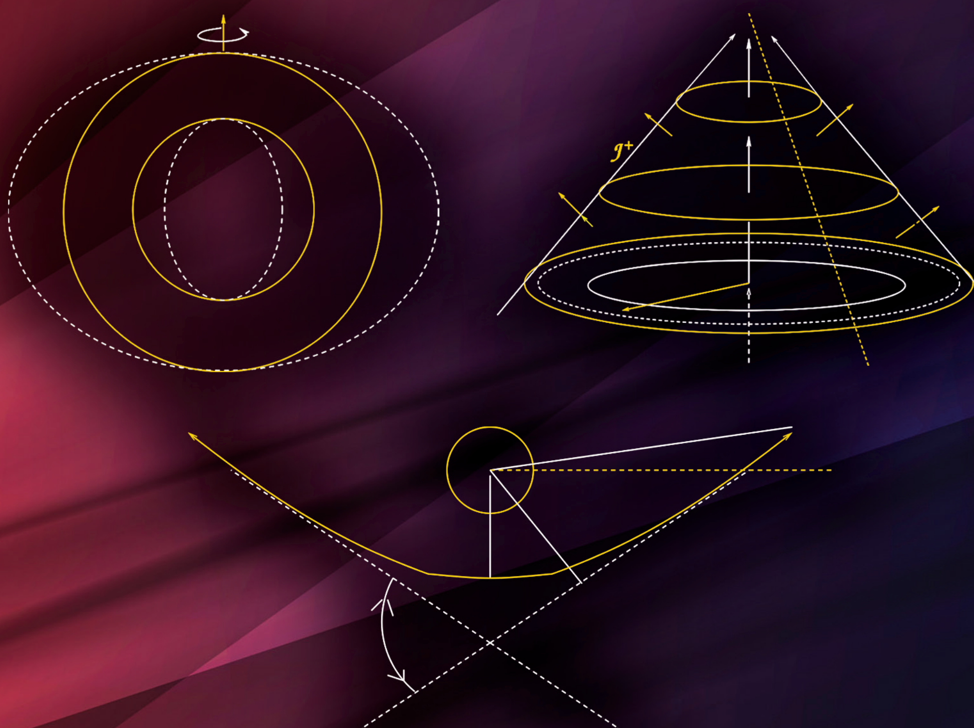
General Relativity

Basics and Beyond



Ghanashyam Date

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

General Relativity

Basics and Beyond

General Relativity

Basics and Beyond

Ghanashyam Date

The Institute of Mathematical Sciences

Chennai, India

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20141007

International Standard Book Number-13: 978-1-4665-5272-2 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

It fills me with a sense of joy and humility to present this book on the eve of the centenary year of the publication of Albert Einstein's General Theory of Relativity. When general relativity arrived, it had an aura of mystery due to its sophisticated view of space, time and gravitational phenomena. From the early phase when primary elaboration of the theory was mathematical in nature, it has evolved into a phase where it is being confronted by increasingly sophisticated experiments that have been successful so far. Students are often attracted to the theory and want to know what yet can be done with it. The book is envisaged as an attempt to familiarize students and prospective researchers with the basic features of the theory and offer a perspective on its more advance features.

There are many excellent textbooks from the classics by Misner–Thorne–Wheeler, Weinberg and Wald to the more recent ones by Sean Carroll, James Hartle and Thanu Padmanabhan, with differing styles and emphasis and there are excellent review articles on frontline topics. The idea here is to combine the 'textbook' and 'the review'. Thus, I have tried to adopt the pedagogical style of a textbook while avoiding an emphasis on detailed treatments, and at the same time, tried to present the essential ideas and just enough background material needed for students to appreciate the issues and current research.

There was also a conscious effort to emphasize the physical ideas and motivations, contrasting the mathematical idealizations which are important in appreciating the scope and limitations of the theory. Consequently, requisite mathematical background of differential geometry is summarized in the last chapter while the main text emphasizes the physical aspects.

The first five chapters usually form the core of an introductory course on General Relativity (GR) and constitute the "Basics" part of the book. The first chapter traces Einstein's arguments and informally motivates the mathematical model for space-time. In the second chapter, we first discuss the basic physical quantities related to space-time measurements and their relation to a metric in an arbitrary coordinate system. This is followed by examples of space-times corresponding to different types of gravitational fields. Some of these are revisited subsequently for further elaboration. Chapter 3 discusses adaptation of dynamics in a Riemannian geometry framework while the next chapter presents the Einstein equation together with its elementary properties. The fifth chapter discusses different phenomena either predicted by GR or influenced by GR. This also contains the classic tests of general relativity.

The "Beyond" part of the book, takes a look at some of the more sophisticated features of GR. Chapter 6 discusses the physical requirements of a well-defined deterministic framework for non-gravitational dynamics and the constraints it puts on the global structure of space-times. Surprisingly, the singular features seen in physically motivated examples turn out to have more general presence. The structure of the physically acceptable space-times is such that if certain conditions—such as complete gravitational collapse or an everywhere expanding universe—are realized in nature, then space-time will necessarily have regions where GR will cease to be applicable.

Not all physical situations are as grim. There are physical bodies of finite extent and it becomes necessary to look at the space-time geometry far away from them. This is especially relevant in the context of energy being carried away in the form of gravitational waves. Chapter 7 discusses the characterization of the appropriate asymptotic space-times.

In the next three chapters, we revisit black holes, gravitational waves and cosmological space-times. Apart from considering the general definition of black holes, we examine and discuss their quasi-local generalization in terms of the trapping, isolated and dynamical horizons. In the second look at gravitational waves, we trace the issues that were involved in settling the 'reality' of gravitational waves and briefly discuss the basic features of the challenge involved in their direct detection. The cosmological space-times are discussed primarily to get a glimpse of the possible nature of the space-like singularities.

Chapter 11 discusses the evolutionary interpretation for the class of globally hyperbolic space-times and reviews the initial value formulation. This forms a basis for numerical relativity presented in the next chapter. The Hamiltonian formulation paves a way for canonical quantization of gravity. While the book is focused on classical general relativity, introductory summaries of the main approaches to a quantum theory of gravity are included in Chapter 13. An alternative view of emergent gravity is also briefly mentioned.

There were many topics I wanted to include in this book, but could not. These are listed in the fourteenth chapter together with some concluding remarks. The Epilogue contains a summary of the requisite differential geometry and some of the results used in the main text.

There are many people to whom I owe a debt of gratitude. My understanding and appreciation of GR have been shaped by many influences over several years which are hard to demarcate. I must mention Naresh Dadhich and thank him for the numerous discussions and his generous encouragement. Within the context of this book, I would like to acknowledge critical feedback from my former teacher, Arvind Kumar on an earlier draft of Chapter 2 and my former student Alok Laddha for his comments on Chapter 7. I would also like to thank Thanu Padmanabhan for his help on the emergent gravity view and Sudipta Sarkar for a discussion on Jacobson's work. I must not forget the

students of my institute who had taken my courses on GR and those from places other than India who took short-term courses on various occasions under the SERC Schools in Theoretical High Energy Physics (India). The book has grown out of various lecture notes. I thank all of these students. I thank my friend and colleague, Gautam Menon, for help proofreading and for his helpful suggestions. There are times of meeting deadlines where responsibilities get shuffled and prioritized. This cannot be done without support from the family. I thank Nisha, Aditya, and my parents for it.

Ghanashyam Date

Part I

The Basics

Chapter 1

From Newton to Einstein: Synthesis of General Relativity

1.1 Space, Time, Observers

We all have an intuitive sense of what space is and what time is. Space is something in which 'bodies move' and time is something that sequences these movements. To make these notions quantitative we need to adopt a procedure to assign numbers to 'locations' and put time stamps on events. It is in terms of these assignments or *coordinates* that we make the space time explicit and it is this explicit model that is used in physics. All the tourist maps we use and the scheduling we struggle to achieve are based on precisely such 'made explicit' space and time. There is *no unique way to assign coordinates and time stamps*. Herein enters an *observer* (= adopted procedure).

With such a procedure at hand, it is possible to formulate the phenomenon of motion of bodies in terms of *kinematics* - description of motion and *dynamics* - laws of motion. The key point to note is that there is always an observer implicit directly in kinematics and indirectly in dynamics.

Einstein now observes several examples of relationships between classes of observers and the phenomena being described. Consider the problem of determining the distance between two points say by laying down meter sticks. The answer will evidently depend on how the meter sticks are laid. Drawing on the experience of measuring distances along short straight lines and using the procedure of assigning the *Cartesian coordinates* an observer can determine the distance between two points with Cartesian coordinates (x^1, y^1, z^1) and (x^2, y^2, z^2) to be given by

$$\text{Distance}^2 = (x^2 - x^1)^2 + (y^2 - y^1)^2 + (z^2 - z^1)^2 \ .$$

Now the interesting observation is that all observers assigning Cartesian coordinates will *verify* that the distance between two given points is numerically the same (assuming the same units are used!). Hence, as far as the problem of determining distance between points is concerned, *any* of this class of observers will do fine. *Mathematically*, the coordinates assigned by any two observers are related by the *transformation law*:

$$(x')^i = \sum_{j=1}^{3} A^i{}_j x^j + B^i, \quad \text{where } A^i{}_j \text{ is a 3-by-3 orthogonal matrix.}$$

These leave the Cartesian nature of coordinates unchanged as well as the expression for distance invariant. For $B^i = 0$, Einstein calls this *relativity of orientation.*

The next example he considers is the phenomenon of motion of particles, governed by Newton's laws formulated in the so-called inertial frames. The class of observers whose descriptions are equivalent are those who are in uniform relative motion, possibly differing in the orientation of the axes of the Cartesian frames and possibly with difference in the 'zero' of their clocks. This is of course Galilean relativity. What is left invariant is the *mass* $\times$ *acceleration.*

When phenomenon of motion is extended to include electromagnetic field and the motion of charges under their influence, a contradiction arises. Analysis of the famous moving magnet and conductor problem in the magnet's rest frame and the conductor's rest frame presents two alternatives. *Either* have *Galilean transformations* among the electric and magnetic fields so as to get the same force in both the frames *or,* allow a *new transformation law for the force* so as to be consistent with the Lorentz transformations which leave the Maxwell's equation invariant. Which one of these is 'correct'?

On the one hand, confirmation of constancy of speed of light puts Lorentz transformations on a firmer ground and on the other hand Galilean transformations contain an unwarranted assumption of observer independence of simultaneity. Einstein chooses Lorentz transformations and we have the *theory of special theory.* What two observers in uniform relative motion must agree on is the *same value of the speed of light in vacuum.*

This affects the kinematics in a profound manner. We will discuss the derivations a little later but let us note at this stage that length of a stick measured by a moving observer is a little *less* than that measured by an observer at rest with respect to the stick. Likewise when an observer compares the successive ticks of a moving clock with a stationary clock, the moving clock always ticks *slower.* These consequences of the demand of invariance of the speed of light go by the names *length contraction* and *time dilation* respectively.

The new kinematics does not leave invariant the other Newtonian law, namely the law of gravitational force. Once again we face a similar dilemma as before: Do we limit the applicability of the new kinematics or do we modify the law of gravitational force?

There is a peculiarity with the law of gravitation. The 'charge' that enters in the force law, the *gravitational mass*, happens to be numerically equal to the measure of the inertia of a body, its *inertial mass.* This makes different bodies of varied compositions, weights fall to the ground with the same acceleration.

There is no 'reason' for such conceptually widely different quantities to be numerically equal, except perhaps it is a clue to the *nature* of *gravitational interaction.*

All bodies fall at the same rate also means that an observer does so too and therefore, relative to the observer, the bodies continue to maintain their state of uniform motion. In the absence of any force of any other origin, this just means that the freely falling observer is the *Newtonian inertial observer*! The clue of equality of the two masses provides us with a *definition of inertial frames* as precisely those in which gravitational field *cannot* be detected. Furthermore, an observer who detects gravitational field, is *accelerated* relative to an inertial frame. Thus we can trade-off a gravitational field, for an observer accelerated relative to an inertial observer. Since relatively accelerated observers are involved, Lorentzian kinematics is not immediately applicable. Rotating platforms provide a convenient 'laboratory' for a thought experiment.

Imagine determining the circumference and the radius of a rotating platform. The measuring sticks tangential to the circumference will undergo Lorentz contraction while those along the radial direction will not be contracted. Thus the ratio of the circumference to radius of the *rotating* platform, obtained by taking the ratio of the number of measuring sticks along the circumference and the number along the radius, will be *greater than* 2π [1] while that of a non-rotating platform will be 2π. Hence, the *geometry* on a rotating platform will be *non-Euclidean.* But by equivalence principle, acceleration is equivalent to a gravitational field (locally) and therefore one must infer that gravity affects the geometry. This gravitational field is of course inferred by the observer who is co-rotating with the platform. We will return to the rotating platform later again.

Thus the *response* (motion) of bodies to a gravitational field is *independent* of their masses and the gravitational field also changes the geometry of space. Since a gravitational field is produced by masses, the spatial geometry is also influenced by the masses. Thus, *geometry of space is changeable.* This is quite a novel inference! Does *space-time* geometry also change with distribution of masses?

This could be so if clocks tick at different rates in a gravitational field. Consider an observer stationed at a height of h from the ground and another observer freely falling. The freely falling observer will have a speed $v = gt$ relative to the stationary observer after a time t and will have a fallen through a distance of $s = \frac{1}{2}gt^2$. As per Lorentzian kinematics, the rate of freely falling clock will be,

$$\Delta\tau_{\text{falling}} = \Delta\tau_{\text{fixed}}\sqrt{1 - g^2t^2} = \Delta\tau_{\text{fixed}}\sqrt{1 - 2gs} = \Delta\tau_{\text{fixed}}\sqrt{1 - 2\Delta\Phi_{\text{grav}}}$$

The final expression is depends only on the gravitational potential difference between the stationary observer and instantaneous position of the freely falling observer.

It is clear from this argument that the gravitational potential affects the rates of clocks and since gravitational potential changes with the distribution of masses, so does the clock rates and hence the *space-time geometry* too is affected by distribution of masses.

Thus, replacing gravitational field by an accelerated observer and the Lorentzian kinematics leads us to a space-time geometry which is affected by presence of gravitational field which in turn depends on distribution of masses. One puzzle still remains. If gravitational field can be 'gotten rid off' as in a freely falling lift, is gravity 'fictitious'? It can't be. After all Earth *is* freely falling in the gravitational field of the Sun and real tides - which are effects of Newtonian gravity - do exist! So, while metrical property within a freely falling lift will be that in the absence of gravitational field, something else must remain encoded in the geometry that will account for the tides.

From the examples of two-dimensional surfaces, we know that the non-Euclidean geometries have non-zero *curvature*. This is most easily seen on the surface of a sphere. Consider a triangle made up of sides which are portions of great circles on the sphere. If a triangle is 'large', with two points on the equator and the third one the north pole (say), then the sum of angles is *greater* than 180^0 degrees. Now bring the two equatorial points closer to the pole. Note that the generic latitude is *not* a great circle (the longitudes always are). So the small triangle will look more and more 'distorted', but the sum of its angles will get closer and closer to 180^0. In short, non-zero curvature is detectable as deviation from Euclidean geometry, only for larger triangles. The same is true for tidal forces in Newtonian gravity. The differential forces on two extremes of a body are larger when the separation of the two extremes is larger. Thus we see a parallel between the effects of curvature in geometry and the tidal forces of gravity.

At a qualitative level then, we see that effects of gravitational field can be mimicked by a space-time geometry which has curvature which in turn must depend on the distribution of masses since Newtonian gravitational potential does. The observed equality of gravitational mass and inertial mass, combined with Lorentzian kinematics leads to replacing gravitational interaction as revealing a space-time geometry which is curved in general and is changeable. Space-time is a dynamical entity. In the process, the principle of relativity also gets extended to *all* observers regardless of their state of motion. As Einstein says [1]: *"Theory of relativity is intimately connected with a theory of space and time ..."* In the subsequent chapters we will formalize and make these arguments precise and quantitative.

1.2 General Relativity and Space-Time Arenas

We will proceed somewhat informally and heuristically to arrive at the mathematical model for space-time. The precise details are given in chapter 14.

We have already alluded to the assignment of coordinates (and time stamps) as the a defining character of an observer. We are quite familiar with assignment of *Cartesian coordinates* on a plane: choose a point (origin) and a pair of orthogonal directions at that point (we use a protractor to determine orthogonality) call them the x-axis and the y-axis; go 'x'-units along the x-axis and then 'y'-units along the y-direction and assign the coordinates (x, y) to the point reached, 'P'. Repeat for other points. For the same choice of origin and the axes, we may reverse the order of traversal from origin to the same point i.e. first go 'y-'-units along the y-axis and then 'x'-units in the direction of the x-axis. From experience, we know that we will reach the same point and assign the same coordinates to it. A different observer may choose the same origin but a different pair of axes, can still reach the same point, 'P', but now with different values for its coordinates. Another observer may even choose a different origin. Nevertheless, each observer is able to follow this procedure for arbitrary values of (x, y) and thus label the points on a plane in an unambiguous and on-to-one manner. We even know how to relate the coordinates assigned by different observers, namely, $x' = O_{1,1}x + O_{1,2}y + C_1, \ y' = O_{2,1}x + O_{2,2}y + C_2$, where, the matrix O_{ij} is an orthogonal matrix, $O^T O = \mathbb{1}$. We can see readily that if we follow the same procedure on the surface of a sphere, then even for the same choice of an origin and the same pair of axes, the point reached depends on the order of traversal! Secondly, the relation between the coordinates assigned by two observers not a simple linear one as before. We also recognize this as a feature of the 'curved' nature of the sphere. We can attempt a similar exercise on the surface of a saddle and discover the same features. Clearly, the plane surface is rather an exception in the class of two-dimensional surfaces and therefore the ambiguities in the procedure for assigning coordinates is quite generic. we may have to be content with (i) *any arbitrary procedure of assigning coordinates* - but in a one-to-one manner and (ii) allow arbitrary (invertible) relations among different coordinates.

Of course labeling the points is only a first step an observer has to undertake. An observer has to observe and describe phenomena in terms of the reference system of coordinates chosen. How can different observers be sure that they are describing the same phenomena and compare notes to evolve a consensus on the laws of nature? Is it possible at all? Let us keep in mind the surface of the Earth as a concrete example. We know that temperatures at various locations have their specific values, irrespective of the labeling of the locations. Likewise, the wind patterns or ocean currents are described by a field of arrows, again independent of the labeling of the locations. Therefore there exist quantities on the sphere which have an existence independent of the labelling of the location. However, when we want to describe the variations of these quantities with the locations in quantitative terms, each observer can only do so using his or her reference system. Clearly, for the same quantity, we will have multiple descriptions in terms of multiple coordinates systems. Using the relations among the coordinates, we can transform one description into another one. Consistency requires that the quantities describing the phenomena must transform in a *specific manner* reflecting the fact that the quantities *exist independent of the coordinates.* For example, if a point P has two sets of coordinates (x, y) and (x', y') and the temperature in the vicinity is described by two functions $T(x, y)$ and $T'(x', y')$, then we must have, $T'(x', y') = T(x, y)$ at P. Similarly, if we have two descriptions of wind velocities as $(dx/dt, dy/dt)$ and $(dx'/dt, dy'/dt)$, then we must have the relations,

$$\frac{dx'}{dt} = \frac{\partial x'}{\partial x}\frac{dx}{dt} + \frac{\partial x'}{\partial y}\frac{dy}{dt}, \quad \frac{dy'}{dt} = \frac{\partial y'}{\partial x}\frac{dx}{dt} + \frac{\partial y'}{\partial y}\frac{dy}{dt}.$$

We have only used the chain rule of differentiation and the assumption that the relation among different coordinates is not only invertible but also differentiable. In a similar manner, we can see that the *gradients* of the temperature distribution must be related as,

$$\frac{\partial T'}{\partial x'} = \frac{\partial x}{\partial x'}\frac{\partial T}{\partial x} + \frac{\partial y}{\partial x'}\frac{\partial T}{\partial y}, \quad \frac{\partial T'}{\partial y'} = \frac{\partial x}{\partial y'}\frac{\partial T}{\partial x} + \frac{\partial y}{\partial y'}\frac{\partial T}{\partial y}$$

Here we have also used the fact that $T'(x', y') = T(x, y)$ in applying the chain rule. If we use a more compact notation of denoting the coordinates as x^i, the coordinate relations as $x^{'i}(x^j)$ then we can write the equations as,

$$T'(x') = T(x), \quad \frac{dx'^i}{dt} = \frac{\partial x'^i}{\partial x_j}\frac{dx^j}{dt}, \quad \frac{\partial T'}{\partial x'^i} = \frac{\partial x^j}{\partial x'^i}\frac{\partial T}{\partial x^j}$$

We have also introduced the Einstein summation convention, namely, repeated indices in an expression imply summation over the values of the indices. What we see is the beginning of *tensors*—sets of quantities that transform in a specific manner which imply that they represent entities that exist independent of assignments of coordinates. The temperature is a *scalar*, the velocities are *contravariant tensor of rank 1* and the gradients are *covariant tensor of rank 1.* Generalizations to multi-index quantities and details are given in the chapter 14.

This is similar to the case of special relativity's *4-tensor notation*, except that the implicit transformations are not the *Lorentz transformations* but the *general coordinate transformations.* This also means that the partial derivatives are evaluated at the same point where two sets of quantities are related and that these vary from point-to-point, the transformations being *non-linear* in general. Hence elementary algebraic operations such as addition, multiplications of tensors can only be defined pointwise.

One of the first casualty of allowing arbitrary assignment of coordinates is that the relation between coordinate differences and physically measured length is more remote. From the example of Cartesian coordinates on a plane, we know that the distance between two points, measured by using meter sticks (say) is related to the coordinate differences by the square root of the sum of their squares. If we were to use the polar coordinates, (r, θ), then the expression is, $(\Delta s)^2 = (\Delta r)^2 + r^2(\Delta\theta)^2 := \sum_{ij} g_{ij}(r,\theta)(\Delta x)^i(\Delta x)^j$. For points on a sphere, the coordinate differences have to be sufficiently small (ideally infinitesimal) to match with the length obtained by putting small measuring sticks along the surface of the sphere. The matching would be necessarily approximate as no finite length measuring stick can be confined to the curved surface. Even after restricting to small enough coordinate differences, we need to ensure that the measured length, Δs^2, is numerically the same if computed using differences from a different coordinate system. This can possibly be true, if the coefficients g_{ij} in the second coordinate system are different in just the right manner: $\sum_{ij} g'_{ij}\Delta x'^i \Delta x'^j = \sum_{ij} g_{ij}\Delta x^i \Delta x^j$. In the light of the discussion of tensors, this demand just means that (a) Δx^i transform as the contravariant rank 1 tensor and (b) g_{ij} transform as covariant rank two tensor. This will make the distance an invariant (coordinate independent) quantity. Since $\Delta x^i \approx (dx^i/dt)\Delta t$ and the velocity is a tensor while Δt is manifestly independent of the coordinates, the (a) above is satisfied. The requirement of Δx^i being sufficiently small comes about because for sufficiently small Δt, there is a unique 'straight' path along which we may lay the measuring sticks. The point to note is that we must have an quantity such as g_{ij} so that measured lengths can be computed using coordinate differences. This quantity is called a *metric tensor* while the expression $(\Delta s)^2$ is called the *line element.* There are infinitely many possible choices for a metric tensor.

