gravity) has to satisfy, essentially incorporate the qualitative feature of gravitational interactions that these are *always attractive* (for "positive masses" , at least classically). This means that nearby future directed causal geodesics (i.e. future directed time-like or light-like) tend to come closer. From analysis of families of such geodesics via the *Raychoudhuri equations*, this translates in to the statement that $R_{\mu\nu}k^\mu k^\nu \geq 0$, for all time-like or light-like vectors. Using the Einsteins equations, this is transferred to a statement about $T_{\mu\nu}$. There are *three different* conditions that are stipulated and various results use one or the other of these in the proofs. These are:

$T_{\mu\nu}v^\mu v^\nu \geq 0 \ \forall$ time-like v^μ		*Weak* energy condition
$T_{\mu\nu}v^\mu v^\nu \geq \frac{T^\alpha_\alpha}{2} \ \forall$ *normalized* time-like v^μ		*Strong* energy condition
$T_{\mu\nu}v^\nu$	be a future directed time-like or null vector $\forall$ future directed time-like v^μ	*Dominant* energy condition

For a given $T_{\mu\nu}$ in terms of density, pressure etc these conditions are expressed as conditions on density/pressure etc.

Problem: *For the perfect fluid stress-energy tensor used in the cosmological solution, express the weak and the strong conditions as conditions on the densities and pressures. What about the dominant condition?*

Chapter 10

Exercises

The Following set of problems are chosen to help you develop a feel for basic practical computations used in GTR. Some will give you further factual information.

Covariant Derivatives and Killing Vectors

Notation: $(Tensor);\mu \leftrightarrow \nabla_\mu(Tensor)$

These have following basic properties:

$$\begin{aligned}
\nabla_\mu \Phi &= \partial_\mu \Phi \quad \text{where } \Phi \text{ is a scalar} \\
\nabla_\mu (T_1 T_2) &= (\nabla_\mu T_1) T_2 + T_1 (\nabla_\mu T_2) \\
\nabla_\mu A^\nu &\equiv \partial_\mu A^\nu + \Gamma^\nu_{\mu\lambda} A^\lambda \\
\nabla_\mu B_\nu &\equiv \partial_\mu B_\nu - \Gamma^\lambda_{\mu\nu} B_\lambda
\end{aligned}$$

We choose the affine connection Γ by requiring that,

$$\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}, \qquad \nabla_\mu g_{\nu\lambda} = 0.$$

This gives us the Riemann-Christoffel connection.

Problem 1 By considering a coordinate transformation of the form,

$$x'^\mu = x^\mu + C^\mu_{\alpha\beta} x^\alpha x^\beta + ...$$

Show that $\Gamma'^\lambda_{\mu\nu}$ can be made zero at any given point x^μ. Conclude that the metric can always be expressed in the form:

$$g_{\mu\nu} = \eta_{\mu\nu} + \circ(x^2)$$

for sufficiently small x^μ.

Vector fields ξ^μ which satisfy the Killing equation:

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0$$

are called Killing vector fields.

Problem 2 Consider the usual 2-sphere with metric,

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

Find all possible Killing vectors on the sphere by solving the Killing equations.

Geodesics

Let $x^\mu(\lambda)$ denote a geodesic and $u^\mu \equiv \frac{dx^\mu}{d\lambda}$ denote the geodesic tangent vector. The geodesic equation can then be written as, $u^\nu \nabla_\nu u^\mu \ = \ 0$

Problem 3 Let ξ^μ be a Killing vector and let $K(\xi) \equiv u^\mu \xi_\mu$. Show that K is constant along the geodesic.

Problem 4 Consider Schwarzschild space-time ($r > 2m, t, r, \theta, \phi$ coordinates). We have 4 Killing vectors:

$$\xi^\mu_{(0)} = (1,\ 0,\ 0,\ 0), \qquad \xi^\mu_{(i)} = (0,\ 0,\ \xi^\theta_{(i)},\ \xi^\phi_{(i)}), \qquad i = 1, 2, 3.$$

The $\xi^\mu_{(i)}$ are the Killing vectors obtained in the problem 2 above. Let $J_i \equiv u^\mu \xi_\mu$ where u^μ is a geodesic tangent. Show that $J_i = 0$ for all i implies u^μ is a radial geodesic while $J_3 \neq 0$ implies equatorial geodesic. ($\xi^\mu_{(3)} = (0, 0, 0, 1)$).

Curvatures and Identities

Problem 5 Show that

$$\begin{aligned} \nabla_\mu \nabla_\nu A^\lambda - \nabla_\nu \nabla_\mu A^\lambda &= R^\lambda_{\alpha\mu\nu} A^\alpha \ , \\ \nabla_\mu \nabla_\nu B_\lambda - \nabla_\nu \nabla_\mu B_\lambda &= -R^\alpha_{\lambda\mu\nu} B_\alpha \end{aligned}$$

(Warning: Depending on how you have defined the Riemann tensor, the index positions as well signs on the right hand sides may be different. The relative sign is correct though. These expressions are sometimes used to *define* the Riemann tensor in terms of the Christoffel connections. Note also that the Riemann tensor is antisymmetric in the last two indices. The above expressions are known as the *Ricci identities*. One could generalize these for higher rank tensors.)