An observer cannot be satisfied by just making observations at one point. We will want to set up differential equations, ordinary and partial, to make theoretical predictions. So we need to define derivatives of tensors which should also be tensors. Differentiation involves comparing values at neighboring points and tensors forbid such comparisons. Suppose we are given a tensor field, $A^i(q)$, for points q in the vicinity of a point p. If we consider the derivatives, $\frac{\partial A^i(x)}{\partial x^j}$, in two different coordinate systems, we see immediately that the derivatives do *not* transform as a tensor, due to an offending term containing double derivatives of the form $\frac{\partial^2 x'^i}{\partial x^j \partial x^k}$. For linear transformations such as Lorentz transformations, we don't encounter this, but for general coordinates, we cannot escape it. To construct a tensorial derivative, called a *covariant derivative*, we need to introduce a quantity $\Gamma^i{}_{jk}$ with appropriate transformations and define: $\nabla_j A^i := \frac{\partial A^i}{\partial x^j} + \Gamma^i{}_{jk}A^k$, with $\Gamma'^i{}_{jk}(x') := \frac{\partial x'^i}{\partial x^l}\frac{\partial x^m}{\partial x'^j}\frac{\partial x^n}{\partial x'^k}\Gamma^l{}_{mn}(x) + \frac{\partial x'^i}{\partial x^m}\frac{\partial^2 x^m}{\partial x'^j \partial x'^k}$. This quantity is called an *affine connection.* Notice that a *choice of* Γ is constrained only by the transformation rule and there are infinitely many choices possible. For every choice we can define covariant derivatives for all tensor fields (see chapter 14). Now, unlike the usual coordinate derivatives, the covariant derivatives *do not commute* i.e. $\nabla_i \nabla_j A^k - \nabla_j \nabla_i A^k := R^k{}_{lij}(\Gamma)\ A^l \neq \mathbf{0}$. The 4 index quantity $R^k{}_{lij}$ is manifestly a tensor (since the left-hand side is) and depends only on Γ and its first derivatives. This is the famous *Riemann Curvature Tensor*. We are thus naturally lead to a framework involving tensors, an arbitrarily chosen tensor - the metric tensor g_{ij} and an arbitrarily chosen affine connection - $\Gamma^i{}_{jk}$, with the associated Riemann curvature tensor. It turns out that the arbitrariness in the choice of the connection can be completely removed by demanding that $\Gamma^i{}_{jk} = \Gamma^i{}_{kj}$ and $\nabla_k g_{ij} = 0\ \forall i, j, k$. The connection so restricted is called the Riemann–Christoffel connection which is dependent on the metric tensor and the corresponding Riemann tensor is also determined by the metric. We now have a model for a space-time: It is a collection of 'events', made explicit by arbitrarily assigned coordinates, an arbitrarily chosen metric with a non-vanishing Riemann tensor in general. All determinable physical quantities of interest being tensors of appropriate ranks satisfying differential equations involving covariant derivatives. This model is nothing but a Riemannian manifold, defined more precisely in chapter 14.

To familiarize ourselves, we will discuss several examples of Riemannian manifolds in the next chapter.

Chapter 2

Examples of Space-Times

We will take a specification of a space-time as a set of coordinates x^μ with a non-singular metric $g_{\mu\nu}(x)$ with Lorentzian signature, given as *an infinitesimal invariant interval*, also known as *line element*, and study some of its properties[1]. Specifically, we consider,

Minkowski (No gravity)	$\Delta s^2 = -\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$
Rindler (Uniform)	$\Delta s^2 = -g_0^2 z^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2, \quad z > 0$
Rotating Disk (Centrifugal)	$\Delta s^2 = -f(\rho)\Delta t^2 + 2h(\rho)\Delta t \Delta\phi + g(\rho)\Delta\phi^2 + \Delta\rho^2 + \Delta z^2$ $f(\rho) := e^{-\omega^2\rho^2} - \rho^2\omega^2 e^{+\rho^2\omega^2}$, $h(\rho) := -\omega g(\rho)$, $\quad g(\rho) := \rho^2 e^{+\rho^2\omega^2}$
Schwarzschild (Spherical)	$\Delta s^2 = -\left(1 - \frac{2GM}{r}\right)\Delta t^2 + \left(1 - \frac{2GM}{r}\right)^{-1}\Delta r^2 + r^2\Delta\Omega^2$
FRW (Cosmological)	$\Delta s^2 = -\Delta t^2 + a^2(t)\left\{\frac{\Delta r^2}{1-\kappa r^2} + r^2\Delta\Omega^2\right\}$ where, $\Delta\Omega^2 := \left(\Delta\theta^2 + \sin^2\theta\Delta\phi^2\right)$
Plane wave (Undulating)	$\Delta s^2 = (\eta_{\mu\nu} + h_{\mu\nu})\Delta x^\mu \Delta x^\nu$, where, $h_{\mu\nu}(x) = \epsilon_{\mu\nu}(k)e^{ik\cdot x} + \bar{\epsilon}(k)_{\mu\nu}e^{-ik\cdot x}$ and $\eta_{\mu\nu} =$ diag (-1, 1, 1, 1) .

In order to appreciate interpretation of physical consequences of the space-time model, we will focus on familiar quantities such as *physical lengths, elapsed times measured by clocks, local speed ('speedometer reading'), local acceleration ('acceleration due to gravity')*. We take as given, a line element

[1] Our notation: *space-time* coordinates are indexed by Greek letters, $\mu, \nu, \ldots$ taking values $0, 1, 2, 3$; *space* coordinates are indexed by Roman letters, $i, j, \ldots$ taking values $1, 2, 3$. The invariant interval $\Delta s^2 = g_{\mu\nu}\Delta x^\mu \Delta x^\nu$. The metric signature is $- + ++$ and speed of light $c = 1$.

and its *physical interpretation.* We view the given coordinate system as a 'map' *on* the space-time with the metric coefficients as giving the rule to link coordinate intervals with physical quantities.

On a pseudo–Riemannian manifold, an infinitesimal *invariant interval* can be positive (*space-like*), negative (*time-like*) or null (*light-like*). Time-like intervals are given by *elapsed time on a physical clock* while space-like intervals are the *lengths measured by a physical measuring stick.* This could be deduced from the principle of equivalence applied to a freely falling lift [2], but we will simply take it as part of the interpretational scheme. We can classify smooth curves in the manifold into time-like, space-like and light-like according as the nature of infinitesimal intervals along the curves. Motion of small bodies ('point particles') in space is represented by time-like curves (or world lines) while propagation of light, in the geometrical optics approximation, is represented by light-like curves in the space-time manifold.

Elapsed Times: Any small physical clock is represented by a time-like curve. Define the clock's *coordinate velocity,* $V^i := \frac{\Delta x^i}{\Delta t}$. Consider two events on the clock's world-line defined by two consecutive 'ticks' of the clock. Let the coordinate intervals for these two events be $(\Delta t, \Delta x^i := V^i \Delta t)$. The corresponding *invariant interval* is given by,

$$\Delta\tau^2_{\vec{V}} := -\Delta s^2 = -g_{00}\Delta t^2 \left(1 + \frac{2g_{0i}V^i}{g_{00}} + \frac{g_{ij}V^iV^j}{g_{00}}\right) \tag{2.1}$$

By definition, the invariant interval is the elapsed time measured by this clock. Notice that for a clock 'at rest' ($V^i = 0$), $\Delta\tau^2 > 0$ implies that $g_{00} < 0$ and then for $V^i \neq 0$, the expression in parenthesis must be positive.

For a clock at rest, $V^i = 0$, we get $\Delta\tau_{\vec{0}} = \sqrt{-g_{00}}\Delta t$ and this provides the *interpretation of the coordinate interval: it is the elapsed time as measured by a clock at rest, divided by* $\sqrt{-g_{00}}$. It has a dependence on the *location* of the clock, through the metric coefficient. It follows,

$$\Delta\tau_{\vec{V}} = \Delta\tau_{\vec{0}} \left[1 + \frac{2g_{0i}V^i}{g_{00}} + \frac{g_{ij}V^iV^j}{g_{00}}\right]^{\frac{1}{2}} \tag{2.2}$$

For the Minkowski line element, $g_{00} = -1, g_{0i} = 0, g_{ij} = \delta_{ij}$, and we infer the *special relativistic time dilation* by noting that $\Delta\tau_{\vec{V}}$ is the time measured by the moving clock while $\Delta\tau_{\vec{0}}$ is the time measured by the stationary clock.

This appears to be 'opposite' to the usual special relativistic time dilation. It is not. The two events whose invariant interval is given by $\Delta\tau_{\vec{V}}$ are defined by the two consecutive ticks of the *moving clock.* This would usually be denoted by '$\Delta\tau_0$' ('proper time'). The *same* interval as measured by a clock at rest, would usually be denoted by Δt and we have denoted it by $\Delta\tau_{\vec{0}}$. Thus the time dilation derived above is the same one as obtained in special relativity when the metric is Minkowskian.

Next, consider two clocks A and B, both with coordinate velocities zero. Choose two events A and B on their respective world-lines such that the coordinates are (t, x^i_A) and (t, x^i_B) respectively. Consider two subsequent points

A' and B' on their world-lines with the *same* time coordinate $t' = t + \Delta t$. The spatial coordinates will remain the same since the coordinate velocities are zero. The elapsed physical times are then related as,

$$\Delta\tau_A = \sqrt{\frac{g_{00}|_A}{g_{00}|_B}}\ \Delta\tau_B \tag{2.3}$$

Notice that for Minkowski line element (and the FRW line element), $g_{00} = -1$ at both locations and hence the two elapsed times are the same. This ratio gives the *gravitational time dilation.* Taking the invariant time intervals to define inverses of frequencies, we get the prediction that frequencies undergo a change in a gravitational field. This was indeed first measured and verified by Pound and Rebka in 1959 [3–5]. The quantitative estimate is obtained using the Schwarzschild line element. The choice of pairs of events with the same coordinate interval, can be achieved in practice by clock A sending consecutive pulses to clock B. The coordinate time intervals at both clocks will be the same when the coordinate velocities are the same *and* the metric is assumed to be almost time independent over the flight time interval. This is of course realized in the near Earth space-time. We discuss the general case of frequency shifts in section 3.2.

Physical Lengths: Similar considerations apply to spatial invariant intervals ($\Delta s^2 > 0$) as well, in particular physical length intervals are also 'observer dependent'. To see this, recall an argument in the context of *special relativity.*

Imagine two events A and B defined by a car crossing two ends of a road. The coordinates assigned by a road observer will be $(0, \vec{0}), (\Delta t, L_{road}\vec{n})$. The coordinates assigned by the car observer will be $(0, \vec{0}), (\Delta t', 0)$. Let the speed of the car relative to road be β_{car} so that $L_{road} = \beta_{car}\Delta t$. Let the speed of the road relative to the car be β_{road} (in the opposite direction of course) so that $L_{car} = \beta_{road}\Delta t'$. Since the two events are the same, the invariant interval must be the same i.e.

$$\begin{aligned} -\Delta t^2 + L_{road}^2 &= -\Delta t'^2 \\ L_{road}^2(1 - \beta_{car}^{-2}) &= -L_{car}^2\beta_{road}^{-2} \\ \therefore L_{car} &= L_{road}\sqrt{1 - \beta_{car}^2}\left(\frac{\beta_{road}}{\beta_{car}}\right) \end{aligned} \tag{2.4}$$

The usual length contraction formula results when we assert that $\beta_{road} = \beta_{car}$ i.e. speed of road measured by car observer is the same as the speed of car measured by the road observer. Had we insisted on the lengths L_{car}, L_{road}, of the road as measured by the two observer are same, we would have got the two speeds to be different, with the β_{road} being not bounded by 1. Clearly, we should interpret the above equation (2.4) as implying that the physical lengths measured by two observers *can* be different while the two speeds are the same: $\beta_{road} = \beta_{car} < 1$. Notice that this identification makes the *velocity truly relative.*

The *necessity* of length contraction can also be seen in the explanation of the observation of muons at the ground level after traversing the atmosphere even though, naively, the rest-frame-life-time of 2.2 μsec would not be sufficient to travel through the thickness of the atmosphere. The ground observer explains this by invoking time dilation to stretch the half life while the muon-rest-frame observer gets the simplest explanation by invoking length contraction to squeeze the thickness of the atmosphere.

We have discussed length contraction using the Minkowski metric. Its generalization to general space-times is given below by a different argument using the definition of local speed.

Local Speed: Imagine a spaceship going from a location A to another one B. The duration of the journey can be measured by an on-board clock and we can ask for an average speed for the journey. How is this to be determined in terms of the arbitrary local coordinates (and the metric coefficients)? For the everyday experience of going in a car the speed shown by speedometer denotes the physical distance traversed in a time shown by a clock, either on board or on the ground. The natural definition of speed would thus be the ratio of a physical distance to a proper time. The problem is to identify, for a given time-like curve, the *spatial* distance covered in some physical time - we *need a definition* of splitting the space-time into space and time.

Recall that space-time coordinates are just labels and it is only in conjunction with metric coefficients that physical meanings are ascribed. Thus to properly identify a split as space and time, we have to specify a form of metric as well, apart from simply labelling $t := x^0$. This is achieved by taking a form for the metric as,

$$\Delta s^2 \ = \ - N^2\Delta t^2 + \bar{g}_{ij}(\Delta x^i + N^i\Delta t)(\Delta x^j + N^j\Delta t) \quad \text{where,} \tag{2.5}$$

$\bar{g}_{ij}$ is positive definite with inverse $\bar{g}^{ij}$. As matrices,

$$g_{\mu\nu} \ = \ \begin{pmatrix} -N^2 + \bar{g}_{ij}N^iN^j & \bar{g}_{ij}N^i \\ \bar{g}_{ij}N^j & \bar{g}_{ij} \end{pmatrix} \leftrightarrow$$

$$g^{\mu\nu} \ = \ \begin{pmatrix} -N^{-2} & N^{-2}N^j \\ N^{-2}N^i & \bar{g}^{ij} - N^{-2}N^iN^j \end{pmatrix} \tag{2.6}$$

Such a form can always be taken locally and serves to identify time-like directions.

And now, *for a given coordinate system*, we define 'space' to be the 't = constant' hypersurface. We had already deduced $g_{00} < 0$ just below eq. (2.1) and g^{00} is manifestly negative. The *unit* (time-like) normal to such a hypersurface is given by $n_\mu = \frac{1}{\sqrt{-g^{00}}}(1, 0, 0, 0)$ and the corresponding $n^\mu = g^{\mu 0}/\sqrt{-g^{00}}$. Define the associated projector, $P^\mu_{\ \nu} := \delta^\mu_{\ \nu} + n^\mu n_\nu$ which projects any vector to a *space-like vector*. Consider two points on the world line of the spaceship, with coordinate interval: $\Delta x^\mu := v^\mu \Delta\tau$, $v \cdot v := g_{\mu\nu}v^\mu v^\nu = -1$.

Let $\Delta x^\mu_{\parallel} := P^\mu_{\ \nu}\Delta x^\nu \ = \ \Delta\tau(v^\mu - \frac{g^{\mu 0}}{g^{00}}v^0)$. The *spatial physical interval* is then,

$$\begin{aligned}
\Delta\ell^2 \ &:= \ g_{\mu\nu}\Delta x^\mu_{\parallel}\Delta x^\nu_{\parallel} \ = \ -\Delta\tau^2\left(1 + \frac{(v^0)^2}{g^{00}}\right) \\
&= \ \Delta\tau^2(v^0)^2\left(g_{00} + 2g_{0i}\frac{v^i}{v^0} + g_{ij}\frac{v^i}{v^0}\frac{v^j}{v^0} - \frac{1}{g^{00}}\right) \qquad (2.7) \\
&= \ \Delta t^2\left(g_{00} + 2g_{0i}V^i + g_{ij}V^iV^j - \frac{1}{g^{00}}\right) \qquad V^i := \frac{v^i}{v^0} \ = \ \frac{\Delta x^i}{\Delta t}
\end{aligned}$$

Likewise, let $\Delta x^\mu_{\perp} \ := \ n^\mu n_\nu \Delta x^\nu \ = \ \frac{n^\mu \Delta t}{\sqrt{-g^{00}}}$ and the corresponding physical *time-like* interval is,

$$\Delta T^2 \ := \ -g_{\mu\nu}\Delta x^\mu_{\perp}\Delta x^\nu_{\perp} \ = \ -\frac{\Delta t^2}{g^{00}} \qquad (2.8)$$

$\Delta\ell^2 > 0$, together with $g^{00} < 0$, implies,

$$g^{00}(g_{00} + 2g_{0i}V^i + g_{ij}V^iV^j) \leq 1.$$

A *physical, local four velocity* can now be defined as $\beta^\mu := \Delta x^\mu_{\parallel}/\Delta T$ with the corresponding physical, local speed given by $\beta^2 = g_{\mu\nu}\beta^\mu\beta^\nu$. Explicitly,

$$\begin{aligned}
\beta^\mu \ &:= \ \sqrt{-g^{00}}\frac{\Delta\tau}{\Delta t}\left(v^\mu - \frac{g^{\mu 0}}{g^{00}}v^0\right) \ \Longrightarrow \ \beta^0 = 0 \ \text{ and} \qquad (2.9) \\
\beta^i \ &= \ V^i\sqrt{-g^{00}} + \frac{g^{0i}}{\sqrt{-g^{00}}} \ \longleftrightarrow \ V^i \ = \ \frac{\beta^i}{\sqrt{-g^{00}}} + \frac{g^{i0}}{g^{00}} \ , \qquad (2.10) \\
\beta^2 \ &:= \ \frac{\Delta\ell^2}{\Delta T^2} \ = \ g^{00}\left(\frac{1}{g^{00}} - g_{00} - 2g_{0i}V^i - g_{ij}V^iV^j\right) \\
&= \ 1 - \left(g^{00}g_{00}\right)\left(1 + \frac{2g_{0i}V^i}{g_{00}} + \frac{g_{ij}V^iV^j}{g_{00}}\right) \ \geq \ 0 \qquad (2.11) \\
\therefore \beta \ &= \ \sqrt{1 - g^{00}(g_{00} + 2g_{0i}V^i + g_{ij}V^iV^j)} \ < \ 1. \qquad (2.12)
\end{aligned}$$

That $\beta^2 < 1$ follows because, $v \cdot v = -1$ and $g^{00} < 0$, ensure that the second term under the square root is *positive*. Thus the physical speed is less than the speed of light.

We also have an explicit relation between the physical velocity β^i and the coordinate velocity V^i. For Minkowski line element in particular, both the speeds are equal. In fact for all the line elements listed above, except the rotating platform case, $g^{00}g_{00} = 1, g_{0i} = 0$ and $\beta^2 = -g_{ij}V^iV^j/g_{00}$.

Observe that the local speed vanishes with the coordinate speed only if $g^{0i} = 0$, which also implies $g^{00}g_{00} = 1$. Conversely, when the physical speed vanishes, the coordinate speed equals g^{i0}/g^{00}. These are special world-lines which are *orthogonal* to a local t = constant hypersurface i.e. $P^\mu_{\ \nu}V^\nu = 0$. Non-zero

g^{0i} (and g_{0i}) arises naturally in the context of the space-times near steadily rotating bodies or black holes. There these observers are known as *locally non-rotating observers*, (see section 2.3). We may call them generally *locally pinned* observers.

We have thus obtained a definition of a physical, *local speed* relative to an arbitrary coordinate system. By integrating the spatial invariant $\Delta\ell$ and the temporal invariant ΔT along the world line of the spaceship, we can also obtain the *average speed*. These are speeds determined by the observer associated with the local coordinates - roughly the 'on-road observer'. What about speed determined by the 'on board observer', e.g. the speedometer reading?

The elapsed time on the on-board clock is just the proper time $\Delta\tau$. The spatial distance travelled needs to be defined. *If* we were to use the same physical distance defined by the 'on-road' observer, $\Delta\ell^2$, we would get,

$$\begin{aligned}
\beta^2_{speedometer} &:= \frac{\Delta\ell^2}{\Delta\tau^2} = \frac{\left(\frac{1}{g^{00}} - g_{00} - 2g_{0i}V^i - g_{ij}V^iV^j\right)}{(g_{00} + 2g_{0i}V^i + g_{ij}V^iV^j)} \quad \Rightarrow \\
\beta^2_{speedometer} &= \frac{\beta^2}{1-\beta^2} \quad \longleftrightarrow \quad \beta^2 = \frac{\beta^2_{speedometer}}{1+\beta^2_{speedometer}} \qquad (2.13)
\end{aligned}$$

We see that the speedometer speed *can be* larger than the speed of light. This however is not a correct definition because it ignores the 'length contraction effect' - we used the *same* spatial distance in both the definitions. From the discussion of length contraction, we should demand that *the speedometer reading is the same as the speed β defined in (2.12)*. As a consequence, we obtain the *on-board odometer reading* or the physical distance travelled as measured by the on-board observer, namely,

$$\begin{aligned}
\Delta\ell_{odometer} &:= \beta\Delta\tau = \frac{\Delta\ell}{\Delta T}\Delta\tau \\
&= \Delta\ell\sqrt{g^{00}g_{00}\left(1 + 2\frac{g_{0i}V^i}{g_{00}} + \frac{g_{ij}V^iV^j}{g_{00}}\right)} \qquad (2.14) \\
&= \Delta\ell\sqrt{1 - g_{ij}\beta^i\beta^j} \quad \text{in terms of physical speed.}
\end{aligned}$$

In the last equation we used the relation between V^i and β^i and also used the metric form (2.6), with $g_{ij} = \bar{g}_{ij}$. This is the generalization of the length contraction formula for a general space-time.

Local Gravity: We would like to identify a quantity which can be interpreted as the *local acceleration due to gravity*, the analogue of the familiar $g = 981.0 cm/sec^2$ on Earth. A freely falling object has no weight, so we have to consider an object which is 'held in place' and find the 'force' needed to achieve this. Thus, the world-line of such an object should be a time-like curve which is not a geodesic (not freely falling). The idea of holding in place would be captured naively by requiring that the *spatial coordinates* along the world-line do not change. However, as observed in the discussion of local speed, coordinate speed can be different from the physical speed for a generic metric. It turns out to be more appropriate to use *vanishing local speed* as the characterization of 'holding-in-place'. Once again, we need to identify the spatial coordinates which we take to be the coordinates on the $t = constant$ surfaces. Such a world-line with zero local speed, is *normal* to the $t = constant$ surfaces i.e. $v^\mu = n^\mu$.

For any time-like curve, we have its *absolute acceleration* defined as: $a^\mu := v^\nu \nabla_\nu v^\mu$, $v^\mu := dx^\mu/d\tau$, $g_{\mu\nu}v^\mu v^\nu = -1$. Observe that $g_{\mu\nu}a^\mu v^\nu = 0$. Hence, a^μ is space-like and for $v^\mu = n^\mu$, $a^\mu = P^\mu_{\ \nu} a^\nu$ also holds. For such a curve we have, $v^\mu = g^{0\mu}/\sqrt{-g^{00}}$ and $a \cdot v = 0$ implies $a^0 = 0$. The spatial components of the acceleration are obtained as,

$$
\begin{aligned}
a^i &= \frac{g^{0\nu}}{\sqrt{-g^{00}}}\partial_\nu \frac{g^{0i}}{\sqrt{-g^{00}}} + \Gamma^i_{\mu\nu}\frac{g^{0\mu}}{\sqrt{-g^{00}}}\frac{g^{0\nu}}{\sqrt{-g^{00}}} \qquad (2.15)\\
&= \frac{1}{2}\left(g^{i\mu}g^{00} - g^{i0}g^{0\mu}\right)\partial_\mu \frac{1}{g^{00}} \quad = \quad \frac{1}{2}\left(g^{ij}g^{00} - g^{i0}g^{0j}\right)\partial_j \frac{1}{g^{00}}\\
&= \frac{1}{2}g^{00}\bar{g}^{ij}\partial_j \frac{1}{g^{00}} \quad = \quad -\frac{1}{2}\bar{g}^{ij}\partial_j \ln|-g_{00} + \bar{g}_{ij}N^iN^j|\\
&= -\frac{1}{2}\bar{g}^{ij}\partial_j \ln|-g_{00} + \bar{g}_{ij}V^iV^j| \qquad (2.16)
\end{aligned}
$$

In the first line, we have used the definition of the covariant derivative, in the second line the definition of the Γ and in the last equation we have used the inverse metric coefficients from eqn.(2.6) and also used the relation $N^i = -\ g^{0i}/g^{00} = -V^i$, the coordinate velocity of the 'locally pinned' world-line.

Notice that the acceleration of these special class of observers is determined entirely by the metric and its derivatives (in the given coordinate system). Specialising to the Schwarzschild metric (2.24), we see that $a^r = GM/(r^2)$. This is directed radially outward while force of gravity is radially inward. Therefore we *define the local acceleration due to gravity as the spatial components of minus the absolute acceleration of the normalized normal to the constant time foliation.* In equation,

$$
\begin{aligned}
g^i := -a^i &= -\frac{1}{2}\left(g^{ij}g^{00} - g^{i0}g^{0j}\right)\partial_j \frac{1}{g^{00}} \qquad (2.17)\\
&= +\frac{1}{2}\bar{g}^{ij}\partial_j \ln|-g_{00} + \bar{g}_{ij}V^iV^j|, \quad V^i := g^{0i}/g^{00}.
\end{aligned}
$$

Although we used the locally pinned worldlines to motivate the definition, the final expression for local gravity is in terms of the metric and its derivatives. Applying the definition to the Minkowski and the cosmological line elements, we see that local gravity vanishes in both cases.

The definition of local gravity is *not* a coordinate independent characterization and is *not* an intrinsic property of the space-time. Nevertheless, under purely spatial coordinate transformations, $t \to t, x^i \to \tilde{x}^i(x)$, which preserves the identification of space, the components g^i transform as a three-dimensional vector.

2.1 No Gravity (Minkowski Space-Time)

This is the space-time of special relativity i.e. in the absence of gravity. As a manifold, it is $\mathbb{R}^4$ and thus a single chart suffices. Let the coordinates be denoted as (t, x, y, z), each ranging over $\mathbb{R}$. The metric on it is then given by,

$$\Delta s^2 := -\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 \quad \leftrightarrow \quad \eta_{\mu\nu} := \text{diag}\ (-1, 1, 1, 1)\ .$$

As the metric is independent of the coordinates, the Riemann–Christoffel connection vanishes everywhere and the space-time is *Riemann flat.* We can choose another set of global coordinates x'^μ in terms of which the line element remains invariant, related to x^μ as, $x'^\mu = \Lambda^\mu{}_\nu x^\nu$, where the Λ's are independent of the coordinates and additionally satisfy: $\Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \eta^{\alpha\beta} = \eta^{\mu\nu}$. These are the *Lorentz transformations* forming the matrix group, $O(1, 3)$. We will refer to these coordinates in which metric is constant as defining *Lorentz frames.*

Since the Riemann–Christoffel connection vanishes in *all* Lorentz frames, the geodesics of Minkowski space-time are just straight lines: $x^\mu(\lambda) = a^\mu\lambda + b^\mu,\ \lambda \in \mathbb{R},$ and a^μ, b^μ are constants. They are time-like, space-like, light-like according as $a^\mu a^\nu \eta_{\mu\nu} =: a^2$ is negative, positive, zero respectively. Evidently, for any two points along a geodesic, the *invariant interval* is given by $a^2(\lambda_2 - \lambda_1)^2$. Since between any two pints there is a unique geodesic up to orientation, we can divide up the space-time relative to any point, P, into time-like region, space-like region and light-like region. The light-like region is called the *light cone* with the vertex of the cone at P. It is a cone because it is defined by $(t - t_P)^2 = (x - x_P)^2 + (y - y_P)^2 + (z - z_P)^2$. The points, Q, *inside or on* a light cone are said to be *causally connected* to P, since these can be connected by a time-like or a light-like curve which physically represents world-line of a material particle or a 'massless particle'. Observe that the light cones and the divisions into causal and space-like regions, is preserved by Lorentz transformations.

Can we arbitrarily restrict to a sub-region, say each coordinate in some interval, and take it as a space-time? Mathematically, it certainly defines a Riemann-flat space-time, still has geodesics which are straight lines and still has divisions into time-like, space-like and light-like regions relative to any point. However, now all geodesics will be 'incomplete' - their affine parameter λ will be limited to an interval. Furthermore, the region will *not* be preserved by Lorentz transformations. In other words, restriction to any bounded region is *observer-dependent.*

The fact that curves could have tangent vectors which are causal or space-like makes the approach to asymptotic regions more complex, indeed the nature of asymptotic region itself is more complex. This will be discussed later in chapter 7.

There is another natural coordinate system for Minkowski space-time, namely the usual spherical polar coordinates for the spatial metric:

$$\Delta s^2 := -\Delta t^2 + \Delta r^2 + r^2\Delta\theta^2 + r^2\sin^2\theta\Delta\phi^2$$

In terms of these of course, the Lorentz group does not act linearly.

We have already discussed time dilation, length contraction, physical speed in the general context. The definition of local gravity applied to the Minkowski line element, immediately gives the local gravity to be zero. We will see in the next section that with a different choice of coordinates, the local gravity will *not* be zero even though the Riemann tensor vanishes.

2.2 Uniform Gravity (Rindler Space-Time)

Freely falling lifts is an important laboratory in the thought experiments in Einstein's arguments. Near the surface of Earth, the (Newtonian) gravitational field is uniform and the freely falling lifts are seen as *uniformly accelerated* frames or observers. We ask the question: what is the appropriate metric which will describe the uniform gravitational field of Newton?

Let us assume that the uniform gravitational field is along z axis and is *time independent*. Furthermore, we may introduce coordinates x, y labelling the points of constant t, constant z plane which may be taken mutually orthogonal as well as orthogonal to the t, z directions. The metric then takes a diagonal form with metric coefficients being function of z only [6, 7], i.e.,

$$\Delta s^2 \ := \ -F(z)\Delta t^2 + G(z)\Delta z^2 + A(z)\Delta x^2 + B(z)\Delta y^2, \tag{2.18}$$

where, F, G, A, B are all *positive* for the implicit spatial and temporal nature of the coordinates. Observe that by redefining the coordinate $z \to z'$ as $dz' := \sqrt{G(z)}dz$, we can absorb away $G(z)$, but we will not do so at this stage for reasons that will be clear later. Since, a Newtonian uniform gravitational field does not produce any tidal forces and in general relativity absence of tidal forces implies vanishing Riemann tensor, we *require* the metric to have zero Riemann tensor. A straightforward calculation [6] then shows that by constant rescaling of coordinates, we can take

$$A(z) = B(z) = 1, \ \ \text{and} \ \ G(z) = \frac{(F'(z))^2}{4g_0^2F(z)}, \ \ F'(z) := \frac{dF(z)}{dz} \ ,$$

and $g_0 > 0$ (say) is a constant. In the above, it has been assumed that $F'(z), \ g_0 \neq 0$. For the special case of $F' = 0$, we go back to the line element, set $F = 1$ and then redefine z to set G $= 1$. This gives the Minkowski metric with no restrictions on the coordinates.

We have thus Riemann flat space-time modelling a Newtonian uniform gravitational field, parametrized by one free function $F(z) > 0$. Our definition of local gravity gives it to be directed along the negative z-axis with a magnitude, $g(z) = |-2g_0^2/F'(z)|$.

A coordinate transformation of the form $(x' = x, y' = y)$,

$$\left.\begin{array}{lcl} g_0 z' & := & \sqrt{F(z)}\cosh(g_0 t) \\ g_0 t' & := & \sqrt{F(z)}\sinh(g_0 t) \end{array}\right\} \quad \leftrightarrow \quad \left\{\begin{array}{lcl} F(z) & := & g_0^2(z'^2 - t'^2) \\ \tanh(g_0 t) & := & \frac{t'}{z'} \end{array}\right.$$

implies,

$$\Delta s^2 \ = \ -\Delta t'^2 + \Delta z'^2 + \Delta x'^2 + \Delta y'^2$$

The restriction $F(z) > 0$ implies *a region* of the Minkowski chart (t', z', x', y'). Thus the freedom contained in the function $F(z)$ corresponds to picking out different, proper subsets of the full Minkowski space-time of the previous section. The full Minkowski space-time is recovered only by the special case of $F' = 0$.

There are two natural choices for the $F(z)$ namely, (a) $F(z) = (1 + g_0 z)^2$ and (b) $F(z) = 1 + 2g_0 z$. The choice (a) gives $G(z) = 1$ corresponding to the possibility of redefining z to absorb $G(z)$ as mentioned before. For this choice, $z \neq -g_0^{-1}$ must hold and the local gravity is *not* uniform. It will even change sign across $z = -\ g_0^{-1}$ and we may restrict to $z > -g_0^{-1}$. For the choice (b), $G(z) = F^{-1}(z)$ and the local gravity is precisely $-g_0$. The z is now restricted to be larger than $-(2g_0)^{-1}$. We will stick to the choice (b).

Notice that an observer at rest relative to the uniform gravitational field (i.e. with respect to the (t, z, x, y) coordinates), will have a hyperbolic trajectory with respect to the (t', z', x', y') coordinates which are *inertial* since the metric is the Minkowskian metric.

A freely falling observer will be a geodesic of the line element (2.18) and depends on the choice of $F(z)$. To find the geodesics, we have to compute the Riemann–Christoffel connection defined in eq. (14.9), using the metric. For the special class of lifts which fall only along the z-axis, the relevant non-zero connection components are: ${\Gamma^t}_{tz} = F'/(2F)$, ${\Gamma^z}_{zz} = G'/(2G)$, ${\Gamma^z}_{tt} = -F'/(2G)$. The geodesics equations then become (over-dot denotes derivative with respect to proper time),

$$\ddot{t} + \frac{F'}{F}\dot{t}\dot{z} = 0, \quad \ddot{z} + \frac{G'}{2G}\dot{z}^2 - \frac{F'}{2G}\dot{t}^2 = 0. \tag{2.19}$$

The first equation gives $\dot{t}F$ =a constant which can be taken to be 1. Using this, we can convert the derivatives with respect to the proper time to those with respect to t. The solution of the second equation, $z(t)$ is then obtained from [6],

$$\frac{1}{\sqrt{F}} = \cosh(g_0(t - t_0)), \quad F(t_0) = 1$$

where we have made a convenient choice for integration constants.

Thus we know how the trajectory of an observer at rest in the un-primed coordinate system (say S), looks in the inertial coordinate system (say S' – it is hyperbolic), and we also know the trajectory of a freely falling observer in the un-primed coordinate system. The former is a *Rindler observer* which is uniformly accelerated relative to the inertial observer. The Rindler line element, given at the beginning of the chapter, corresponds to the choice, $F(z) = g_0^2 z^2$, giving $G(z) = 1$ and the restriction of excluding the $z = 0$ plane.

The equivalence of uniform gravitational field and uniformly accelerated frame, poses a puzzle for radiation from an electric charge. A charge at rest in a gravitational field, is accelerated relative to an observer who sees no gravitational field (i.e. a freely falling or inertial observer) and thus must radiate as per Maxwell theory valid in the inertial frame. But the observer in the gravitational field sees a stationary charge and is led to conclude that the charge should not radiate. What is the correct conclusion?

Briefly, the answer is that both statements are correct and there is no contradiction. It is always true that an observer will not detect any radiation from an electric charge if the charge is at rest relative to the observer and conversely, if a charge is accelerated relative to an observer, that observer will detect radiation. This is true for *both* cases where the observer is at rest relative to the gravitational field (i.e. charge at rest relative to a gravitational field) or it is freely falling in the gravitational field. These statements are checked by using the Maxwell theory in the inertial frame, S' and transforming the fields to the S frame. There is no contradiction here since radiation is not a general relativistically invariant notion. The details may be seen in [6, 7].

2.3 Centrifugal Gravity (Uniformly Rotating Platform)

Uniformly rotating platform has played an important role in Einstein's argument for linking accelerated observers, gravity and geometry. Recall that an inertial observer infers that the geometry on a rotating platform is non-Euclidean in that the ratio of circumference to diameter is greater than π. Equivalently, the geometry described by a co-rotating observer is non-Euclidean. With the general framework that we have developed, we ask: *what is the explicit metric description of the geometry to be used by the co-rotating observer?*

To arrive at such a metric, we consider two observers, both located at the center of a platform which is rotating with respect to one of them - *the inertial observer* and the other is co-rotating with the platform (i.e. 'spinning about the z-axis'). Both the observers use same coordinate labels (t, ϕ, ρ, z) but of course different metric coefficients. The platform has $z = 0$ which will play no role and hence will be suppressed. The inertial observer uses the line element, $\Delta s^2 = -\Delta T^2 + \rho^2 \Delta\phi^2 + \Delta\rho^2$ which is the Minkowski line element in cylindrical coordinates. Because of the circular symmetry of the system and the fact that apart from rotation, no other change is taking place in time, we assume the metric used by the co-rotating observer is time independent and angle independent but may depend on ρ. However, sense of rotation distinguishes motion of test particles, so the line element should *not* be *time symmetric* i.e. invariant under $t \leftrightarrow -t$.

The general form of metric can then be written as,

$$\Delta s^2 := -f(\rho)\Delta t^2 + 2h(\rho)\Delta t\Delta\phi + g(\rho)\Delta\phi^2 + \Delta\rho^2, \quad f(\rho),\ g(\rho) > 0. \quad (2.20)$$

Our task is to determine the functions f, g, h appropriately. We have set the possible $g_{\rho\rho} = 1$ invoking the freedom to choose the label ρ. This is also convenient because the physical radial length inferred by co-rotating observer is just ρ consistent with the inertial observer's expectation that radial direction suffers no length contraction. As $\rho \to 0$ we expect the rotation to have no effect and therefore require that as $\rho \to 0$, the functions $f \to 1$ and $g \to \rho^2$ while $h \to 0$. This also means that the t, ρ, ϕ used by the co-rotating observer, match with the inertial coordinates.

We require that the elapsed time and length contractions inferred by the inertial observer are precisely also inferred by the co-rotating observer using the metric above. We have to input the information that the metric above corresponds to a platform rotating with an *angular velocity*, ω, relative to the inertial observer. Lastly, we have to accommodate the possibility that, even for the inertial observer, the tangential speed $v(\rho)$ of a clock or a rod, may not be related to the angular speed as $v(\rho) = \rho\omega$, but could be a more general form so that the light speed limit is not violated.

This is the first case where we have a non-zero $g_{t\phi}$. Such space times admit a class of observers called *locally non-rotating observers*. These are defined by the property that their 4-velocity, $u^\mu(\rho)$ is orthogonal to the local, $t =$constant hypersurface. This means that for such an observer at some ρ, ϕ, its 4-velocity is given by $u^\mu = \alpha g^{\mu t}$, α determined by requiring $u^2 = -1$. Notice that $\Omega := \frac{d\phi}{dt} = \frac{d\phi}{d\tau}/\frac{dt}{d\tau} = u^\phi/u^t = g^{t\phi}/g^{tt} = -g_{t\phi}/g_{\phi\phi}$. The last equality follows from the form of the inverse metric (2.6).

These observers have an angular (coordinate) velocity Ω relative to the inertial observer at the origin. The physical speed as defined in (2.12) also turns out to be zero, implying that this observer is 'at rest relative to the local geometry'. It is thus appropriate to set the angular velocity, $\Omega(\rho) = \omega,\ \forall\ \rho$. This determines $h(\rho) = -\omega g(\rho)$.

Consider next the ratio of circumference to radius as inferred by the inertial observer and the co-rotating observer. By Einstein's argument, the inertial observer gets the ratio to be $\frac{2\pi\rho}{\sqrt{1-v^2(\rho)}}\rho^{-1}$. The co-rotating observer obtains the ratio as $2\pi\sqrt{g(\rho)}\rho^{-1}$. Equating the two ratios, determines $g(\rho) = \rho^2/(1 - v^2(\rho))$.

To determine $f(\rho)$ we use the time elapsed computation. According to the inertial observer a clock at ρ will have tangential speed $v(\rho)$ and will give

$\Delta\tau_\rho = \Delta\tau_{\rho=0}\sqrt{1-v^2(\rho)}$. The co-rotating observer on the other hand will use the eqn. (2.1) to get,

$$\Delta\tau(\rho) = \Delta t\sqrt{-g_{tt} - 2g_{t\phi}\omega - g_{\phi\phi}\omega^2} = \Delta\tau_{\rho=0}\sqrt{-g_{tt} + \omega^2 g_{\phi\phi}}$$

Equating $\Delta\tau_\rho = \Delta\tau(\rho)$ we get,

$$g_{tt} - \omega^2 g_{\phi\phi} = -1 + v^2(\rho) \implies f(\rho) = \frac{(1-v^2(\rho))^2 - \omega^2\rho^2}{(1-v^2)} \tag{2.21}$$

We have thus determined all three metric coefficients in terms of an unknown function $v^2(\rho)$. To deduce this function we use the local gravity expression for the locally non-rotating observers and equate it to the Newtonian *centrifugal* acceleration: $\rho\omega^2$. The absolute acceleration is computed easily.

$$\begin{aligned} a^\mu &:= u^\nu \nabla_\nu u^\mu = u \cdot \partial u^\mu + \Gamma^\mu{}_{\nu\lambda} u^\nu u^\lambda \\ \therefore\ g^\rho &= -\frac{1}{2}\frac{\partial_\rho(g_{tt} - \omega^2 g_{\phi\phi})}{g_{tt} - \omega^2 g_{\phi\phi}} = -\frac{1}{2}\frac{d}{d\rho}\ln(1 - v^2(\rho)) \end{aligned} \tag{2.22}$$

This is the only non-vanishing component of the acceleration. Equating it to the centrifugal acceleration, $g^\rho = \rho\omega^2$, determines the function $v^2(\rho) = 1 - e^{-\rho^2\omega^2}$. We have chosen the constant of integration so that $v(0) = 0$. Observe that for small $\rho\omega$ we recover the non-relativistic relation $v = \rho\omega$. For large $\rho\omega$ though, this relation is modified such that the speed is always less than the speed of light.

In summary, we have determined,

$$\begin{aligned} f(\rho) &= e^{-\omega^2\rho^2} - \omega^2\rho^2 e^{+\omega^2\rho^2}, \\ g(\rho) &= \rho^2 e^{+\rho^2\omega^2}, \\ h(\rho) &= -\rho(\rho\omega)e^{+\rho^2\omega^2}. \end{aligned} \tag{2.23}$$

This space-time may be used to model a centrifugal gravity attributable to a rotation relative to an inertial, non-rotating observer.

2.4 Spherical Gravity (The Schwarzschild Space-Time)

To get glimpses of the refined theory of gravity, let us take an example of the space-time corresponding to an idealized solar system in which the Sun is viewed as 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. This leads to the line element[2],

$$\begin{aligned} \Delta s^2 &= -f(r)\Delta t^2 + g(r)\Delta r^2 + r^2(\Delta\theta^2 + sin^2\theta\Delta\phi^2) \\ f(r) &= 1 - \frac{R_s}{r}, \quad g(r) \ = \ f(r)^{-1}. \end{aligned} \tag{2.24}$$

The particular choice, $f(r) = 1 - R_S/r$, $g(r) = f^{-1}(r)$ is known as the *Schwarzschild space-time.* This is not Riemann flat but its Ricci tensor is zero. The R_s is a parameter having dimensions of length and can be related to the mass, M of the spherical body as: $R_S := \frac{2GM}{c^2}$ and is known as the *Schwarzschild radius* of the body. This is completely determined by its mass.

On the two-dimensional surfaces defined by t = constant, r = constant, the line element gives to the *induced metric* which is the standard metric on a sphere. The area of such a sphere is given by,

$$Area \ = \ \int \sqrt{g_{ind}}d\theta d\phi \ = \ \int \sqrt{r^4 sin^2\theta}d\theta d\phi \ = \ 4\pi r^2 \tag{2.25}$$

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

On the three-dimensional space defined by t = constant, the induced metric is similar to the standard Euclidean metric expressed in the spherical polar coordinates. It would be exactly so, if $g(r) = 1$ and then r also gives the radius of the sphere. However $g(r)$ is yet to be determined, so we cannot interpret r as the usual radius.

There is another length scale, namely, the size of the spherical body. We denote it by $r = R$. The region $r \geq R$ defines the *exterior Schwarzschild space-time* and is the arena for the solar system tests of general relativity. For $r \gg R_S$, the line element takes the form,

$$\begin{aligned} \Delta s^2 &= -\left(1 - \frac{R_S}{r}\right)\Delta t^2 + \left(1 + \frac{R_S}{r} + \frac{R_S^2}{r^2} + \cdots\right)\Delta r^2 + r^2\Delta\Omega^2 \\ &= \left[-\Delta t^2 + \Delta r^2 + r^2\Delta\Omega^2\right] \ + \ \left[\frac{R_s}{r}\left(\Delta t^2 + \Delta r^2\right) + o\left(\left(\frac{R_S}{r}\right)^2\right)\right] \\ &= \text{Minkowski metric} \quad + \quad \text{deviations} \end{aligned} \tag{2.26}$$

The deviations are controlled by the small parameter R_S/r. This feature of

[2]We will see later, in section 5.1, that this form can always be chosen for spherically symmetric, static space-times.

the space-time, particularly the leading fall-off $o(R_s/r)$, illustrates the idea of *asymptotically flat space-time* which is discussed in more details in chapter 7.

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

$$R_S \approx \frac{2 \times (6.67 \times 10^{-8}) \times (2 \times 10^{33})}{(3 \times 10^{10})^2} \approx 3 \text{ km} \tag{2.27}$$

For contrast, the physical radius of the Sun is about 6,95,500 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.

Extending the domain of validity of the line element inward to smaller r, we notice that for $r = R_S$, $f = 0$ while g is infinite. The determinant of the metric however continues to be non-zero. Although some of the metric coefficients vanish/blow up, it turns out that the curvature components and hence the tidal forces, remain perfectly *finite* as $r \to R_S$. This particular 2-sphere turns out to be a 'coordinate singularity' or failure of the coordinates to describe the geometry. It marks the location of a *horizon* and is the first example of a black hole space-time. The region $r \leq R_S$ is discussed further in the section 5.4.