Problem 6 For any 1-form ω_μ show by direct computation,

$$\begin{aligned} \nabla_{[\mu} \nabla_\nu \omega_{\lambda]} &\equiv \frac{1}{3!}\{\nabla_\mu \nabla_\nu \omega_\lambda + \nabla_\nu \nabla_\lambda \omega_\mu + \nabla_\lambda \nabla_\mu \omega_\nu - \nabla_\nu \nabla_\mu \omega_\lambda - \nabla_\mu \nabla_\lambda \omega_\nu - \nabla_\lambda \nabla_\nu \omega_\mu\} \\ &= 0 \end{aligned}$$

Problem 7 Using the above and the Ricci identity deduce that,

$$R^\alpha_{\lambda\mu\nu} + R^\alpha_{\mu\nu\lambda} + R^\alpha_{\nu\lambda\mu} = 0$$

This is the *cyclic identity.* Alternative proof will also be OK.

Problem 8 Applying the Ricci identity to $g_{\mu\nu}$ show that,

$$R_{\mu\nu\alpha\beta} = -R_{\nu\mu\alpha\beta}$$

and deduce further that,

$$R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu}$$

Problem 9 Prove the *Bianchi identity*:

$$R^{\alpha}_{\beta\mu\nu;\lambda} + R^{\alpha}_{\beta\nu\lambda;\mu} + R^{\alpha}_{\beta\lambda\mu;\nu} = 0$$

The *Ricci tensor* and the *Ricci scalar* are defined as:

$$R_{\mu\nu} \equiv R^{\alpha}_{\mu\alpha\nu} \quad , \quad R \equiv g^{\mu\nu} R_{\mu\nu}$$

Problem 10 Contracting the Bianchi identity show that the Einstein tensor $G_{\mu\nu}$ satisfies,

$$\nabla^{\mu} G_{\mu\nu} = 0$$

Geodesic Deviation etc.

Consider a family of time-like geodesics i.e. $x^{\mu}(\tau, \sigma)$ where for each $\sigma, x^{\mu}(\tau)$ is a time like geodesic. Define,

$$\begin{aligned} u^{\mu}(\tau,\sigma) &\equiv \frac{\partial x^{\mu}(\tau,\sigma)}{\partial \tau} \quad \text{(geodesic tangent vector which is time like,)} \\ X^{\mu}(\tau,\sigma) &\equiv \frac{\partial x^{\mu}(\tau,\sigma)}{\partial \sigma} \quad \text{(geodesic “displacement” or deviation vector).} \end{aligned}$$

Such a family can always be chosen to satisfy further,

$$u^{\mu} u_{\mu} \ = \ \text{constant (= 1, say) and} \quad u^{\mu} X_{\mu} \ = \ 0$$

Terminology:

$$\begin{aligned} v^{\mu} &\equiv u \cdot \nabla X^{\mu} \quad \text{relative velocity (of nearby geodesics)} \\ a^{\mu} &\equiv u \cdot \nabla v^{\mu} \quad \text{relative acceleration (of nearby geodesics)} \end{aligned}$$

Problem 11 Noting that $\frac{\partial}{\partial \tau} = u^{\mu} \nabla_{\mu}$ and $\frac{\partial}{\partial \sigma} = X^{\mu} \nabla_{\mu}$, show that:

$$u \cdot \nabla X^{\mu} \ = \ X \cdot \nabla u^{\mu}.$$

Problem 12 Show that,

$$a^{\mu} \ = \ - R^{\mu}_{\nu\alpha\beta} u^{\nu} X^{\alpha} u^{\beta}.$$

This is the geodesic deviation equation. Clearly, the relative acceleration is zero *iff* the Riemann tensor vanishes.

Problem 13 In the Schwarzschild space-time consider a family of radial time-like geodesics with, say, $\phi = 0$ and θ playing the role of σ. Consider two geodesics in this family, with $X^\mu = (0, 0, X^\theta, 0)$. Compute the relative acceleration.

NOTE: You will need to compute some of the components (which ones?) of the Riemann tensor. Also notice that in the conventional units (CGS say), $g_{00} = 1 - \frac{2GM}{c^2 r}$. Taking, in CGS units,

$$G \sim 6 \cdot 10^{-8}, M \sim 10^{24}, c \sim 3 \cdot 10^{10} \text{ and } r \sim 6 \cdot 10^8$$

estimate a^μ. This corresponds to the relative acceleration of two test bodies dropped from the same height, same longitude but different latitude near the surface of Earth.

Red shifts

Problem 14 Using the definition of the electromagnetic field tensor,

$$\begin{aligned} F_{\mu\nu} &\equiv \partial_\mu A_\nu - \partial_\nu A_\mu = \nabla_\mu A_\nu - \nabla_\nu A_\mu \ , \\ \nabla^\mu F_{\mu\nu} &= 0 \quad \text{Maxwell equations,} \\ \nabla^\mu A_\mu &= 0 \quad \text{the gauge condition,} \end{aligned}$$

obtain the curved space wave equation for the potential A_μ.