The exterior Schwarzschild space-time is the arena for the solar system tests of general relativity, discussed in section 5.1.

2.5 Cosmological Gravity (Robertson–Walker Space-Times)

The Schwarzschild space-time of the previous section was static and spherically symmetric and was a prototype for an isolated, highly idealized body. Another prototype is a *spatially homogeneous and isotropic space-time* which is an idealisation for our universe.

A *spatially homogeneous* space-time can be viewed as a stack of three dimensional spatial slices. Homogeneity means the geometry does not vary from point-to-point on any of the slices and *isotropy* at a point means that it 'looks the same' in every direction or equivalently has *no* distinguished direction. The absence of distinguished direction implies that the Riemann tensor must be a multiple of the combination of the metric consistent with its symmetries while homogeneity requires that the multiple must be constant over the slice. Thus, the spatial Riemann tensor must have the form: $R^i{}_{jkl} = \lambda(\delta^i{}_k g_{jl} - \delta^i_l g_{jk}) \ \Rightarrow \ R = 6\lambda$. Here λ can vary from slice-to-slice. Such three dimensional Riemannian spaces are completely classified and come in three varieties depending on the sign of the λ. Labelling each of the slices by a time coordinate, τ, and denoting the *normalized constant* curvature by $k = \pm 1, 0$, one can write the form of the metric for the universe as:

$$\begin{aligned} \Delta s^2 &= -\Delta\tau^2 + a^2(\tau) \left\{ \begin{array}{ll} \Delta\psi^2 + sin^2\psi\Delta\Omega^2 & \text{Spherical, } k=1 \\ \Delta\psi^2 + \psi^2\Delta\Omega^2 & \text{Euclidean, } k=0 \\ \Delta\psi^2 + sinh^2\psi\Delta\Omega^2 & \text{Hyperbolic, } k=-1 \end{array} \right\}, \\ d\Omega^2 &:= \Delta\theta^2 + sin^2\theta\Delta\phi^2 \qquad (2.28) \\ \Delta s^2 &= -\Delta\tau^2 + a^2(\tau)\left(\frac{\Delta r^2}{1-kr^2} + r^2\Delta\Omega^2\right) \text{ (Alternative form)} \qquad (2.29) \\ \Delta s^2 &:= -\Delta\tau^2 + a^2(\tau)\Delta s_3^2 \qquad (2.30) \end{aligned}$$

There is no terms of the form $g_{\tau i}\Delta\tau\Delta x^i$, because such $g_{\tau i}$ coefficients will distinguish a spatial direction which is disallowed by isotropy. The spatial metric with normalized constant curvature is called the *comoving spatial metric* while with the $a^2(\tau)$ factor included, it is the *physical spatial metric.*

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. Generically, these are *neither Riemann flat nor Ricci flat.* Most of modern cosmography – mapping of the cosmos – is based on these geometries. Dynamics of this idealized universe is encoded in the evolution of the scale factor which will be discussed in the section 5.2.

The points on any particular slice are supposed to represent galaxies (or clusters of galaxies) at a *cosmic time*, τ. The *physical distance* between two such galaxies is the scale factor $a(\tau)$ times the *comoving distance* - the distance computed from the comoving metric. As we change the slice, the physical distance changes just by the scale factor evolution and thus the ratio of physical distance at two times is the same as the ratio of the scale factors at those two times.

Consider now the physical distance, $R(\tau)$ between two galaxies at the same value of τ. Its change with τ can be obtained as:

$$v(\tau) := \frac{dR(\tau)}{d\tau} = \frac{R(\tau_0)}{a(\tau_0)}\frac{da(\tau)}{d\tau} = \frac{R(\tau)}{a(\tau)}\dot{a} =: H(\tau)R(\tau) \qquad (2.31)$$

Thus, the relative speeds of galaxies is proportional to their separation.

The scale factor although multiplying only the spatial comoving metric, also affects the frequency of light pulses received from distant galaxies. Imagine a galaxy at a point p on a slice at time τ_{em} emitting successive pulses of light with period T_{em} and a galaxy at point Q on a slice at a later time τ_{re} receiving them at intervals T_{re}. Since the comoving distance between the galaxies remains the same and invariant interval along a light path is zero, $\Delta\tau = a(\tau)\Delta s_3$, we get,

$$\int_{\tau_{em}}^{\tau_{re}} \frac{d\tau}{a(\tau)} = \int_{\tau_{em}+T_{em}}^{\tau_{re}+T_{re}} \frac{d\tau}{a(\tau)} \Rightarrow \frac{T_{em}}{a(\tau_{em})} = \frac{T_{re}}{a(\tau_{re})} \Rightarrow \qquad (2.32)$$

$$\frac{a(\tau_{re})}{a(\tau_{em})} = \frac{\nu_{em}}{\nu_{re}} := 1 + z \Rightarrow z = \frac{a(\tau_{re}) - a(\tau_{em})}{a(\tau_{re})} \approx \frac{\dot{a}(\tau_{em})}{a(\tau_{re})}(\tau_{re} - \tau_{em})$$
$$\therefore z = \left[\frac{\dot{a}(\tau_{em})}{a(\tau_{re})}\right] R(\tau_{em}) := H(\tau_{em})R(\tau_{em}) \tag{2.33}$$

Here we have replaced the periods T by ν^{-1} and have assumed that the difference of the τ labels is small. Speed of light being 1, this difference equals the *physical distance* between the galaxies.

Thus, in Robertson–Walker space-times, there *is* a frequency shift in the light exchanged between two galaxies thanks to the time dependent scale factor and the shift is proportional to the physical separation between them.

This matches naturally with the famous conclusion drawn by Hubble from observation of spectra from distant galaxies. He consistently observed a *red-shift* which was also proportional to their distance from us, now known as the *Hubble Law*. Putting the two equations (2.31, 2.33) together, he concluded that our universe must be expanding.

2.6 Undulating Gravity (Gravitational Waves)

Consider a small ripple of geometry on the background of the Minkowski space-time. By this we mean a metric of the form $g_{\mu\nu}(x) := \eta_{\mu\nu} + h_{\mu\nu}(x)$ where h is treated as a first order quantity i.e. raising/lowering of indices is done by the background metric and only the leading, non-vanishing terms are kept. The coordinates denote the standard Cartesian coordinates of the Minkowski space-time so that the background metric takes the form $\eta_{\mu\nu} := diag(-1, 1, 1, 1)$. As an example, we take,

$$h_{\mu\nu}(x) = \epsilon_{\mu\nu}(k)e^{ik\cdot x} + \bar{\epsilon}_{\mu\nu}e^{-ik\cdot x}, \tag{2.34}$$
$$k^2 := k^\mu k^\nu \eta_{\mu\nu} = 0, \quad \epsilon_{\mu\nu}\eta^{\mu\nu} = 0, \quad k^\mu \epsilon_{\mu\nu} = 0;$$

The conditions on the $\epsilon_{\mu\nu}$ and $k^2 = 0$ are equivalent to $h_{\mu\nu}(x)$ satisfying the equations: $\eta^{\mu\nu}\partial_\mu\partial_\nu h_{\alpha\beta} = 0$, $h_{\mu\nu}\eta^{\mu\nu} = 0$, and $\partial_\mu h^{\mu\nu} = 0$. The $h_{\mu\nu}(x)$ thus represents a plane wave propagating along a direction $\vec{k}$ with frequency $\omega := |k_0| = |\vec{k}|$ and the (complex) *polarization tensor*, $\epsilon_{\mu\nu}$ satisfying those conditions. The polarization tensor is said to be *transverse, traceless.*

It is easy to verify that these conditions do *not* determine the polarization tensor completely: two sets satisfying the conditions can differ as, $\epsilon'_{\mu\nu} - \epsilon_{\mu\nu} = ik_\mu\zeta_\nu + ik_\nu\zeta_\mu$, $k \cdot \zeta = 0$. The 10 components of the polarization tensor satisfy 5 conditions and have a further freedom worth 3 parameters (since $k \cdot \zeta = 0$). This leaves two independent components of polarization tensor.

Given the wave 4-vector k, we can define a set of 4, *complex*, linearly independent null (light-like) vectors: $k, \ell, m, \bar{m}$ with non-zero scalar products

$k \cdot \ell = -1, m \cdot \bar{m} = 1$. Using this basis of null vectors, we can write $\epsilon_{\mu\nu} := \Phi_{ab} e^a_\mu e^b_\nu$, $\{e^a\} = (k, \ell, m, \bar{m})$. Using the 5 conditions on the polarization tensor as well as the exploiting the 3 parameter freedom, all Φ_{ab} can be chosen to be zero except $\Phi_{mm}, \Phi_{\bar{m}\bar{m}}$. We will thus work with,

$$h_{\mu\nu}(x) = e^{ik\cdot x}(\Phi_{mm} m_\mu m_\nu + \Phi_{\bar{m}\bar{m}} \bar{m}_\mu \bar{m}_\nu)$$

Following properties are immediately verified: (a) these waves are transverse i.e. the polarization tensor is orthogonal to the wave vector ($k \cdot m = 0 = k \cdot \bar{m}$); (b) if we perform a rotation through an angle θ, in the plane transverse to $\vec{k}$, then the Φ_{mm} amplitude changes by $e^{-2i\theta}$ while the other amplitude changes by $e^{2i\theta}$. The waves are then said to have *helicities* ± 2. The Riemann–Christoffel connection and the Riemann tensor, to first order in h, are given by,

$$\Gamma^\lambda{}_{\mu\nu} = \frac{i}{2} e^{ik\cdot x} \left(k_\mu \epsilon^\lambda{}_\nu + k_\nu \epsilon^\lambda{}_\mu - k^\lambda \epsilon_{\mu\nu} \right) + \text{Complex Conjugate} \tag{2.35}$$

$$R^\alpha{}_{\lambda\mu\nu} = \frac{1}{2} e^{ik\cdot x} \left\{ k^\alpha (\epsilon_{\lambda\nu} k_\mu - \epsilon_{\lambda\mu} k_\nu) - k_\lambda (\epsilon^\alpha{}_\nu k_\mu - \epsilon^\alpha{}_\mu k_\nu) \right\} + \text{C.C.} \tag{2.36}$$

The see how such a wave may affect test particles, we can use the geodesic deviation equation. The relative acceleration between two neighboring time-like geodesics is given by (14.11),

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

where, u^λ is a reference geodesic while X^μ is a deviation vector ($u \cdot X = 0$) to a nearby geodesic. In the rest frame of the reference geodesic, $u^0 = 1/\sqrt{-g_{00}}, u^i = 0$, we get $u \cdot m = 0 = u \cdot \bar{m}$. Substituting for our polarization tensors and using $\omega := k \cdot u$, the frequency of the wave as measured by the freely falling observer u, we get,

$$a^\alpha = -\frac{1}{2}\omega^2 \left\{ \Phi_{mm}(m \cdot X) m^\alpha + \Phi_{\bar{m}\bar{m}}(\bar{m} \cdot X) \bar{m}^\alpha \right\} e^{ik\cdot x} + \text{C.C.}$$

Notice that the relative acceleration is in the plane transverse to the wave vector, $\vec{k}$. Furthermore, if a deviation vector is along $\vec{k}$, then there is no tidal acceleration. Taking a ring of test masses in the plane perpendicular to the direction of the wave, one can develop a detailed picture of how these masses respond to a passing gravitational wave [8].

This is our second example of a time varying geometry. It describes a *ripple of curvature*. It has a characteristic effect on test bodies implied by its helicity 2 nature.

Chapter 3

Dynamics in Space-Time

In this chapter, we will consider motions of *test objects* i.e. objects which will carry mass, energy etc. whose influence on the space-time geometry however can be neglected. We take as given, some space-time and study some generic properties of motion of point particles, small rotating objects and wave motion. These equations are generically obtained by appealing to *Principle of general covariance* and *Principle of equivalence.*

3.1 Particle Motion Including Spin

Having gotten a more general framework for a space-time, a natural question is: *how are the (classical) non-gravitational laws of physics to be adapted to this generalized framework?*. The laws of physics here, refer to the laws of point particle mechanics and laws of dynamics of fields. We already made such an adaptation while going from the Newtonian framework to the special relativistic one: the kinematical quantities associated with a particle e.g. its position, velocity, acceleration are to be described as 4-vectors, its intrinsic attributes are to be the rest-mass (for a massive particle) - a scalar, its intrinsic angular momentum - a space-like 4-vector and force to be a 4-vector. The Newton's laws of motion are to be expressed as a Lorentz-covariant equations. In going to the general relativistic adaptation, the Lorentz tensors are promoted to *general tensors* and so are the equations. The position of a particle ceases to be a vector. Wherever there are derivatives, these are to be replaced by *covariant derivatives.* This procedure is taken as a statement of *Principle of General Covariance.* Since the covariant derivatives reduce to ordinary derivatives in a locally inertial coordinate system, the procedure amounts to stipulating that *Laws of physics should take the special relativistic form in a local inertial coordinate system.* This formulation is sometimes referred to as the *Principle of Equivalence.* By contrast, the assumed equality of gravitational and inertial masses is referred to as the *weak principle of equivalence.* There are caveats to the procedure of replacing coordinate derivatives by covariant derivatives when higher order derivatives are involved. These stem from the fact that while coordinate derivatives commute, the covariant derivatives do not and different orderings differ by curvature terms via the Ricci identity. More on this below.

Free Motion: In special relativity (and in Newtonian theory too), all free particles follow straight line trajectories (world lines) regardless of their non-zero mass i.e. in a locally inertial coordinate system, its trajectory satisfies the equation: $\frac{d^2x^\mu}{d\tau^2} = 0$. This can be manipulated as,

$$\begin{aligned} 0 &= \frac{d^2x^\mu}{d\tau^2} = v^\nu \partial_\nu v^\mu & & (v^\mu := \frac{dx^\mu}{d\tau} \text{ is used}) \\ &= v^\nu \nabla_\nu v^\mu & & (\text{general covariance}) \\ &= v^\nu \partial_\nu v^\mu + v^\nu \Gamma^\mu{}_{\nu\lambda} X^\lambda & & (\text{definition of covariant derivative}) \\ &= \frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu{}_{\nu\lambda} \frac{dx^\nu}{d\tau}\frac{dx^\lambda}{d\tau} & & (\text{the geodesic equation!}). \end{aligned} \tag{3.1}$$

Thus a trajectory of a free particle is described by a geodesic.

In special relativity, the invariant interval on a world line of a massive particle is time-like while that on a massless particle (light) is light-like or null. This is equivalent to their *velocities* being time-like or null vectors. The same distinction holds for general space-time by general covariance. Since the norm of velocity, $g_{\mu\nu}v^\mu v^\nu$ is constant along a geodesic, the entire geodesic can be labelled as being time-like or null. Thus, both massive and massless, free particles follow time-like (respectively null) geodesics in any space-time.

What if a particle has an 'intrinsic spin'? This would be a model for a small gyroscope having spin angular momentum. In special relativity, such a spin will be described by a 4-vector which is space-like. This is because, in the rest frame of a particle (necessarily massive)[1], the spin is a 3-vector pointing in some direction and having some magnitude $\sqrt{\vec{s}\cdot\vec{s}}$. Its 4-norm is positive and hence space-like. The velocity in the same frame is a time-like vector with no spatial component i.e. $\eta_{\mu\nu}v^\mu S^\nu = 0$. In a torque-free, free motion, this 4-vector will be preserved along the time-like geodesic: $\frac{dS^\mu}{d\tau} = 0$, which generalizes into the equation,

$$\frac{dS^\mu}{d\tau} = 0 = v^\nu \partial_\nu S^\mu + \Gamma^\mu{}_{\nu\lambda} v^\nu S^\lambda, \quad g_{\mu\nu}v^\mu S^\nu = 0, \quad s^2 := g_{\mu\nu}S^\mu S^\nu. \tag{3.2}$$

For a non-free motion with a force being $\mathcal{F}^\mu$, the equations of motion for a single, non-spinning particle then take the form,

$$m_0 v^\nu \nabla_\nu v^\mu = \mathcal{F}^\mu, \quad g_{\mu\nu}\mathcal{F}^\mu v^\nu = 0\ ; \tag{3.3}$$

An example of such a force, for a charged particle, is the Lorentz force $\mathcal{F}^\mu := qF^{\mu\nu}v_\nu$. Another example would be $\mathcal{F}^\mu \sim A^{\mu\alpha\beta}v_\alpha v_\beta$ where the third rank tensor A, is symmetric in the last two indices and anti-symmetric in the first two indices.

[1] For a massless particle, there is no notion of a spin, but one has the notion of *helicity*.

If the body is spinning as well, then the force may act with/without generating a torque. In such a case of torque-free, accelerated motion of a small spinning body, we write $\frac{dS^\mu}{d\tau} = \xi v^\mu$ since in the rest frame, the spin should *not* change its direction. Preservation of $S \cdot v = 0$ then determines, $\xi = S \cdot \mathcal{F}/m^0$ and we get,

$$v^\nu \nabla_\nu S^\mu \quad = \quad (S_\alpha v \cdot \nabla v^\alpha)\, v^\mu \quad = \quad \left(\frac{S \cdot \mathcal{F}}{m_0}\right) v^\mu. \tag{3.4}$$

Thus under torque-free, *accelerated* motion, the spin vector satisfies an equation (the first equality) known as the *Fermi Transport Equation.* For geodesic motion, ($\mathcal{F}^\mu = 0$) it reduces to the parallel transport equation for the spin vector.

For a small spinning body or an idealized point spin, we may have only torque-free motion.

Even for the free fall motion, we should appreciate that the spin vector will 'precess' in general *even though it is non-precessing in the rest frame.* This precession - or change of direction of the spin - is defined relative to some fixed direction defined by a distant star or quasar. This can be computed by solving the parallel transport equation for S^μ [2] and is sensitive to the curvature[2] ('geodetic precession/De Sitter precession') as well as the spin of the rotating body ('frame dragging effect/Lense-Thirring effect') warping the space-time geometry. An experiment to detect these precessions in the near Earth geometry, thereby testing general relativity was proposed by Pugh and Schiff in 1959 [9] and was realized some 45 years later by the Gravity Probe B mission [10].

For an extended body though, a torque will in general be induced due to the differential forces on parts of the extended body and these can be obtained from the deviation equation (14.11). For instance, even though Earth's motion around the Sun may be well approximated as a free fall (geodesic), there is a torque induced on the Earth's spin by the tidal forces causing *precession of the equinoxes* [8]. For analysis of general motion of an extended body, please see [11–15].

3.2 Wave Motion

Electromagnetic waves, especially light, forms an important means of probing and learning about nature. In Minkowski space-time, their propagation is governed by the wave equation, $\eta^{\mu\nu}\partial_\mu\partial_\nu F_{\alpha\beta} = 0$ which follows from the Bianchi identity and vacuum equation. The generalization of source free

[2]Even in the absence of curvature i.e. in special relativity, the spin does precess relative to the distant stars and is know as the *Thomas precession.*

Maxwell equations to general space-time is obtained by replacing the coordinate derivatives by covariant derivatives:

$$\nabla_\mu F^{\mu\nu} \ = \ 0, \quad \nabla_\lambda F_{\mu\nu} + \nabla_\mu F_{\nu\lambda} + \nabla_\nu F_{\lambda\mu} \ = \ 0 \tag{3.5}$$

In the second equation, the Bianchi identity, the covariant derivative is redundant in our space-times with zero torsion. It can be solved identically by $F_{\mu\nu} := \nabla_\mu A_\nu - \nabla_\nu A_\mu$, with A_μ defined to within an addition of a term $\nabla_\mu \Lambda$ for an arbitrary scalar Λ. This is the usual gauge freedom of electromagnetism. Substitution in the first equation leads to,

$$\Box A_\mu - R_{\mu\alpha} A^\alpha - \nabla_\mu(\nabla \cdot A) \ = \ 0, \quad \Box := g^{\mu\nu}\nabla_\mu\nabla_\nu \tag{3.6}$$

and we have used the Ricci identity in getting the last two terms. Fixing the gauge by imposing $\nabla \cdot A = 0$, the equation reduces to an *inhomogeneous* wave equation with the Ricci tensor of the background space-time serving as a non-electromagnetic source.

We can also derive a wave equation directly for the gauge invariant $F_{\mu\nu}$ by operating on the Bianchi identity by ∇^μ and using the first equation together with the Ricci identity to get,

$$\Box F_{\mu\nu} - R_{\mu\alpha} F^\alpha_{\ \nu} + R_{\nu\alpha} F^\alpha_{\ \mu} + R_{\mu\nu\alpha\beta} F^{\alpha\beta} \ = \ 0. \tag{3.7}$$

For typical applications in observational astronomy, one uses the *geometrical optics approximation* which is developed assuming a form of solution whose amplitude varies very slowly compared to the variation of its phase. The scale of variation of the geometry e.g. inverse of square root of non-zero curvature components, is also assumed large compared to the scale of variation of the phase. Thus, if λ denotes the scale of variation of the phase and L denotes the smaller of the scales of variations of the geometry, the amplitude and polarization, then $\lambda \ll L$. The approximation is developed as a formal expansion in the parameter $\epsilon := \lambda/L$ assuming that the phase $\Phi(x)$ has no 'correction terms' while the amplitude has an expansion in power series in ϵ. Thus, we consider solution of the form,

$$F_{\mu\nu}(x) \ = \ \left\{\epsilon^0_{\mu\nu}(x) + \epsilon\ \epsilon^1_{\mu\nu}(x) + o(\epsilon^2)\right\}\sin(\epsilon^{-1}\Phi(x)) \ := \ \epsilon_{\mu\nu}\sin(\epsilon^{-1}\Phi) \tag{3.8}$$

It is more common to take the ansatz as a (complex amplitude)$\times$ $\exp(i\Phi)$ and then take real parts. We have taken a real form directly and choice of sine vs the cosine form does not matter. Substituting in the (3.5, 3.7) and denoting $k_\mu := \nabla_\mu \Phi$, we get,

$$0 \ = \ \epsilon^{-1}\cos(\epsilon^{-1}\Phi)(\sum_{(\lambda\mu\nu)} k_\lambda \epsilon_{\mu\nu}) + \sin(\epsilon^{-1}\Phi)(\sum_{(\lambda\mu\nu)} \nabla_\lambda \epsilon_{\mu\nu}) \tag{3.9}$$

$$0 \ = \ \epsilon^{-1}\cos(\epsilon^{-1}\Phi)(k^\mu \epsilon_{\mu\nu}) + \sin(\epsilon^{-1}\Phi)(\nabla^\mu \epsilon_{\mu\nu}) \tag{3.10}$$

$$0 \ = \ -\epsilon^{-2}\sin(\epsilon^{-1}\Phi)((k\cdot k)\epsilon_{\mu\nu} \tag{3.11}$$

$$+ \ \epsilon^{-1}\cos(\epsilon^{-1}\Phi)(2k\cdot\nabla\epsilon_{\mu\nu} + \epsilon_{\mu\nu}\nabla\cdot k\ ) \tag{3.12}$$

$$+ \sin(\epsilon^{-1}\Phi)(\ \Box\epsilon_{\mu\nu} - \ R_{\mu\alpha}\epsilon^\alpha_{\ \nu} + R_{\nu\alpha}\epsilon^\alpha_{\ \mu} + R_{\mu\nu\alpha\beta}\epsilon^{\alpha\beta}\ ) \tag{3.13}$$

Equating terms singular as $\epsilon \to 0$ and noting that the sine and cosine dependences have to vanish separately, we get the *defining equations* of the geometrical optics approximation:

$$\sum_{(\lambda\mu\nu)} k_\lambda \epsilon^0_{\mu\nu} = 0, \quad k^\mu \epsilon^0_{\mu\nu} = 0, \quad k \cdot k = 0 \tag{3.14}$$

$$2k \cdot \nabla \epsilon^0_{\mu\nu} + \epsilon^0_{\mu\nu} \nabla \cdot k = 0. \tag{3.15}$$

The first of these equations can be solved identically by taking $\epsilon^0_{\mu\nu} := k_\mu \epsilon_\nu - k_\nu \epsilon_\mu$ with ϵ_μ being defined to within $k_\mu \zeta$. The second equation then gives $k \cdot \epsilon = 0$ (transversality) and this is preserved under the ζk_μ addition. The transversality implies that ϵ_μ cannot be time-like and must be space-like modulo addition of ζk_μ. Evidently, the norm $\epsilon \cdot \epsilon =: a^2$ is preserved under the shift and is *positive* for a non-trivial solution. It is called the *scalar amplitude* [8]. Substituting $\epsilon^0_{\mu\nu}$ in the last equation leads to

$$\{(2k \cdot \nabla k_\mu)\epsilon_\nu + k_\mu(2k \cdot \nabla \epsilon_\nu + \epsilon_\nu \nabla \cdot k)\} - \{(\mu \leftrightarrow \nu)\} = 0 \tag{3.16}$$

The first term is zero because $k \cdot \nabla k_\mu = k^\nu \nabla_\nu \nabla_\mu \Phi = k^\nu \nabla_\mu \nabla_\nu \Phi = \frac{1}{2}\nabla_\mu (k \cdot k) = 0$. Hence k^μ *is tangent to a null geodesic.* Putting $\epsilon^\mu := a\mathcal{E}^\mu, \ \mathcal{E} \cdot \mathcal{E} = 1$ and substituting in (3.16) leads to

$$2k \cdot \nabla a + a \nabla \cdot k = 0, \quad k \cdot \nabla \mathcal{E}^\mu = 0. \tag{3.17}$$

In getting the second equation we observe that $k_\mu k \cdot \nabla \mathcal{E}_\nu - \mu \leftrightarrow \nu = 0$ implies that $k \cdot \nabla \mathcal{E}_\nu = \eta k_\nu$ and exploiting the freedom to change $\mathcal{E}_\mu \to \mathcal{E}_\mu + \frac{\zeta}{a} k_\mu$, we can arrange $\eta = 0$. The resultant $\mathcal{E}^\mu$ is called the *polarization vector.*

To summarize, the geometrical optics approximation applied to Maxwell equations and the wave equation imply that *the wave propagates along a null geodesic with its scalar amplitude satisfying the transport equation and its polarization vector parallelly transported along the geodesic.* This forms a basic ingredient in astronomical observations. One of the main applications is the computation of red-shifts.

Application to frequency shifts: Consider a source following a *time-like* trajectory, emits light at a point P which propagates along a null geodesic. It is received by an detector, following its own *time-like* trajectory, at a point Q. The frequencies at the emission and reception points are in general different and we would like to know the relation between them.

Let S^μ, D^μ and k^μ denote the 4-velocities of the source, detector and the light respectively. We have $S^2 = -1, D^2 = -1, k^2 = 0$. Furthermore the frequencies of the light, measured at P, Q are given by, $\omega_P := k \cdot S$ and $\omega_Q := k \cdot D$. The light vector satisfies $k \cdot \nabla k^\mu = 0$.

We have have already noted while discussing local speed, for time-like world-lines, that the local (physical) velocity β^i and the coordinate velocity V^i are related by $V^i = \beta^i/\sqrt{-g^{00}} - g^{0i}/(-g^{00})$. Defining $\gamma := 1/\sqrt{1-\beta^2}, \beta^2 := g_{ij}\beta^i\beta^j$, we can express a normalized, time-like 4-vector as: $v^\mu = \gamma\sqrt{-g^{00}}(1, V^i)$. In a similar manner, for a light-like world-line, we define $K^i := k^i/k^0$ and introduce $\hat{k}^i := \sqrt{-g^{00}}K^i + g^{0i}/\sqrt{-g^{00}} \leftrightarrow K^i = \hat{k}^i/\sqrt{-g^{00}} - g^{0i}/(-g^{00})$. It follows that $k \cdot k = 0 \Rightarrow \hat{k}^2 := \hat{k}^i\hat{k}^j g_{ij} = 1$. This allows us to write: $k^\mu = k^0(1, K^i)$. It is straight forward to obtain, $\omega := k \cdot v = -\frac{\gamma k^0}{\sqrt{-g^{00}}}(1 - \beta\cos\theta)$ where, $\beta := \sqrt{\beta^2}$ and $\cos\theta$ is defined through $g_{ij}\hat{k}^i\beta^j := \beta\hat{k}\cos\theta, \hat{k} := \sqrt{\hat{k}^2}$. With these, we now write,

$$\omega(P) \;:=\; k \cdot S \;=\; -\left.\frac{k^0\ \gamma_S(1 - \beta_S\cos\theta_{kS})}{\sqrt{-g^{00}}}\right|_P \tag{3.18}$$

$$\omega(Q) \;:=\; k \cdot D \;=\; -\left.\frac{k^0\ \gamma_D(1 - \beta_D\cos\theta_{kD})}{\sqrt{-g^{00}}}\right|_Q \tag{3.19}$$

$$\frac{\omega(Q)}{\omega(P)} = \left(\frac{k^0(Q)}{k^0(P)}\right)\left[\frac{\gamma_D(1 - \beta_D\cos\theta_{kD})(Q)}{\gamma_S(1 - \beta_S\cos\theta_{kS})(P)}\right]\left[\frac{\sqrt{-g^{00}(P)}}{\sqrt{-g^{00}(Q)}}\right] \tag{3.20}$$

The first factor is the ratio of the k^0 which are defined up to a constant scaling due to the affine parametrization of the null geodesic. This constant drops out in the ratio. It is the geodesic equation satisfied by the light ray that will determine this ratio. The second factor involves the direction of the light ray as well as the physical local speed of the source and the detector and corresponds in the special relativistic context, to the *Doppler shifts* due to motions of the source and the detector relative to their local coordinates. The last factor is the ratio of the metric coefficients and denotes the contribution of the *gravitational shift*.

We will consider three specific types of space-times and obtain the general frequency shifts. These are (i) static space-times, relevant for stellar scale red-shifts, (ii) cosmological space-times which are spatially homogeneous, isotropic and non-stationary, and (iii) stationary but non-static space-times, specifically the Kerr black hole. For these, we will obtain the first factor by using the geodesic equation: $k \cdot \partial k^0 + {\Gamma^0}_{\mu\nu}k^\mu k^\nu = 0$. Special case of Minkowski space-time will reproduce the special relativistic frequency shifts.

Static space-times: These have $\partial_0 g_{\mu\nu}$ and $g_{0i} = 0$. This immediately gives ${\Gamma^0}_{00} = 0 = {\Gamma^0}_{ij}$ and ${\Gamma^0}_{0i} = \frac{1}{2}\partial_i \ln|g_{00}|$. We have used $g^{00} = 1/g_{00}$ when $g_{0i} = 0$. Therefore,

$$\begin{aligned} 0 &= k \cdot \partial k^0 + k^0 k^i \partial_i \ln|g_{00}| \\ &= k \cdot \partial\left\{\ln k^0 + \ln|g_{00}|\right\}, \quad \because k^i\partial_i = k \cdot \partial - k^0\partial_0 \end{aligned} \tag{3.21}$$

$$\therefore \frac{k^0(Q)}{k^0(P)} = \frac{g_{00}(P)}{g_{00}(Q)} \quad \text{and}$$

$$\frac{\omega(Q)}{\omega(P)} = \left[\frac{\gamma_D(1 - \beta_D\cos\theta_{kD})(Q)}{\gamma_S(1 - \beta_S\cos\theta_{kS})(P)}\right]\left[\frac{\sqrt{-g_{00}(P)}}{\sqrt{-g_{00}(Q)}}\right] \tag{3.22}$$

Cosmological space-times: We choose the form of the metric as, $\Delta s^2 =$

$-\Delta t^2 + a^2(t)\bar{g}_{ij}\Delta x^i \Delta x^j$ where the 3-metric $\bar{g}$ is independent of t and is homogeneous. $a(t)$ is the scale factor and $g_{0i} = 0$. This leads to $\Gamma^0_{ij} = a\dot{a}\bar{g}_{ij} = \frac{\dot{a}}{a}(k^0)^2 g_{ij}\hat{k}^i\hat{k}^j \ = \ (k^0)^2\partial_0 \ln(a)$.

$$\begin{aligned}
0 &= k\cdot\partial k^0 + k^0 k^0 \partial_0 \ln a(t) \\
&= k\cdot\partial\left\{\ln k^0 + \ln a(t)\right\}, \quad \because k^0\partial_0 = k\cdot\partial - k^i\partial_i \qquad (3.23)\\
\therefore \frac{k^0(Q)}{k^0(P)} &= \frac{a(P)}{a(Q)} \text{ and} \\
\frac{\omega(Q)}{\omega(P)} &= \left[\frac{\gamma_D(1-\beta_D\cos\theta_{kD})(Q)}{\gamma_S(1-\beta_S\cos\theta_{kS})(P)}\right]\left[\frac{a(P)}{a(Q)}\right] \qquad (3.24)
\end{aligned}$$

In the cosmological context, the light originates (and is observed) in a cluster of galaxies and there is no sense in ascribing a local speed to the source or a detector. Consequently the Doppler factor is one and the second factor is a new type of frequency shift called the *cosmological frequency shift*. We have seen a different derivation of it in the section 2.5.

Rotating black hole: This case is little involved for a direct computation as above, because the geodesic equation for k^0 does not de-couple from the other components, notably the k^ϕ component. However, we can exploit the available *isometries* of the Kerr black hole, namely stationarity which is synonymous with existence of a time-like Killing vector field, ξ^μ, which provides a natural time coordinate and that of axisymmetry which implies existence of a space-like Killing vector field, ψ^μ, which provides the angular coordinate ϕ. It is a simple result that given a *Killing vector field* ξ^μ (i.e. satisfying $\nabla_\alpha\xi_\beta + \nabla_\beta\xi_\alpha = 0$) and a geodesic $k\cdot\nabla k^\mu = 0$, the quantity $k\cdot\xi$ is constant along the geodesic: $k\cdot\nabla(k\cdot\xi) = 0$. Thus, for the Kerr space-time, we have the constants, $\mathcal{E} := k^\mu\xi^\nu g_{\mu\nu}$ and $\ell := k\cdot\psi$. Using the space-time metric for the Kerr solution, given in equation (5.110) with $Q = 0$, we get,

$$\mathcal{E} \ := \ k^\mu g_{\mu t} = k^0 g_{tt} + k^\phi g_{t\phi}, \quad \ell \ := \ k^\mu g_{\mu\phi} = k^0 g_{t\phi} + k^\phi g_{\phi\phi}. \qquad (3.25)$$

We can evaluate both the equations at the source point P and the observation point Q to eliminate the two constants and obtain the $(k^0, k^\phi)(Q)$ in terms of $(k^0, k^\phi)(P)$. Simple algebra gives,

$$\begin{aligned}
\frac{k^0(Q)}{k^0(P)} &= \left\{\frac{g_{\phi\phi}(Q)g_{tt}(P) - g_{t\phi}(Q)g_{t\phi}(P)}{(g_{tt}g_{\phi\phi} - g^2_{t\phi})(Q)}\right\} \qquad (3.26)\\
&\quad + \left\{\frac{g_{\phi\phi}(Q)g_{t\phi}(P) - g_{t\phi}(Q)g_{\phi\phi}(P)}{(g_{tt}g_{\phi\phi} - g^2_{t\phi})(Q)}\right\}\Omega(P)
\end{aligned}$$

where, $\quad \Omega(P) := \dfrac{k^\phi(P)}{k^0(P)}$ is the local angular velocity of the light ray.

Substitution in the general red-shift equation (3.20), gives the desired expression.

We could have derived the previous two cases in the same manner as well. Specializing to staticity, $g_{t\phi} = 0$ immediately reproduces the previous expression.

For the cosmological example, we have only 'spatial homogeneity' and isotropy which has 6 space-like Killing vectors whose time components are all zero. Thanks to isotropy about every point, we can choose three of the Killing vectors to vanish at any given point leaving us with three non-trivial constants from the remaining Killing vectors e.g. $C_A := (k^i \xi^j_A \bar{g}_{ij})a^2$, $A = 1, 2, 3$. We have used the fact that the Killing vectors have no time component and $\bar{g}$ is the time independent comoving metric. Since the Killing vectors are determined by the comoving metric, they too are independent of time and hence we deduce that $k^i = a^{-2}(t)k^i(C_A)$. Now $k^2 = 0$ gives the relation derived earlier. Another derivation using a 'Killing tensor' may be seen in [16].

Chapter 4

Dynamics of Space-Time

In the last chapter we saw how equations determining motion of matter are adapted to any given space-time. We do not know yet how to determine the space-time consistent with a distribution of matter. This link is provided by the Einstein equation whose heuristic derivation and elementary properties are discussed below.

4.1 Einstein Equation

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,

$$\frac{d^2x^i}{dt^2} = -\frac{\partial}{\partial x^i}\Phi \tag{4.1}$$

$$\nabla^2\Phi = 4\pi G\rho \tag{4.2}$$

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'. 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,

$$\text{Non-relativistic test particle} \quad \Rightarrow \quad \left|\frac{dx^i}{d\lambda}\right| << \left|\frac{dt}{d\lambda}\right| \quad \Rightarrow$$

$$\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} \approx \Gamma^\mu_{00}\left(\frac{dt}{d\lambda}\right)^2$$

$$
\begin{aligned}
\text{time independence of geometry} &\Rightarrow \\
\Gamma^{\mu}_{00} &= -\frac{1}{2} g^{\mu\rho}\partial_\rho g_{00} \\
\text{Close to Minkowskian geometry} &\Rightarrow \quad g^{\mu\nu} \approx \eta^{\mu\nu} - h^{\mu\nu} \quad \Rightarrow \\
\Gamma^{\mu}_{00} \approx -\frac{1}{2}\eta^{\mu\rho}\partial_\rho h_{00} &
\end{aligned} \tag{4.3}
$$

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

$$
\frac{d^2x^i}{dt^2} = \frac{1}{2}\eta^{ij}\partial_j h_{00} = +\delta^{ij}\partial_j\left(\frac{1}{2}h_{00}\right) \tag{4.4}
$$

Comparing with the Newtonian equation (4.1), we see that the metric component g_{00} gets identified with $-1 - 2\Phi/c^2$. Note that the Newtonian equations have usual units while in the metric we have used $c = 1$ units which is the reason for the c^2 factor in the denominator. Thus we obtain a relation between metric and the Newtonian potential. Newton's theory determines the potential given a mass density ρ via the Poisson equation. Using special relativity, the energy density is ρc^2, 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,

$$
\nabla^2 g_{00} = -\frac{8\pi G}{c^4} T_{00} \tag{4.5}
$$

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,

$$
\mathcal{F}_{\mu\nu}(g) = -\frac{8\pi G}{c^4} T_{\mu\nu} \tag{4.6}
$$

where, $\mathcal{F}_{\mu\nu}$ is a tensor constructed from the metric and should satisfy the following properties [2]:

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 (4.2).

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. Conveniently, the Riemann tensor already satisfies the differential Bianchi identity:

$$\begin{aligned} R^{\rho}_{\sigma\mu\nu;\lambda} + R^{\rho}_{\sigma\nu\lambda;\mu} + R^{\rho}_{\sigma\lambda\mu;\nu} &= 0 \quad \Rightarrow \\ R_{\mu}{}^{\nu}{}_{;\nu} &= \frac{1}{2}R_{;\mu} \end{aligned} \tag{4.7}$$

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,

$$a(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}) + \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^4}T_{\mu\nu} \tag{4.8}$$

We have not yet used 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 $aR_{00} \approx -\frac{4\pi G}{c^4}T_{00}$. 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:

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{4.9}$$

A number of remarks are in order.

4.2 Elementary Properties and Peculiarities

(1) The coefficient in front of $T_{\mu\nu}$ is about $2 \times 10^{-48} cm^{-1}.gm^{-1}.sec^2$. 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) The derivation above follows Weinberg [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 neighboring 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 $R_{\mu\nu} = \frac{4\pi G}{c^4} T_{\mu\nu}$. 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 [17]. Weinberg [2] 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 (see section 14.6). 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.

(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 higher rank 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), see section (14.6). When a metric is available, one has a natural invariant volume element available and one can also 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:

$$\begin{aligned}
\sqrt{g}\nabla_\mu J^\mu &= \sqrt{g}\partial_\mu J^\mu + \sqrt{g}\Gamma^\mu_{\mu\nu} J^\nu \\
&= \sqrt{g}\partial_\mu J^\mu + \sqrt{g}(\partial_\nu \ell n \sqrt{g}) J^\nu \\
&= \partial_\mu(\sqrt{g} J^\mu) \\
&= \partial_\mu(\sqrt{g}\epsilon^{\mu\nu_1\cdots\nu_{n-1}}\omega_{\nu_1\cdots\nu_{n-1}}) \\
&= \mathcal{E}^{\nu_1\cdots\nu_n}\partial_{\nu_1}\omega_{\nu_2\cdots\nu_n} \\
&= d\omega
\end{aligned} \tag{4.10}$$

Here, $\mathcal{E}$ is the Levi–Civita symbol (section (14.6).

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 leaves the metric invariant, then one can define a conserved quantity. 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 $Q(\Sigma) := \int_\Sigma J^\mu \xi_\mu$ where Σ is a hypersurface orthogonal to the Killing vector, is conserved. However, generic space-times do not admit any Killing vectors.

4.3 The Stress Tensor and Fluids

Let us consider the right-hand side of the Einstein equation – the Stress Tensor, $T_{\mu\nu}$. Only two properties are stipulated, namely, it is a *symmetric tensor* and it is covariantly conserved. One more property is implicit: if variables representing matter are 'set to zero', the stress tensor vanishes. Hence,

$T_{\mu\nu}$ is more accurately called as the *matter stress tensor*. Coupling of matter to the metric is *only* through its 'energy' and its 'strength' is determined by appealing to the Newtonian limit and has the specific $8\pi G$ coefficient. How is the $T_{\mu\nu}$ to be determined for a given type of matter?

It is postulated ab initio as in the case of a perfect fluid and is derived for matter which can be described by an action containing the metric. The precise coefficients of coupling are a matter of convention in the sense, it depends on the coefficients in front of the Einstein–Hilbert action. These coefficients are chosen so that the equations of motion derived give the Einstein equation. Let us note some examples.

Perfect fluid: A perfect fluid is characterized by a fluid *velocity*, $v^\mu(x)$ (time-like), and two functions, $P(x)$ and $\rho(x)$, representing *pressure and energy density* respectively. Its stress tensor is given by,

$$T^{\mu\nu} \quad := \quad \rho v^\mu v^\nu + P(g^{\mu\nu} + v^\mu v^\nu) \tag{4.11}$$

This is the same form as in special relativistic formulation.

For matter which can be described by an action, It is most naturally derived from an action functional including matter denoted generically by Φ, coupled to a metric, via the definition: $T_{\mu\nu}(\Phi, g) := -\frac{2}{\sqrt{|g|}}\frac{\delta S_{\mathrm{matter}}(\Phi, g_{\mu\nu})}{\delta g^{\mu\nu}}$. The specific sign and the factor of 2 are chosen so that for the standard normalization for actions on Minkowski space-time, the definition matches with the special relativistic definitions[1]. Thus, in using this definition, we assume standard normalization of the flat space-time actions[2].

Here are some examples.

Free scalar field: The action is taken as

$$\begin{aligned} S[g,\phi] \quad &= \quad \int d^4x\sqrt{|g|}\mathcal{L} := \int d^4x\sqrt{|g|}\left(-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}m\phi^2\right) \quad \Rightarrow \\ T_{\mu\nu} \quad &:= \quad -\frac{2}{\sqrt{|g}}\frac{\delta S}{\delta g^{\mu\nu}} \quad = \quad \partial_\mu\phi\partial_\nu\phi + g_{\mu\nu}\mathcal{L} \end{aligned} \tag{4.12}$$

Electrodynamics:

$$\begin{aligned} F_{\mu\nu}(A) \quad &:= \quad \partial_\mu A_\nu - \partial_\nu A_\mu, \\ S[g,A] \quad &= \quad \int d^4x\sqrt{|g|}\mathcal{L} \ := \ \int d^4x\sqrt{|g|}\left(-\frac{1}{4}g^{\mu\alpha}g^{\nu\beta}F_{\mu\nu}F_{\alpha\beta}\right) \quad \Rightarrow \\ T_{\mu\nu} \quad &:= \quad -\frac{2}{\sqrt{|g}}\frac{\delta S}{\delta g^{\mu\nu}} \quad = \quad F_{\mu\alpha}F_{\nu\beta}g^{\alpha\beta} + g_{\mu\nu}\mathcal{L} \ . \end{aligned} \tag{4.13}$$

Let u^μ, a time-like unit vector, represent an observer. Let $E_I{}^\mu$, $I = 0, 1, 2, 3,$

[1] Recall, our metric signature is $(-, +, +, +)$.

[2] The action for metric and matter together will be written as: $\alpha_g S_{\mathrm{Gravity}} + \alpha_m S_{\mathrm{matter}}(\Phi, g)$, the constant coefficients are chosen so that Einstein equation results. The gravitational action is taken to be: $\int d^4x\sqrt{|g|}R(g)$. This implies, $\alpha_m = 16\pi G\alpha_g$.

$E_0^\mu := u^\mu$ denote an orthonormal basis, so that $g^{\mu\nu} = \eta^{IJ} E_I{}^\mu E_J{}^\nu$. Then $\mathcal{E} := T_{\mu\nu}u^\mu u^\nu (=: T_{00})$ is the *energy density* measured by the observer in his/her rest frame. For the three examples above, we get,

$$\begin{aligned}
\mathcal{E}_{\text{scalar}} &= (E_0 \cdot \nabla\phi)^2 + \frac{1}{2}\left\{-(E_0 \cdot \nabla\phi)^2 + (E_i \cdot \nabla\phi)^2 + m^2\phi^2\right\} \\
&= \frac{1}{2}\left\{(E_0 \cdot \nabla\phi)^2 + (E_i \cdot \nabla\phi)^2 + m^2\phi^2\right\} && (4.14) \\
\mathcal{E}_{\text{em}} &= F_{0I}F_{0J}\eta^{IJ} + \frac{1}{4}F_{IK}F_{JL}\eta^{IJ}\eta^{KL}, \quad F_{IJ} := E_I{}^\alpha F_{\alpha\beta} E_J{}^\beta, \\
&= \frac{1}{2}\left(\vec{E}\cdot\vec{E} + \vec{B}\cdot\vec{B}\right), \quad E_i := F_{0i}, \quad B_i := \frac{1}{2}\mathcal{E}_i{}^{jk}F_{jk} && (4.15) \\
\mathcal{E}_{\text{fluid}} &= \rho && (4.16)
\end{aligned}$$

The example of electrodynamics also shows that $T_{0i} := T_{\mu\nu}u^\mu E_i{}^\nu = \mathcal{E}_i{}^{jk}E_jB_k$ which is the Poynting vector representing the flux of electromagnetic field energy.