Problem 15 For the ansatz $A_\mu = \mathcal{E}_\mu e^{i\Phi}$ where the phase Φ is a scalar, rewrite the wave equation in terms of $\mathcal{E}_\mu$ and Φ.

Problem 16 Neglect the Ricci tensor term and covariant derivatives of $\mathcal{E}_\mu$ and show that,

$$(\nabla_\mu \Phi)(\nabla^\mu \Phi) = 0 \text{ and } \nabla^2 \Phi = 0$$

This approximation is called the "geometrical optics approximation".

Thus if $k_\mu \equiv \nabla_\mu \Phi$ then $k^2 = 0$ and $\nabla \cdot k = 0$. Since $\nabla_\mu \Phi$ is normal to the hypersurface (3 dimensional) $\Phi =$ constant, k_μ is indeed the wave propagation vector.

Problem 17 Considering the gradient of k^2, show that

$$k^\nu \nabla_\nu k^\mu = 0 \quad \text{i.e. } k^\mu \text{ is a null geodesic tangent}$$

Thus *in the geometrical optics approximation, light propagates along null geodesics.*

Let u^μ ($u^2 = 1$) denote an observer using his/her proper time as the time coordinate. The *frequency* of a light wave as determined by this observer is given by,

$$\omega(u) \equiv k \cdot u \ (= u^\mu \nabla_\mu \Phi = u^\mu \partial_\mu \Phi = \frac{d\Phi}{d\tau_{proper}} \ )$$

Problem 18 In the Schwarzschild geometry, consider two *stationary* observers i.e. observers whose four velocities are proportional to the Killing vector. Observer O_1 at r $= r_1$ send a light wave of frequency ω_1 which is then received by observer O_2 at r $= r_2$ as a light wave of frequency ω_2. The respective frequencies are of course defined as $\omega_i = k \cdot u_i$ where

k^μ is the propagation vector for the light wave and u_i are the four velocities of the observers satisfying $u_i^2 = 1$. Using the result of problem 3 show that,

$$\frac{\omega_2}{\omega_1} = \frac{\sqrt{g_{00}}|_1}{\sqrt{g_{00}}|_2}$$

For Schwarzschild solution corresponding to Sun (solar mass $\sim 10^{33} gms$, solar radius $\sim 6 \cdot 10^{10} cms$ and Earth-Sun distance about 8 light minutes estimate the red shift.

All red shift calculations essentially proceed similarly.

Bibliography

[1] A. Einstein, *The Meaning of Relativity*, Princeton Univ, Press, 5th edition, 1954;

[2] G F Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill Book Company, Inc, 1963;
S. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge Univ. Press, 1973, first 50 pages;

[3] B. F. Schutz, *A First Course in General Relativity*, Cambridge Univ. Press, 1985;

[4] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, John Wiley & Sons, Inc., 1972;

For a detail analysis of physical basis of GR, particularly with regards to local aspects this is highly recommended. This is excellent for cosmological aspects.

[5] R. M. Wald, *General Relativity*, Univ. of Chicago Press, 1984;

for a somewhat more modern treatment and more advanced topics dealing with *global* aspects Wald's book is highly recommended.

[6] S. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge Univ. Press, 1973;

This contains the rigorous treatment and proofs in the global context. This is advance material but is terse.

[7] C. M. Will, *Living Rev. Relativity*, **4**, 4, 2001,
online article at: http://www.livingreviews.org/Articles/Volume4/2001-4will/ ;

[8] G. Date, *Gen. Rel. & Grav.*, **29**, 953, 1997;

See the appendix of this for the organization of the equations.

[9] S. Chandrasekhar, *The Mathematical Theory of Black Holes*, Oxford:Clarendon Press, 1983;

For very detailed and explicit treatment of black hole solutions, their properties and perturbations about these (gravitational waves) Chandrasekhar's book is THE resource material.

[10] M. Heusler, *Black Hole Uniqueness Theorems*, Cambridge University Press, 1996.

[11] A. Ashtekar, C. Beetle, S. Fairhurst, *Class. Quant. Grav.*, **16** , L1, 1999;

A. Ashtekar, S. Fairhurst, B. Krishnan, *Phys. Rev.*, **D 62**, 104025, 2000;

G. Date, *Class. Quant. Grav.*, **17**, 5025, 2000 ;

G. Date, *Class. Quant. Grav.*, **18**, 5219, 2001 ;

There are many more papers on this subject, this is only a sample.

[12] The astrophysical and cosmological parameters are available at the Particle Data Group site at *http://pdg.lbl.gov/*;

[13] G. Smoot and K. Davidson, *Wrinkles in Time*, William Morrow and Company, New York, 1993;

This is an interesting account of the efforts of the COBE team and the discovery of anisotropies in the Cosmic Microwave Background.

[14] Here are some informative web-sites on the CMB anisotropies, acoustic peaks etc:

http://background.uchicago.edu/ whu/ ;

http://map.gsfs.nasa.gov/

[15] T. Padmanabhan, *Theoretical Astrophysics, Volume III: Galaxies and Cosmology*, Cambridge University Press, 2002.