Not every stress tensor which is symmetric and covariantly conserved need represent *physical matter*. For instance, we expect all physical matter to have positive energy density as measured by *any* observer in his/her rest frame. In our experience with Newtonian gravity, it is an attractive force only among two positive masses and we have never seen it being repulsive. This suggests that in the physical world, masses are only positive.

Since the Ricci tensor is determined by the matter stress tensor through the Einstein equation, the properties of stress tensor have a direct bearing on the geometrical properties manifested through the Ricci tensor. Even though we can have very complicated physical stress tensor arising out of different species of matter, we expect certain general properties to be satisfied by any such tensor. These stipulations go under the name of *energy conditions*. Below are commonly used energy conditions [17].

Weak	$T_{\mu\nu}v^\mu v^\nu \geq 0$	$\forall\ v\cdot v < 0$
Strong	$T_{\mu\nu}v^\mu v^\nu \geq\ -\frac{1}{2}T_{\alpha\beta}g^{\alpha\beta}$	$\forall\ v\cdot v = -1$
Dominant	$-T^\mu_{\ \nu}v^\nu$ is future directed causal	$\forall\ v$ future directed causal.

There is also the *Null energy condition* which replaces the time-like v^μ by a light-like k^μ.

To appreciate their implications, it is useful to have an *algebraic classification* of the stress tensor, which allows us to put the stress tensor in some canonical forms. The classification holds point-wise in the space-time. Mathematically, one considers the eigenvalue equation: $T^\mu_{\ \nu}X^\nu = \lambda X^\mu$ or equivalently, $T_{\mu\nu}X^\nu = \lambda g_{\mu\nu}X^\nu$. Note that $T^\mu_{\ \nu} = g^{\mu\alpha}T_{\alpha\nu}$ is *not* a symmetric matrix, only $T_{\mu\nu} = T_{\nu\mu}$. Hence, diagonalizability of $T^\mu_{\ \nu}$ is not assured. Secondly, the metric is Lorentzian which means that eigenvectors X^μ can be time-like and light-like as well. Nevertheless, symmetric nature of the stress tensor implies that *eigenvectors corresponding to distinct eigenvalues are necessarily orthogonal.* This implies that among the eigenvectors, there can be at the most only one time-like eigenvector. Likewise, since no two distinct light-like vectors can be orthogonal, if they are eigenvectors, they must have the same eigenvalue. The classification is now arranged according to *number of distinct eigenvalues* [18].

Type I: Four distinct eigenvalues. This implies 4 orthogonal eigenvectors. If one of these is light-like, then the remaining three must be space-like and this is impossible. Hence the eigenvectors must form a (pseudo-)orthonormal basis, E^a_μ, $a = 0, 1, 2, 3$, $E^a \cdot E^b = \eta^{ab}$. The $T^\mu_{\ \nu}$ is diagonalizable and can be expressed as,

$$T_{\mu\nu} = -\rho E^0_\mu E^0_\nu + \sum_{i=1}^{3} p_i E^i_\mu E^i_\nu.$$

The p_i are called *principle pressures* and ρ is called the *rest energy density.* Stress tensors of all observed massive and massless fields are of type-I, except for a very special case below.

Type II: Three distinct eigenvalues. So one eigenvalue is repeated once. The eigenspace of the repeated eigenvalue is two-dimensional. If every vector in this space is space-like *or* if one of these vectors is time-like, then we are back to the previous case. So a new case arises when the eigenvector for the repeated eigenvalue is light-like and with no other eigenvector in the subspace. This case thus has a double light-like eigenvector and two space-like ones.

Two get a canonical form, let k^μ be the double light-like eigenvector with eigenvalue λ and let E^2_μ, E^3_μ be the remaining two eigenvectors. Introduce another light-like vector ℓ_μ which is orthogonal to E^2, E^3 and $k \cdot \ell = 2$. Then the stress tensor can be written as,

$$T_{\mu\nu} = \sigma k_\mu k_\nu + \frac{\lambda}{2}(k_\mu \ell_\nu + k_\nu \ell_\mu) + p_2 E^2_\mu E^2_\nu + p_3 E^3_\mu E^3_\nu, \quad \sigma \neq 0 \text{ is arbitrary.}$$

Only known situation is a massless field with radiation going along k_μ (all eigenvalues are zero i.e. $\lambda, p_2, p_3 = 0$) This is said to represent 'null fluids'.

Type III: Two distinct eigenvalues. By similar argument as above, a new case arises when the repeated eigenvalue has a triple, light-like eigenvector and of course one space-like one. Let k^μ be the triple light-like eigenvector with eigenvalue λ. Introduce ℓ as before and another orthogonal vector e_μ which is space-like and is also orthogonal to E^3_μ. The canonical form is then expressed as,

$$T_{\mu\nu} = \frac{\lambda}{2}(k_\mu \ell_\nu + k_\nu \ell_\mu) + (k_\mu e_\nu + k_\nu e_\mu) + \lambda e^2_\mu e^2_\nu + p_3 E^3_\mu E^3_\nu.$$

This ensures that only k_μ and E^3_μ are the eigenvectors.

There are no physical examples of this type.

Type IV: There is no time-like or light-like eigenvector at all. There are no physical examples of this type either. This happens when there is a pair of complex conjugate eigenvalues. This cannot happen in the space-like subspaces hence it suffices to have the canonical form as,

$$T_{\mu\nu} = (\sigma-\rho)k_\mu k_\nu - (\sigma+\rho)\ell_\mu\ell_\nu - \rho(k_\mu\ell_\nu + k_\nu\ell_\mu) + p_2 E^2_\mu E^2_\nu + p_3 E^3_\mu E^3_\nu, \ \ \rho^2 < \sigma^2$$

Note that if all eigenvalues are equal, then $T_{\mu\nu} = \Lambda g_{\mu\nu}$ and this falls in the type I of diagonalizable stress tensors.

For *diagonalizable* stress tensor, the energy conditions take the form:

Weak	$\rho \geq 0, \quad \rho + p_i \geq 0$;
Strong	$\rho + \sum p_i \geq 0, \quad \rho + p_i \geq 0$;
Dominant	$\rho \geq \lvert p_i \rvert$

These energy conditions play a role in the discussion of singularity theorems.

4.4 Operational Determination of the Metric

Is it possible to operationally determine the metric tensor? It is and as follows [18]:

Let us assume that we have a procedure to determine which events, location and time stamps, can be physically connected so that it is possible to empirically determine a *local light cone*. From a given location and a time stamp, send out a light wave and after a while assign the same time label to all the spatial locations on the wave front. Having identified the local light cone in space-time, we infer the light-like directions in the tangent space at the given event.

Choose a time-like vector X and a space-like vector Y in the tangent space and consider the equation,

$$(X + \lambda Y) \cdot (X + \lambda Y) = 0 = X \cdot X + 2\lambda X \cdot Y + \lambda^2 Y \cdot Y.$$

We know the roots of this equation and from their product we find the ratio of the magnitudes of X and Y, $\lambda_+\lambda_- = X \cdot X / Y \cdot Y$. Thus, we know the ratio of magnitudes of a space-like and a time-like vector.

Next, for any two non-light-like vectors, Z, W such that $Z + W$ is not light-like, we can use,

$$- Z \cdot W = \frac{1}{2}\left[Z \cdot Z + W \cdot W - (Z + W) \cdot (Z + W)\right]$$

The magnitudes on the right-hand side can be expressed in terms of magnitude of X or Y. Therefore, we know $Z \cdot W$ in terms of magnitude of Y say. If $Z + W$ is light-like, use the magnitude of $Z + 2W$ to get $Z \cdot W$. We see that we can determine the '$\cdot$' product and hence the metric in terms of magnitude of *one* vector i.e. up to a scale. Evidently, we can repeat this process for each point and therefore determine the metric up-to a space-time dependent conformal factor. However, we also have the requirement of conservation of the stress tensor, $\nabla_\mu T^{\mu\nu} = 0$ and this equation is *not* invariant under conformal scaling. Hence the demand of conservation fixes the conformal factor *up to a space-time constant* and this is dependent on the units chosen.

Thus, experimental determination of local light cone and the covariant conservation of the stress tensor determine the space-time metric modulo choice of units.

Chapter 5

Elementary Phenomenology

General relativity brought in a huge conceptual change regarding the nature of gravitation. It introduced a sophisticated model for possible space-times, required it to be *dynamical* and provided a specific equation determining space-times appropriate in various physical contexts. Within this model, the motion of test bodies under Newtonian gravitational force is understood as geodesics of corresponding space-time. This forms the basis for the *solar system tests* of general relativity. As we saw in the discussion of wave motion in geometrical optics approximation, light too responds to gravity following light-like geodesics. Apart from these test bodies implications, general relativity impacts compact stars and their stability, strongly suggests new types of objects called black holes, points to the possibility of a 'singular' beginning for an expanding universe and makes a brand new prediction of gravitational waves. This chapter is arranged according to these different implications of the theory.

In the following, we use the geometrized units: $c = 1, G = 1$ and the Einstein equation is taken in the form,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}. \tag{5.1}$$

5.1 Geodesics and the Classic Tests

The first set of predictions were in the context of solar system where the Newtonian theory was applied and tested extensively. To make new predictions based on the idea of planetary motions being geodesics, we have to first choose a space-time appropriate for our solar system. In the section 2.4 we have already introduced the idealized solar system. We noted that the appropriate space-time should be time independent, spherically symmetric and should satisfy the source-free Einstein equation in the region exterior to the Sun.

Since the coordinates are arbitrary and have no particular physical interpretation, the notion of a symmetry cannot be based on specific coordinate transformation unless suitable coordinates can be singled out. It is convenient to consider first infinitesimal symmetries.

Consider a vector field $\xi^\mu(x)$ which enables us to make an infinitesimal coordinate transformation, $x^\mu \to x'^\mu(x) := x^\mu + \epsilon\xi^\mu(x)$. Under this, the metric transforms as

$$
\begin{aligned}
g'_{\mu\nu}(x+\epsilon\xi) &= \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta}(x) \\
\therefore \delta g_{\mu\nu} := g'_{\mu\nu}(x) - g_{\mu\nu}(x) &\approx -\epsilon(\xi^\alpha \partial_\alpha g_{\mu\nu} + \partial_\mu \xi^\alpha \ g_{\alpha\nu} + \partial_\nu \xi^\alpha \ g_{\alpha\mu}) \\
&= -\epsilon(\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu) \quad =: \quad -\epsilon \mathcal{L}_\xi g_{\mu\nu} \qquad (5.2)
\end{aligned}
$$

If it so happens that $\delta g_{\mu\nu} = 0$ under the infinitesimal transformation, then we say that the vector field is a *Killing vector field* and satisfies the Killing equation $\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0$. The infinitesimal transformation is said to be an *infinitesimal isometry.* The calculation says that if we move along an infinitesimal curve from a point p, in the direction given by $\xi^\mu(p)$, then the metric does not change. It also means that the metric is independent of the parameter, s, labelling points on the *integral curve* of ξ, defined by $\frac{dx^\mu(s)}{ds} = \xi^\mu(x(s))$. This equation being an ordinary differential equation, it always has a local solution and thus integral curves always exist for smooth vector fields. It however is not always possible to find a hypersurface Σ (a surface of n -1 dimension in an n-dimensional manifold), to which a given vector field is *orthogonal.* The condition for a vector field ξ^μ to be *hypersurface orthogonal* is that $0 = \xi_\lambda(\nabla_\mu \xi_\nu - \nabla_\nu \xi_\mu)$ + cyclic permutations of $(\lambda\mu\nu)$. This is a form of the *Frobinius theorem* [17]. We note that linear combinations of Killing vectors is a Killing vector and the *commutator* of two Killing vectors $[\xi^\mu \partial_\mu, \ \eta^\nu \partial_\nu] = (\xi \cdot \nabla \eta^\alpha - \eta \cdot \nabla \xi^\alpha)\partial_\alpha$, is also a Killing vector. We are now ready to characterize static, spherically symmetric space-times.

A space-time is said to be *stationary* if there exists a *time-like Killing vector,* ξ. It is *static*, if the vector field is hypersurface orthogonal. It is said to be *spherically symmetric* if there exist *three space-like Killing vectors*, ξ_a such that $[\xi_a, \ \xi_b] = \epsilon_{ab}{}^c \xi_c$ *and* the set of points reached from a given point by all possible shifts along the Killing vectors ξ_a (i.e. an orbit of $SO(3)$) is a 2-sphere.

Let t be the parameter along the stationary Killing vector. Staticity implies there is a Σ which is orthogonal to ξ and therefore Σ is space-like. For an arbitrary choice of coordinates x^i on Σ, label integral curve of ξ passing through $p \in \Sigma$, by the spatial coordinates of p and assign the same value, t to all points of Σ. For points q on the integral curve through p, assign the coordinates (t', x^i) where t' is the value of the Killing parameter and x^i are the same spatial coordinates of p. ξ being a Killing vector implies the metric $g_{\mu\nu}$ is independent of t. The staticity implies that $g_{ti}(x^j) = 0$. The metric is now invariant also under $t \to -t$.

The orbit spheres of spherical isometries lie within Σ and each sphere has an induced metric on it which much be proportional to the standard metric on an S^2. Label an orbit sphere by its *areal radial coordinate*, $r := \sqrt{\text{area}/(4\pi)}$. Choose an orbit sphere and introduce the standard spherical polar coordinates (θ, ϕ) on it. On this, the metric takes the form $\Delta s_2^2 = r^2(\Delta\theta^2 + \sin^2\theta \Delta\phi^2)$. Consider space-like geodesics emanating orthogonally from this sphere and and carry the angular coordinates of the point along the geodesics. This introduces the spatial coordinates r, θ, ϕ throughout Σ. The spatial metric then takes the form $\Delta s_3^2 = g(r)\Delta r^2 + \Delta s_2^2$. The orthogonality of geodesics implies that $g_{r\theta} = 0 = g_{r\phi}$. This procedure of setting up a coordinate system using the availability of the Killing vectors restricts the form of the metric to [17],

$$\Delta s^2 \;=\; -f(r)\Delta t^2 + g(r)\Delta r^2 + r^2(\Delta\theta^2 + \sin^2\theta\Delta\phi^2).$$

The coordinates themselves are called the Schwarzschild coordinates. Note that there is freedom to scale the time coordinate by a constant which may be absorbed in f. This freedom will be used below. The two unknown functions f, g are determined by the Einstein equation, $R_{\mu\nu} = 0$ since exterior to the Sun, there is no matter.

Straight forward application of the definitions (see section 14.5) 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

$$\begin{aligned}
R_{tt} &= \frac{f''}{2g} - \frac{1}{4}\left(\frac{f'}{g}\right)\left(\frac{g'}{g} + \frac{f'}{f}\right) + \frac{f'}{rg};\\
R_{rr} &= -\frac{f''}{2f} + \frac{1}{4}\left(\frac{f'}{f}\right)\left(\frac{g'}{g} + \frac{f'}{f}\right) + \frac{g'}{rg};\\
R_{\theta\theta} &= 1 - \frac{r}{2g}\left(-\frac{g'}{g} + \frac{f'}{f}\right) - g^{-1};\\
R_{\phi\phi} &= sin^2\theta\ R_{\theta\theta}; \qquad \text{all other components are zero.} \qquad (5.3)
\end{aligned}$$

The condition of Ricci flatness implies, $g^{-1}R_{rr} + f^{-1}R_{tt} = 0 \;\Rightarrow\; fg =$ constant. In view of the scaling freedom in the definition of t we can take this constant to be equal to 1 so that the metric approaches the standard Minkowski metric for r tending to infinity. 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 (see section 4.1) for large r, we see that $f(r) = -\, g_{tt} = 1 + 2\Phi(r)$ which gives the identification, $R_S = 2M$ (or $R = \frac{2GM}{c^2}$). Thus we have the famous Schwarzschild solution (1916). R_S is the *Schwarzschild radius.*

5.1.1 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,

$$0 = \ddot{t} + \frac{f'}{f}\,\dot{r}\,\dot{t} \tag{5.4}$$

$$0 = \ddot{r} + \frac{f'}{2g}\dot{t}^2 + \frac{g'}{2g}\dot{r}^2 - \frac{r}{g}\dot{\theta}^2 - \frac{r\sin^2\theta}{g}\dot{\phi}^2 \tag{5.5}$$

$$0 = \ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} - \sin\theta\,\cos\theta\,\dot{\phi}^2 \tag{5.6}$$

$$0 = \ddot{\phi} + \frac{2}{r}\dot{r}\dot{\phi} + 2\cot\theta\,\dot{\theta}\dot{\phi} \tag{5.7}$$

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

$$0 = \ddot{t} + \frac{f'}{f}\,\dot{r}\,\dot{t} \quad \Rightarrow \quad \dot{t}f =: E \quad (\text{ a positive constant }) \tag{5.8}$$

$$0 = \ddot{r} + \frac{f'}{2g}\dot{t}^2 + \frac{g'}{2g}\dot{r}^2 - \frac{r}{g}\dot{\phi}^2 \tag{5.9}$$

$$0 = \ddot{\phi} + \frac{2}{r}\dot{r}\dot{\phi} \quad \Rightarrow \quad r^2\dot{\phi} =: EL \quad (\text{ a constant. }) \tag{5.10}$$

The radial equation can be integrated once to yield,

$$g\dot{r}^2 + E^2\left(\frac{L^2}{r^2} - \frac{1}{f}\right) =: -E^2\kappa \quad (\ \kappa \text{ is a constant. }) \tag{5.11}$$

It is easy to see by substitution that,

$$\left(\frac{ds}{d\lambda}\right)^2 = -\,E^2\kappa \quad (\ \leq 0\ ) \tag{5.12}$$

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,

$$r^2\frac{d\phi}{dt} = Lf \tag{5.13}$$

$$\frac{g}{f^2}\left(\frac{dr}{dt}\right)^2 - \frac{1}{f} + \frac{L^2}{r^2} = -\kappa \tag{5.14}$$

$$\left(\frac{ds}{dt}\right)^2 = -\kappa f^2 \tag{5.15}$$

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,

$$0 = \frac{g}{r^4}\left(\frac{dr}{d\phi}\right)^2 + \frac{1}{r^2} + \frac{1}{L^2}\left(\kappa - \frac{1}{f}\right) \quad \text{or} \tag{5.16}$$

$$\phi(r) = \pm \int dr \frac{\sqrt{g}}{r^2}\left(\frac{1}{L^2}(f^{-1} - \kappa) - \frac{1}{r^2}\right)^{-\frac{1}{2}} \tag{5.17}$$

These are the general set of equations for geodesics. The geodesics 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.

5.1.2 Deflection of Light

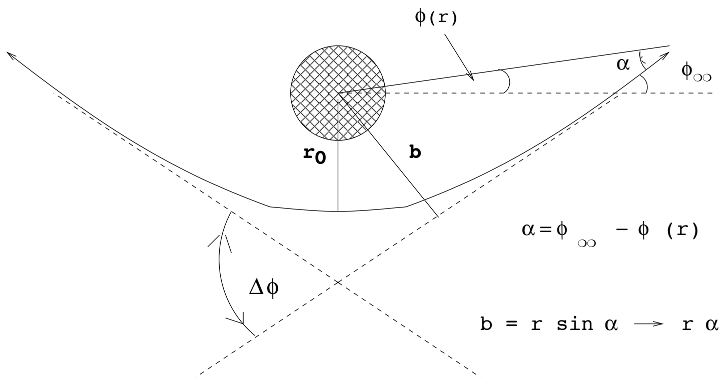


FIGURE 5.1: Deflection of light by a massive body.

Let us consider the scattering problem first. The geometry is shown in the figure (5.1). Asymptotically r is very large and thus $f, g \approx 1$. The incoming radial speed v, defined as $v := -\frac{dr cos\alpha}{dt}$ 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 ($\frac{d\phi}{dr} > 0$),

$$|L| = r_0\sqrt{f(r_0)^{-1} - 1 + v^2} \qquad \text{and the } \phi \text{ integral becomes,} \tag{5.18}$$

$$\phi(r) = \phi_\infty + \int_\infty^r dr \frac{\sqrt{g}}{r^2}\left[\frac{1}{r_0^2}\frac{f(r)^{-1} - 1 + v^2}{f(r_0)^{-1} - 1 + v^2} - \frac{1}{r^2}\right]^{-\frac{1}{2}} \tag{5.19}$$

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_S/r. 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:

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

$$g(r) = \left(1 + \gamma\frac{R_S}{r} + \cdots\right) \tag{5.20}$$

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

$$\Delta\phi = \frac{2R_S}{r_0}\left(\frac{1+\gamma}{2}\right) = \left(\frac{R_\odot}{r_0}\right)\left(\frac{2R_S}{R_\odot}\right)\left(\frac{1+\gamma}{2}\right) \tag{5.21}$$

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

$$\Delta\phi_\odot \approx 1.75''\left(\frac{1+\gamma}{2}\right)\left(\frac{R_\odot}{r_0}\right) \tag{5.22}$$

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} [19].

This phenomenon of phenomenon of 'bending of light', leads to *gravitational lensing.* Light from distant sources would get distorted due to intervening matter distribution producing multiple images and/or distorted images of the same source. The first identification of lensing was in 1979 by Dennis Walsh, Robert F. Carswell and Ray J. Weymann who identified two *quasars* as two images of the same quasar. Subsequently many examples were discovered including the famous Einstein cross, 4 images, in 1985 and the Einstein ring in 1988. There is interesting history of the prediction of gravitational lensing [20] and more aspects of it such as micro-lensing and weak lensing have been identified and are used as observational probes [21, 22].

5.1.3 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,

$$0 = \frac{1}{r_\pm^2} - \frac{1}{L^2 f_\pm} + \frac{\kappa}{L^2} , \qquad f_\pm := f(r_\pm) , \quad \Rightarrow \tag{5.23}$$

$$\kappa = \frac{\frac{r_+^2}{f_+} - \frac{r_-^2}{f_-}}{r_+^2 - r_-^2} ; \tag{5.24}$$

$$L^2 = \frac{\frac{1}{f_+} - \frac{1}{f_-}}{\frac{1}{r_+^2} - \frac{1}{r_-^2}} ; \qquad \text{also,} \tag{5.25}$$

$$\phi(r) = \phi(r_-) + \int_{r_-}^{r} \frac{dr}{r^2}\sqrt{g}\left\{\frac{1}{L^2 f} - \frac{\kappa}{L^2} - \frac{1}{r^2}\right\}^{-\frac{1}{2}} \tag{5.26}$$

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,

$$\text{Precession per revolution} \equiv \Delta\phi := 2|\phi(r_+) - \phi(r_-)| - 2\pi. \tag{5.27}$$

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. For solar system objects, one can compute the precession to first order in R_S. Using the Robertson parameterization ($\gamma = 1, \beta = 1$ for Schwarzschild),

$$g(r) = 1 + \gamma\frac{R_S}{r} + \cdots$$

$$f(r) = 1 - \frac{R_S}{r} + \frac{(\beta - \gamma)}{2}\left(\frac{R_S}{r}\right)^2 + \cdots \qquad \Rightarrow$$

$$f^{-1}(r) = 1 + \frac{R_S}{r} + \frac{(2 - \beta + \gamma)}{2}\left(\frac{R_S}{r}\right)^2 + \cdots , \tag{5.28}$$

leads to the formula [2],

$$\Delta\phi = (2 + 2\gamma - \beta)\pi R_S\left[\frac{1}{2}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)\right] \tag{5.29}$$

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 $\ell = a(1 - e^2)$ and the precession per revolution is given by,

$$\Delta\phi = 3\pi \frac{2GM}{c^2} \frac{1}{\ell} \tag{5.30}$$

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 the Sun also cause precession. Further discussion may be seen in Weinberg's book [2].

5.2 Relativistic Cosmology

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 90 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 what is special about our position and why we occupy it. An alternative is to reject the idea that there is anything special about our location in the universe and propose instead that the 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 elsewhere 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 world line) who will not be able to detect any distinguished direction. The statement that on large scale the universe is spatially homogeneous and isotropic is called the *cosmological principle*[1]. The so-called standard cosmology is based on spatial homogeneity and isotropy and this is what is discussed below.

In order to arrive at a suitable form of the metric, we need to characterize precisely what is meant by spatially homogeneous and isotropic in the context of geometry. The first task is to be able to identify a *spatial slicing* of the space-time. This is achieved by stipulating that there exist *a* one parameter family of space-like hypersurfaces, Σ_τ, foliating the space-time. A space-time is said to be spatially homogeneous if there is a *transitive* action of a group of isometries on each of the spatial slices. Here, transitive action means that given any two points on a Σ_τ, there is a diffeomorphism of Σ_τ on to itself. This being an isometry means that the metric remains the same. There can be more than one such isometries connecting two points.

Isotropy is a stipulation associated with observers. Let $x^\mu(t)$ be a time-like curve representing worldline of an observer. The observer is said to be an *isotropic observer* if at any point $p \in x^\mu(t)$ and for a pair of space-like tangent vectors in the tangent space at p, there exists an isometry which leaves p and the tangent vector $u^\mu := \frac{dx^\mu}{dt}|_p$ unchanged but maps one direction to the other. A space-time is said to be isotropic at every point if there exist a space-time filling congruence of isotropic observers i.e. a time-like vector field, u^μ, whose integral curves represent isotropic observers, variously called as cosmic observers or fundamental observers.

Isotropy implies that the vector field must be orthogonal to surfaces of homogeneity. For if it were not, its projection on the tangent space to Σ_τ will give a distinguished direction which is disallowed by isotropy. If there are more than one family of hypersurfaces of homogeneity, then isotropy implies that at least one of these must be orthogonal to the vector field. Note that, isotropy at each point does not imply/require spatial homogeneity, nor does spatial homogeneity imply isotropy. However, if we have both of these, then the isotropy vector field is orthogonal with the surfaces for homogeneity. We can choose the label τ of the family of hypersurfaces as a time coordinate and given any choice of spatial coordinates, x^i, on a Σ_{τ_0} carry them along the world lines of the isotropic observers. This immediately gives a block diagonal

---

[1] 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*. For a discussion of alternative cosmologies, see [2].

form of the matric with $g_{\tau i} = 0$. We can also relabel the surfaces so that the metric coefficient $g_{\tau\tau} = -1$.

Isotropy restricts the form of the spatial metric severely. The Riemann tensor R_{ijkl} of the spatial metric can be regarded as a symmetric 6×6 matrix in the antisymmetrized pairs of indices $[ij]$ and $[kl]$. If it has distinct eigenvalues, then the corresponding eigenvectors can be uniquely distinguished and these will be 2-forms. From these, we can uniquely obtain a distinguished dual 1-form or equivalently, a tangent vector. This would contradict isotropy. Hence all eigenvalues must be equal i.e. the spatial Riemann tensor must have constant curvature: $R_{ijkl} = \lambda(g_{ik}g_{jl} - g_{il}g_{jk})$. As noted in section 2.5, such constant curvature space are completely classified and lead to the line-elements given in (2.28).

Computation of the Einstein tensor, proceeds as in the case of the Schwarzschild metric, and leads to the non-vanishing components,

$$\Gamma^{\tau}{}_{ij} = \frac{\dot{a}}{a} g_{ij} \quad , \quad \Gamma^{i}{}_{\tau j} = \frac{\dot{a}}{a}\delta^i_j \ , \ \Gamma^{i}{}_{jk} = \hat{\Gamma}^{i}{}_{jk} \tag{5.31}$$

$$R_{\tau\tau} = -3\frac{\ddot{a}}{a} \quad , \quad R_{ij} = g_{ij}\left(\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + 2\frac{k}{a^2}\right) \tag{5.32}$$

Here the hatted Γ denotes the connection corresponding to the comoving metric which is normalized so that the Ricci scalar, $\hat{R} = 6k$, $k = \pm 1,\ 0$.

5.2.1 Friedmann–Lamaitre–Robertson–Walker Cosmologies

The 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 a perfect fluid:

$$T_{\mu\nu} \ = \ \rho(\tau)u_\mu u_\nu \ + \ P(\tau)(u_\mu u_\nu \ + \ g_{\mu\nu}), \tag{5.33}$$

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. The Einstein equations reduce to[2]:

$$3\frac{\ddot{a}}{a} = -4\pi(\rho + 3P) \qquad \text{(Raychaudhuri equation)} \tag{5.34}$$

$$3\frac{\dot{a}^2}{a^2} = 8\pi\rho - \tfrac{3}{a^2}k \quad k = \pm 1, 0\ ; \qquad \text{(Friedmann equation)} \tag{5.35}$$

$$\dot{\rho} = -3(\rho + P)\tfrac{\dot{a}}{a} \qquad \text{(Conservation equation)} \tag{5.36}$$

The first striking implication 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

[2]The equations are with zero cosmological constant. Cosmological constant may be included by the replacements: $\rho \to \rho + \frac{\Lambda}{8\pi}, P \to P - \frac{\Lambda}{8\pi}$.

choice of k. Further, $\ddot{a} < 0$ implies $\dot{a}$ must be monotonically decreasing implies that it cannot 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.

The observed fact of expanding universe immediately implies that the universe must have been extremely small a finite time ago and $\ddot{a} < 0$ implies that it must have been expanding at a faster rate in the past. If it were expanding at today's rate, $H_0 := \frac{\dot{a}_0}{a_0}$, then the scale factor would have been zero at $\tau = H_0^{-1}$. 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 (5.35) and (5.36) imply the second order equation (5.34 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. It characterizes internal dynamical properties of matter at a phenomenological level 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. The different cases are referred to as Friedmann–Robertson–Walker (FRW) cosmologies and are summarized as [17],

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

Explicit computations may be seen in [16].

One can get this far with just cosmological principle, general relativity 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 $P_\Lambda = -\rho_\Lambda = -\Lambda$, with the factor of $(8\pi)^{-1}$ understood to be absorbed in Λ. The total energy density and pressure are 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 when the universe was hotter. 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 ρ_Λ = constant. Thus a non-zero values of the cosmological constant is overwhelmed by the other sources of pressure and energy density in the early universe but can be significant in the late epochs. It is now assumed that these behaviours of the energy densities continues to hold generally to a very good approximation and 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 Raychaudhuri equations can be rewritten as

$$\Omega = 1 + \frac{\kappa}{a^2 H^2} \tag{5.37}$$

$$P = \frac{H^2}{8\pi G}(2q - \Omega) = -\frac{1}{8\pi G}\left(\frac{\kappa}{a^2} + (1 - 2q)H^2\right) \tag{5.38}$$

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

$$\left(\frac{\dot{a}}{a_0}\right)^2 = \frac{8\pi G}{3}\frac{a^2}{a_0^2}\left(\rho_{NR,0}\left(\frac{a_0}{a}\right)^3 + \rho_{R,0}\left(\frac{a_0}{a}\right)^4 + \Lambda\right) - \frac{\kappa}{a_0^2} \tag{5.39}$$

$$= H_0^2\left[\Omega_{NR,0}\left(\frac{a_0}{a}\right) + \Omega_{R,0}\left(\frac{a_0}{a}\right)^2 + \Omega_{\Lambda,0}\left(\frac{a}{a_0}\right)^2 - \Omega_{NR,0} - \Omega_{R,0} - \Omega_{\Lambda,0} + 1\right] \tag{5.40}$$

Here we have used equation (5.37) at present epoch, in going to the second equation above. Putting $a = a_0 x$ gives,

$$\dot{x}^2 = H_0^2\left[\Omega_{NR,0}x^{-1} + \Omega_{R,0}x^{-2} + \Omega_{\Lambda,0}x^2 + 1 - \Omega_0\right] \tag{5.41}$$

The right-hand side of this equation involves the present values of the density parameters. If these are determined observationally, 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,

$$H^2(x) = H_0^2\left[\Omega_{NR,0}x^{-3} + \Omega_{R,0}x^{-4} + \Omega_{\Lambda,0} + (1-\Omega_0)x^{-2}\right] \tag{5.42}$$

Using equation (5.37) and its present epoch version (for $\kappa \neq 0$), we can eliminate the Hubble parameter in favour the total energy density to get,

$$\Omega(x) - 1 = \frac{\Omega_0 - 1}{\left[\Omega_{NR,0}x^{-1} + \Omega_{R,0}x^{-2} + \Omega_{\Lambda,0}x^2 - (\Omega_0 - 1)\right]} \tag{5.43}$$

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 FLRW cosmologies.

One can obtain relations among the present epoch density parameters, Hubble parameter and the deceleration parameter by simply writing the equations (5.37, 5.38) 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,

$$\Omega_{NR,0} + \Omega_{\Lambda,0} = 1 + \frac{\kappa}{a_0^2 H_0^2}, \quad \Omega_{NR,0} - 2\Omega_{\Lambda,0} = 2q_0, \tag{5.44}$$

Thus, if we could determine q_0 and a_0H_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[3]. The distance typically involves the *comoving path length*, ℓ, which is defined as,

$$\ell := \int_{\tau_e}^{\tau_o} \frac{d\tau}{a(\tau)} = \int \frac{da}{\dot{a}a} = \frac{1}{a_0}\int \frac{dx}{x\dot{x}}, \tag{5.45}$$

and we have the equation (5.41) 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 FLRW 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,

$$\begin{aligned} D_L(z) &= \frac{(1+z)}{H_0}\int_0^z dz \times \\ & \left[\Omega_{NR,0}(1+z)^3 + \Omega_{R,0}(1+z)^4 + \Omega_{\Lambda,0} + (1-\Omega_0)(1+z)^2\right]^{-\frac{1}{2}} \end{aligned} \tag{5.46}$$

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), the red-shifts have gone to about 2 in the distance-red-shift plot.

5.2.2 Digression on Big-Bang Cosmology: Thermal History

So far we have sketched qualitative evolution of the gross features of the universe according to the FLRW 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.

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

[3]There are various measures of distances in cosmology - the *luminosity distance* inferred from the apparent brightness 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 $(1+z)$ [2].

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 into 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 made 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.

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) got formed during the expansion of the universe. The question we want to understand is: What determines the products and their abundances during the formation 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 and thermal contact between the species is broken. As the universe decelerates, the reaction rates fall *faster* than the expansion rate thereby switching off some reactions. The number densities (or abundances) of the participants are thus 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.

Let us note another qualitative feature that can be expected. At sufficiently high temperature, we expect a gas of interacting charged (+ neutral) particles together with 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 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, *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 [2].

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 $T_\gamma(t) = A/a(t)$. Since the black body distribution form is preserved during expansion as a consequence of Robertson–Walker geometry independent of dynamics[4], 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 [2]. We know the densities and pressures of matter and photons at temperature T,

$$\begin{aligned}\rho(T) &= bT^4 + mn + \frac{nk_BT}{\gamma - 1}, \quad P(T) \ = \ \frac{1}{3}bT^4 + nk_BT, \\ b &:= \frac{8\pi^5 k_B^4}{15h^3c^3} \ \sim \ 7.6 \times 10^{-15}(\text{cgs}),\end{aligned} \tag{5.47}$$

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 $n(\tau)a^3(\tau))$ is constant while conservation equation (5.36) leads to $a\frac{d(\rho a^3)}{da} = -3Pa^3$. Computing the left-hand side and using (5.47) leads to,

$$\frac{a}{T}\frac{dT}{da} = -\left[\frac{\sigma + 1}{\sigma + \frac{1}{3(\gamma-1)}}\right], \quad \sigma := \frac{4bT^3}{3nk_B} \tag{5.48}$$

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

[4] At equilibrium, the number of photons of frequency ν per unit volume per unit frequency interval is proportional to ν^2 divided by the Planck factor which depends on $h\nu/KT$. Conservation of number of photons during free streaming, means that the number is preserved and therefore the density at a later time scales as $(a_{later}/a_{initial})^2$. In the Planck formula, the ν^2 dependence also scales the same way, thus cancelling the scaling factor. The net result of the expansion is then to retain the shape but scale the temperature by $a_{initial}/a_{later}$ [23].

the large constant value of σ to the ratio of the 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, $n + p \rightleftharpoons D + \gamma$. 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 as the cube root of the ratio of the densities of baryons, 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 1940s! The photons present in this epoch will eventually constitute the cosmic microwave background radiation (CMBR).

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 (5.36) can be written as,

$$\frac{d\{a^3(P+\rho)\}}{d\tau} \;=\; a^3\frac{dP}{d\tau} \;=\; a^3\frac{dT}{d\tau}\frac{dP}{dT} \tag{5.49}$$

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

$$\begin{aligned} dS(T,V) &= \frac{1}{T}(P(T)+\rho(T))dV + \frac{V}{T}\frac{\partial\rho}{\partial T}dT \;\Rightarrow \\ \frac{dP}{dT} &= \frac{P+\rho}{T} \end{aligned} \tag{5.50}$$

These two equations together give the conserved quantity,

$$\frac{d}{d\tau}\left[\frac{a^3}{T}(P(T)+\rho(T))\right] \;=\; 0 \;=:\; \frac{d(a^3 s(T))}{d\tau}, \tag{5.51}$$

where $s(T)$ is the entropy density. This in turn allows us to relate the time variation of the temperature to the entropy production:

$$\frac{\dot{a}}{a} = -\frac{1}{3s}\frac{ds}{dT}\frac{dT}{d\tau} = \sqrt{\frac{8\pi\rho}{3}} \tag{5.52}$$

Here we used the Friedmann equation in the last equality.

4. Vanishing (or negligible) 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 is 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.

Steps 4 and 5 above are where the specific composition of the universe enters. 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 $4000°K$, 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. Details of these computations may be seen in many of the standard cosmology text books e.g. [2, 23].

Here is a summary of the thermal history of the universe, broadly divided into four epochs. It is convenient to describe the epochs in terms of the age (seconds), typical energy scale (GeV/MeV/eV) and corresponding temperature (°K) with $1°K \sim 10^{-4}eV$. As mentioned above, the temperature scales inversely with the scale factor. The age is then inferred from the evolution of the scale factor appropriate in the epochs.

The Very Early Epoch: ($10^{-42} - 10^{-4}$ *seconds)* The first and the earliest epoch which is also the least understood. The birth moment of the classical universe (FLRW space-time) is set at the Planck time. The temperature was about 10^{31} $°K$. This is supposed to be followed up by a period of *exponential expansion* or *inflation*. The main need for postulating inflation comes from the remarkable observed *isotropy* of the CMBR, discussed in the next section, and the corresponding *Horizon Problem*. The currently favoured amount of inflation is about 70 e-folds or expansion by a factor of about 10^{30} while the temperature falls by a factor of about 10^{-5}. The various candidate mechanisms of inflation also suggest a period of *reheating* following the inflationary phase, during which the energy is transferred from the presumed 'inflaton' field to other matter species which we eventually observe. This period is also supposed to set the conditions for the hot big bang e.g. the ratio of photon to baryon number densities being very large. The energy scale drops from the Planck scale ($10^{19}GeV$) to GUT scale ($10^{15}GeV$) to Electro-Weak scale ($10^{2}GeV$) to Hadronic scale($10^{-1}GeV$). By this time, hadrons have formed and hence this era is also known as the *Hadron era*.

The Lepton Era: ($10^{-2}-10^{0}$ *seconds)* Thanks to the electro-weak breaking, the soup has baryons, leptons and photons in approximate equilibrium. By the end of this era, the weak interactions effectively switch off and the neutrinos decouple. The energy scale is a few MeVs and the temperature has fallen to about 10^{10} $°K$.

The Plasma Era: (10^{2} *seconds*-10^{5} *years)* This is perhaps the best understood epoch. The temperature is now low enough ($\sim 10^{9}$ $°K$) for the protons and neutrons to begin forming the light nuclei such as deuterium, helium and lithium. This is called the *primordial nucleo-synthesis* and their observed abundances tell us that the big bang has been 'hot'. Following this phase, the nuclear interactions are no longer relevant and subsequent slower cooking is dominated by the electromagnetic interactions. By about 10^{4} years (10^{4} $°K$), the photons, electrons and protons (the plasma) are in equilibrium and with a further fall of temperature, the photons decouple constituting the CMBR.

The Post-Recombination Era: (10^{9} *years*$-$ *now)* The universe is cool enough to permit gravity to start binding the first structures. The first stars are supposed to have been formed around 250 million years after the big bang. These once again heat matter, locally, to produce ionization and generating light ending the so-called 'Dark Ages'. The subsequent formation of large scale structures such as galaxies began only after a billion years or so. The subsequent variety of structures at various scales are amenable to more direct observations allowing us to look for patterns in their distributions. These then provide constraints on models of the very early universe.

5.2.3 Cosmic Microwave Background Radiation

In building up our knowledge of the universe, we used 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 summary of thermal history discussed above, all these correspond to matter dominated era and our current observations go back to about $z \sim 10$ (about billion 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, the 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 the 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 produced 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 [24]. 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 $3.5 \pm 1^\circ K$. Its theoretical significance (identification with CMBR) was provided by Dicke, Peebles, Roll and Wilkinson [25]. This was of course observation at one frequency. Since then CMBR has been observed at wavelengths ranging from about 100 cm down to about 0.05 cm. The lower wavelengths are observed from balloon, rocket borne instruments and finally from the COsmic Background Experiment satellite. These ranges cover both sides of the Planck distribution curve and the current value of the photon temperature is 2.725 ± 0.001^0K.

One of the 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 provide a 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 that 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 gives 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 conclusive. Finally by 1992, COBE established presence of anisotropies to $l = 30$ multipole. The decade old *W*ilkinson *M*icrowave *A*nisotropy *P*robe (WMAP) made measurements to about $l \sim 600$. The most recent Planck Mission has measured anisotropies to $l \sim 2000$.

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,

$$\Delta(\hat{n}) \ := \ \frac{\Delta T}{\bar{T}}(\hat{n}) \ = \ \sum_{\ell,m} a_{\ell,m} Y_{\ell,m}(\hat{n}) \qquad (5.53)$$

The anisotropies are now encoded in the coefficients $a_{\ell,m}$. In principle, from the observed temperature distribution, one can infer the multipole coefficients $a_{\ell,m}$. However, what is relevant is not particular values of these coefficients, but rather their *statistical properties*.

We do not know the precise origin of the anisotropies. In principle they would be determined from some initial conditions in the plasma generating the CMB. We can make assumptions about the *statistical distribution* of these initial conditions and treat them as fluctuations. Averaging over the fluctuations is called *ensemble average* and we simply postulate the properties of these averages. For instance, we assume that the ensemble average of products of $\Delta(\hat{n})$ is *isotropic* and thus depend only on the rotational invariants constructed from the directions $\hat{n}$'s. This assumption is known as *statistical isotropy.* Furthermore, the fluctuations are generally assumed to be Gaussian so that all n-point correlations are reducible to products of the two point correlations.

Consider the two point angular correlation function which depends only on the cosine of the angle between the two directions by statistical isotropy. This can be expanded in terms of the Legendre polynomials [23],

$$\langle\Delta(\hat{n})\Delta(\hat{n}')\rangle_{\rm ens} := \sum_\ell C_\ell \frac{2\ell+1}{4\pi} P_\ell(\hat{n}\cdot\hat{n}') \quad \Rightarrow \tag{5.54}$$

$$C_\ell = \frac{1}{4\pi}\int\int d\Omega(\hat{n})d\Omega(\hat{n}')P_\ell(\hat{n}\cdot\hat{n}')\langle\Delta(\hat{n})\Delta(\hat{n}')\rangle_{\rm ens} \tag{5.55}$$

The statistical isotropy is equivalently characterized by the stipulation,

$$< a^*_{\ell,m} a_{\ell',m'} >_{\rm ensemble} = C_\ell \delta_{\ell,\ell'}\delta_{m,m'} \tag{5.56}$$

Obviously we cannot perform an ensemble average, but assuming ergodicity, we replace the ensemble average by the manifestation of the particular initial condition realized in our universe i.e. the observed data! Substitute the expansion of the anisotropy, eqn. 5.53, in the expression for the C_ℓ in eqn. 5.55 and use the addition theorem backwards to replace P_ℓ in terms of the spherical harmonics. *Without* performing the averaging, and using orthogonality of the spherical harmonics leads to,

$$C_\ell = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2$$

and a theory is supposed to give a prediction for the C_ℓ's.

A representative plot of measured C_ℓ against ℓ is shown in the figure (5.2). 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.*

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. The precise measurements of the anisotropies of CMBR is regarded as heralding the age of precision (observational) cosmology.

The beautiful isotropy of CMBR however poses a problem. Consider antipodal points in the sky at which the temperatures are equal to within 1 part

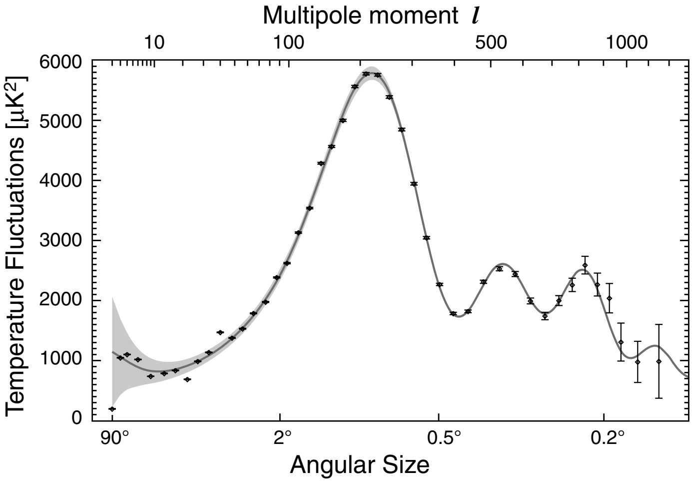


FIGURE 5.2: Sample Power Spectrum C_ℓ vs ℓ. The temperature fluctuation on the y-axis is proportional to the C_ℓ. The angular size is also shown. Credit: NASA / WMAP Science Team, http://map.gsfc.nasa.gov/media/111133/index.html

in 10^5. These regions of the plasma must have been in causal contact to have such a high correlation. However, the finite age of the universe in the FLRW models makes this impossible. Given that the current concordance model of cosmology favours spatially flat universe in which the scale factor vanishes as $\tau^{1/2}$ (radiation dominated era) and the present value of the Hubble parameter imply that regions with angular separation of more than about a degree *could not have been in causal contact*! This is the so-called *Horizon Problem* and was one of the primary motivation for Alan Guth's proposal of *inflation* according to which in the early universe there existed a period of *accelerated expansion.* With this, it is possible to solve the horizon problem. More than just solving the traditional problems of standard big bang cosmology, inflationary phase provides a mechanism for quantum fluctuations in the very early universe to be amplified and provide the seeds for subsequent structure formations. However interesting, it is well beyond the scope of this monograph. Interested readers may begin from [26].

This is an arena where General Relativity, particle physics and possibly quantum gravity have to come together in synthesizing a comprehensive picture.

5.3 Gravitational Waves

One of the distinctive features of GR is the existence of 'propagating solutions' of Einstein equation, known as *gravitational waves.*

This concept has been quite tricky to define because of the general covariance. Naively one could think of it as the metric varying in a region of the manifold which 'moves' to another region of the manifold. However, this could be just be effected by a diffeomorphism. To define a localized configuration of geometry and its propagation we need to have a reference geometry *and/or* a reference defined by some matter distribution.

In this first introduction, we will simply look at the *linearized theory*, identify the wave solutions, study their properties and also obtain the quadrupole formula of Einstein. In the part II of the book, we will return to the conceptual issues in section 10.1.

The linearisation of general relativity begins by postulating existence of a coordinate system in which the metric can be taken to be $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ with $|h_{\mu\nu}| \ll 1$ and deriving all relevant equation to first non-trivial order in h. In particular, $g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}$ and the indices are raised/lowered with the background Minkowskian metric $\eta^{\mu\nu}/\eta_{\mu\nu}$. Under an infinitesimal coordinate change, $x'^{\mu} \ := \ x^{\mu} + \xi^{\mu}(x)$, the metric transforms as,

$$g'^{\mu\nu}(x') \ = \ (\delta^{\mu}{}_{\alpha} + \partial_{\alpha}\xi^{\mu})(\delta^{\nu}{}_{\beta} + \partial_{\beta}\xi^{\nu})g^{\alpha\beta}(x)$$

the form of the metric is again preserved to first order in h, provided h and the infinitesimal transformation parameter, ξ are both taken to be of the same order and the h is transformed as,

$$\begin{aligned} h^{\mu\nu}(x') &= h^{\mu\nu}(x) - \partial^{\nu}\xi^{\mu} - \partial^{\mu}\xi^{\nu}, \\ h_{\mu\nu}(x') &= h_{\mu\nu}(x) - \partial_{\nu}\xi_{\mu} - \partial_{\mu}\xi_{\nu} \end{aligned} \tag{5.57}$$

To first order in h, the connection, the Ricci tensor and the Einstein tensor are given by,

$$\Gamma^{\lambda}{}_{\mu\nu} = \frac{1}{2}\eta^{\lambda\alpha}\left(h_{\alpha\mu,\nu} + h_{\alpha\nu,\mu} - h_{\mu\nu,\alpha}\right) + o(h^2) \tag{5.58}$$

$$\begin{aligned} R_{\mu\nu} &= R^{\lambda}{}_{\mu\lambda\nu} \ = \ \partial_{\lambda}\Gamma^{\lambda}{}_{\nu\mu} - \partial_{\nu}\Gamma^{\lambda}{}_{\lambda\mu} \ + o(h^2) \\ &= \frac{1}{2}\eta^{\lambda\alpha}\left(h_{\alpha\mu,\nu\lambda} + h_{\alpha\nu,\mu\lambda} - h_{\mu\nu,\alpha\lambda} - h_{\lambda\alpha,\mu\nu}\right) \\ &= \frac{1}{2}\left(-\Box h_{\mu\nu} - \partial_{\mu}\partial_{\nu}h^{\lambda}{}_{\lambda} + \partial^2_{\mu\lambda}h^{\lambda}{}_{\nu} + \partial^2_{\nu\lambda}h^{\lambda}{}_{\mu}\right) \end{aligned} \tag{5.59}$$

$$R = -\Box h^{\alpha}{}_{\alpha} + \partial^2{}_{\mu\nu}h^{\mu\nu}$$

$$G_{\mu\nu} = -\frac{1}{2}\left(\Box\tilde{h}_{\mu\nu} - \partial^2_{\lambda\mu}\tilde{h}^{\lambda}{}_{\nu} - \partial^2_{\lambda\nu}\tilde{h}^{\lambda}{}_{\mu} + \eta_{\mu\nu}\partial^2_{\alpha\beta}\tilde{h}^{\alpha\beta}\right) \quad \text{where,}$$

$$\tilde{h}_{\mu\nu} := h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h^{\alpha}{}_{\alpha} \tag{5.60}$$

Since the Einstein tensor is $o(h)$, we treat the stress tensor also to be of the same order. Consequently, its trace is taken with η. Explicitly, the linearized Einstein equation takes the form ($G = 1$ is taken),

$$-\Box\tilde{h}_{\mu\nu} + \partial^2_{\mu\lambda}\tilde{h}^{\lambda}{}_{\nu} + \partial^2_{\nu\lambda}\tilde{h}^{\lambda}{}_{\mu} - \eta_{\mu\nu}\partial^2_{\alpha\beta}\tilde{h}^{\alpha\beta} = 16\pi T_{\mu\nu}, \tag{5.61}$$

Notice that the divergence (∂^{μ}) of the linearized Einstein tensor is identically zero just as the covariant divergence is for the full Einstein tensor. Therefore conservation of the stress tensor also holds, $\partial_{\mu}T^{\mu}{}_{\nu} = 0$.

A couple of properties of this equation are noteworthy, namely, (a) $\delta\tilde{h}_{\mu\nu} := -(\partial_{\mu}\xi_{\nu} + \partial_{\nu}\xi_{\mu} - \eta_{\mu\nu}\partial\cdot\xi)$ is always a solution for every ξ_{μ} and (b) if we further stipulate that $\tilde{h}_{\mu\nu}$ satisfy: $\partial_{\mu}\tilde{h}^{\mu}{}_{\nu} = 0$, then the linearized equation simplifies to an *inhomogeneous wave equation*, $\Box\tilde{h}_{\mu\nu} = -\,16\pi T_{\mu\nu}$. In view of the equation (5.57), which is the linearized form of infinitesimal general covariance of the Einstein equation, we stipulate that two solutions $\tilde{h}'_{\mu\nu}, \tilde{h}_{\mu\nu}$ are to be regarded as *physically the same* if they differ by $\partial_{\mu}\xi_{\nu} + \partial_{\nu}\xi_{\mu} - \eta_{\mu\nu}\partial\cdot\xi$. This follows from the equation (5.57). In the context of the linearized theory, we view $\delta\tilde{h}_{\mu\nu}$ defined in the property (a) above as an infinitesimal *gauge transformation* while the condition in (b) is viewed as a *gauge fixing condition.* This gauge condition, is said to specify the so called *harmonic gauge*. The terminology arises from the *harmonic coordinate condition*, $\nabla_{\mu}\nabla^{\mu}x^{\alpha} = 0$ which translates into $g^{\mu\nu}\Gamma^{\lambda}{}_{\mu\nu} = 0$ and upon linearization, leads to the gauge condition. The harmonic gauge condition, still does not fix the gauge freedom completely: a gauge transformation with $\Box\xi_{\mu} = 0$ leaves both the wave equation and the gauge condition, invariant.

Our aim is to explore the *physical* solutions of *this* linearized theory.

5.3.1 Plane Waves

Let us consider now the homogeneous wave equation and look for *plane wave* solutions of the form,

$$\tilde{h}_{\mu\nu}(x) \;=\; \epsilon_{\mu\nu}e^{ik\cdot x} + \bar{\epsilon}_{\mu\nu}e^{-ik\cdot x}, \quad \text{over-bar denoting complex conjugate.}$$

Substituting in the linearized equations together with the harmonic gauge conditions implies that the wave vector k_{μ} and the (complex) polarization tensor, $\epsilon_{\mu\nu}$ must satisfy: $k\cdot k = 0,\ k_{\mu}\epsilon^{\mu}{}_{\nu} = 0$,. The gauge transformations of $h_{\mu\nu}$ implies a gauge transformation on the polarization tensor as: $\delta\epsilon_{\mu\nu} = -ik_{\mu}\xi_{\nu} - ik_{\nu}\xi_{\mu} + i\eta_{\mu\nu}k\cdot\xi$. Notice that the gauge condition is preserved by these residual gauge transformations. Thus, for a given null vector k, we have 10 components of the polarization tensor, the 4 conditions of the harmonic gauge and a 4 parameter worth of freedom of gauge transformations. This leaves us with *two physical polarization* parameters only.

For the source-free situation we are discussing, The harmonic gauge condition may be specialized further by stipulating that the trace of $\tilde{h}$ be zero[5]. This implies, $\epsilon := \epsilon^{\alpha}{}_{\alpha} = 0$ and in view of the transversality condition, $k_\mu \epsilon^{\mu\nu} = 0$, this further restricted gauge is referred to as the *transverse, traceless gauge* or TT gauge for short. The extra condition of tracelessness, reduces the gauge freedom to those ξ_μ which satisfy $k \cdot \xi = 0$. Note that in *the TT-gauge,* $\tilde{h}_{\mu\nu} = h_{\mu\nu}$.

To make the physical polarizations explicit, it is convenient to introduce an orthonormal tetrad, $e^{\mu}_{I}, I = 0, 1, 2, 3,\ \eta_{\mu\nu} e^{\mu}_{I} e^{\nu}_{J} = \eta_{IJ}$ and further allow complex linear combinations to get a *null tetrad* as: $k_{\pm} := (e_0 \pm e_1)/\sqrt{2},\ m := (e_2 + ie_3)/\sqrt{2},\ \bar{m} := (e_2 - ie_3)/\sqrt{2}$. All these are null vectors and the only non-zero dot products are $k_+ \cdot k_- = -1, m \cdot \bar{m} = 1$. For a given plane wave, we choose an adapted null tetrad so that $k_+ := k$ and denote the corresponding $k_- =: \ell$. The null tetrad is denoted as $e_a, a = k, \ell, m, \bar{m}$. Using the null tetrad, we can write the polarization tensor as: $\epsilon_{\mu\nu}(k) := \Phi_{ab} e^{a}_{\mu} e^{b}_{\nu}$ and the gauge transformation parameter as $\xi^\mu := \zeta^a e^{\mu}_{a}$. The gauge transformations in terms of Φ and ζ are: $\Phi'_{ab} = \Phi_{ab} - i(\delta^k_a \zeta_b + \delta^k_b \zeta_a)$.

The transversality condition on polarizations, $k^\mu \epsilon_{\mu\nu} = 0$, implies $\Phi_{\ell\ell} = \Phi_{\ell k} = \Phi_{\ell m} = \Phi_{\ell \bar{m}} = 0$. The tracelessness condition implies, $\Phi_{\ell k} = \Phi_{m\bar{m}}$ which in turn gives, $\Phi_{m\bar{m}} = 0$. We are left with 5 non-zero Φ's. The gauge parameter is restricted by $k \cdot \xi = 0$ and therefore has only $\zeta_k, \zeta_m, \zeta_{\bar{m}}$ components. These can be exhausted by setting $\Phi_{kk} = \Phi_{km} = \Phi_{k\bar{m}} = 0$ and we are left with two non-zero components, $\phi_{mm}, \Phi_{\bar{m}\bar{m}}$ and no gauge freedom.

The physical plane wave solutions are thus,

$$h(x)^{TT}_{\mu\nu} = e^{ik\cdot x}\left\{\Phi_{mm} m_\mu m_\mu + \Phi_{\bar{m}\bar{m}} \bar{m}_\mu \bar{m}_\nu\right\} + \text{complex conjugate}$$

The adapted tetrad has the plane wave propagating in the x^1 direction with the $x^2 - x^3$ plane being transverse to it. A rotation through an angle θ in the transverse plane, induces a phase change in the transverse null vectors: $m' = e^{-i\theta} m,\ \bar{m}' = e^{+i\theta}\bar{m}$. The null vectors are then said to have *helicity* -1 and $+1$ respectively. Clearly then, the Φ_{mm} represents the amplitude of helicity -2 and $\Phi_{\bar{m}\bar{m}}$ represents helicity $+2$.

There is an alternative way of identifying the independent, physical waves. From the null tetrad basis, we constructed the the corresponding basis for the second rank, symmetric tensors, namely, $\mathcal{E}^{ab}{}_{\mu\nu} := \frac{1}{2}(e^a \otimes e^b + e^b \otimes e^a)_{\mu\nu}$. This may be called the *helicity basis*, since each term has a definite helicity - $0, \pm 1, \pm 2$. A basis analogous to the 'plane polarizations' of electromagnetism, is defined as:

$$\mathcal{E}_+ \ := \ m \otimes m + \bar{m} \otimes \bar{m}, \quad \mathcal{E}_\times \ := \ -i(m \otimes m - \bar{m} \otimes \bar{m}).$$

[5]The transversality condition can always be imposed even in presence of source thanks to the conservation of the stress tensor. The trace-free condition can be imposed only in the source free situation or when the stress tensor is itself traceless. This however does *not* change the counting of the physical polarizations.

These are referred to as the 'plus' and the 'cross' polarizations. Under rotations, these transform as,

$$\mathcal{E}'_+ \ = \ \cos 2\theta \ \mathcal{E}_+ + \sin 2\theta \ \mathcal{E}_\times , \quad \mathcal{E}'_\times \ = \ -\sin 2\theta \ \mathcal{E}_+ + \cos 2\theta \ \mathcal{E}_\times .$$

We have already computed the corresponding Riemann tensor as well as the geodesic deviation equation in section (2.6). The linearized, plane gravitational waves thus propagate with the speed of light ($k^2 = 0$), are transverse ($k \cdot \epsilon = 0$) and induce displacements of test masses only in the transverse direction (from the deviation equation). They have helicities ± 2. Commonly, these are referred to as *gravitons* as they precisely correspond to the massless, helicity 2 irreducible representation of the Poincare group.

5.3.2 Gravitational Radiation

The plane wave solutions correspond to a source free case whereas waves due to a 'source' confined to a compact region of the space-time will typically be 'spherical' and will fall off with the distance from the source. We look at this case now.

The Green functions for the Minkowski d'Alembertian are well known and choosing the retarded Green function, the particular solution of the inhomogeneous equation is written down as,

$$\begin{aligned}
\tilde{h}_{\mu\nu}(t, \vec{x}) &= 4 \int_{\text{source}} d^3x' \frac{T_{\mu\nu}(t - |\vec{x} - \vec{x'}|, \vec{x'})}{|\vec{x} - \vec{x'}|}, \quad \text{or} && (5.62) \\
\tilde{h}_{\mu\nu}(t, \vec{x}) &:= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \tilde{h}_{\mu\nu}(\omega, \vec{x}) e^{i\omega t}, \\
T_{\mu\nu}(t, \vec{x}) &:= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega T_{\mu\nu}(\omega, \vec{x}) e^{i\omega t} \ , \\
\tilde{h}_{\mu\nu}(\omega, \vec{x}) &= 4 \int_{\text{source}} d^3x' \frac{T_{\mu\nu}(\omega, \vec{x'})}{|\vec{x} - \vec{x'}|} e^{-i\omega|\vec{x} - \vec{x'}|} && (5.63) \\
&\approx 4 \int_{\text{source}} d^3x' \frac{T_{\mu\nu}(\omega, \vec{x'})}{r(1 - \frac{\hat{r}\cdot\vec{x'}}{r})} e^{-i\omega r(1 - \frac{\hat{r}\cdot\vec{x'}}{r})}, \quad r := |\vec{x}| \gg |\vec{x'}|, \\
&\approx 4 \frac{e^{-i\omega r}}{r} \int_{\text{source}} d^3x' T_{\mu\nu}(\omega, \vec{x'}) \ + \ o(r^{-2}) && (5.64)
\end{aligned}$$

In the last but one line we have used the property that the source is confined to a compact spatial region and taken the observation point far away from the source. In the last line we have obtained a spherical wave with a source integral over a compact region. The solution of course has to satisfy the transversality condition.

The 4 equations of the transversality condition allows us to eliminate 4 components of the $\tilde{h}^{\mu\nu}$ in favour of $\tilde{h}^{ij}$.

$$\begin{aligned} \partial_\mu \tilde{h}^{\mu\nu} &= 0 & \Rightarrow \quad & \partial_0 \tilde{h}^{0\nu} + \partial_i \tilde{h}^{i\nu} = 0 \;\; \forall \;\; \nu & (5.65) \\ \therefore \partial_0 \tilde{h}^{00} &= -\partial_i \tilde{h}^{i0} & , \quad & \partial_0 \tilde{h}^{0j} = -\partial_i \tilde{h}^{ij} & \\ \therefore \tilde{h}^{00} &= -\frac{1}{\omega^2}\partial_i\partial_j \tilde{h}^{ij} & , \quad & \tilde{h}^{0i} = \frac{i}{\omega}\partial_j \tilde{h}^{ij} & (5.66) \end{aligned}$$

It therefore suffices to determine only $\tilde{h}^{ij}(\omega, \vec{x})$. To manipulate the source integral, we use the conservation of the stress tensor.

$$\begin{aligned} \partial_\mu T^{\mu\nu} &= 0 & \Rightarrow \quad & \partial_0 T^{0\nu} + \partial_i T^{i\nu} = 0 \;\; \forall \;\; \nu & (5.67) \\ \therefore \partial_0 T^{00} &= -\partial_i T^{i0} & \text{and} \quad & \partial_0 T^{0j} = -\partial_i T^{ij} & \\ \therefore T^{00} &= -\frac{1}{\omega^2}\partial_i\partial_j T^{ij} & \text{and} \quad & & (5.68) \\ -\omega^2 \int x^m x^n T^{00}(\omega, \vec{x}) &= & & \int x^m x^n \partial_i \partial_j T^{ij}(\omega, \vec{x}) & (5.69) \\ \Rightarrow \int_{source} T^{ij}(\omega, \vec{x}) &= & & -\frac{\omega^2}{2}\int_{source} x^i x^j T^{00} & (5.70) \\ \therefore \tilde{h}_{ij} &= -2\frac{e^{-i\omega r}}{r}\omega^2 q_{ij} & , \quad & q^{ij}(\omega) := \int_{source} x^i x^j T^{00}(\omega, \vec{x}) & (5.71) \end{aligned}$$

The q^{ij} is called the *quadrupole moment* of the source distribution. There are different conventions for its definition e.g. [8], we will continue with the definition above. In the equation (5.69), we have done partial integration and the integration is taken just out side the source so that the surface terms vanish.

Some remarks are in order. From the (5.68) equation, it follows immediately that the *double time derivative* of (remember the ω^2 factor) the 0^{th} moment of the energy density, $\int T^{00}$ and of the 1^{st} moment, $\int x^i T^{00}$, both vanish. The first is just the statement that if there is no net matter energy flux across the boundary of a region surrounding the source (the source is assumed to remain confined to a bounded region), then the total energy content of that region does not change. For the 1st moment ('mass dipole'), the first time derivative gives the 'momentum' of the source while the second time derivative vanishes due to momentum conservation (the total source momentum is conserved). Consequently, the first non-vanishing contribution to the radiation field comes from the quadrupole moment.

We have obtained the radiation field. We have to find out the energy, momentum, angular momentum etc. carried by the radiation field.

The metric tensor field or its linearized version are not like other fields whose energy contents can be related to their magnitudes or amplitudes. Thanks to the equivalence principle or the general covariance, any space-time metric or even a spatial metric can be made to vanish at any given point. Consequently, the energy content of the field cannot possibly be related to the value of the field at any given point and one does not expect to have a *local, tensorial* definition of gravitational energy.

Can we at least identify a suitable quantity which can be a satisfactory measure of energy, momentum, angular momentum etc. for the *linearized theory*? A natural way is to appeal to an action principle and deduce a conserved stress tensor via the Noether procedure. For this, we should expand the Einstein–Hilbert action to the quadratic order in $h_{\mu\nu}$. The Noether procedure will then give us a conserved quantity which is *quadratic* in $h_{\mu\nu}$. Here we will work at the level of Einstein equations themselves.

We have solutions of the linearized equation - both the homogeneous solution and the particular solution. Since the $h_{\mu\nu}$ is not just any symmetric tensor field in Minkowski space-time, but is part of the metric, we expect it to back react and modify the Minkowski background. We will consider a general decomposition of a metric into a background piece and a ripple, in the next subsection. Here we continue to take the background to be Minkowski metric. We can however incorporate the non-linearities in a *perturbative* manner. For this, let us substitute $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ in the exact equation and obtain the $h_{\mu\nu}$ in an iterated manner [17].

Substitution of $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ in the Einstein equation leads to a power series expansion in h which can be grouped in the form,

$$\begin{aligned} G^{(1)}_{\mu\nu} &:= R^{(1)}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}R^{(1)}_{\alpha\beta}\eta^{\alpha\beta} := 8\pi(T_{\mu\nu} + t_{\mu\nu}) =: 8\pi\tau_{\mu\nu} \quad \text{where} \\ t_{\mu\nu} &:= -\frac{1}{8\pi}\left[G_{\mu\nu} - G^{(1)}_{\mu\nu}\right] \quad \text{(`gravitational stress tensor');} \qquad (5.72) \end{aligned}$$

The above equations hold exactly but is to be solved in a perturbative manner.

Since the first order Einstein tensor is divergence-free, we get a conservation law and corresponding conserved quantities. Defining $\tau^{\mu\nu} := \eta^{\mu\alpha}\eta^{\nu\beta}\tau_{\mu\nu}$,

$$\partial_\mu \tau^{\mu\nu} = 0, \qquad \text{Conservation equation} \qquad (5.73)$$

$$P^\mu := \int_{\text{vol}} d^3x \tau^{0\mu}, \quad \text{Conserved energy momentum} \qquad (5.74)$$

$$\partial_\mu M^{\mu\nu\lambda} = 0, \quad M^{\mu\nu\lambda} := \tau^{\mu\lambda}x^\nu - \tau^{\mu\nu}x^\lambda \ ,$$

$$J^{\mu\nu} := \int_{\text{vol}} d^3x \tau^{0\mu\nu}, \quad \text{Conserved angular momentum} \qquad (5.75)$$

Note that the metric used for raising/lowering indices is the flat metric η.

It is an identity that,

$$G^{(1)}_{\mu\nu} = \partial_\lambda Q^\lambda{}_{\mu\nu}, \qquad \text{where ,} \qquad (5.76)$$

$$Q^{\lambda\mu\nu} := -\frac{1}{2}\left\{(\partial^\mu h\eta^{\lambda\nu} - \partial_\alpha h^{\alpha\mu}\eta^{\lambda\nu} + \partial^\lambda h^{\mu\nu}) - (\lambda \leftrightarrow \mu)\right\} \qquad (5.77)$$

$$\therefore \tau^{\mu\nu} = \frac{1}{8\pi}\partial_\lambda Q^{\lambda\mu\nu} \qquad \text{which implies,} \qquad (5.78)$$

$$P^\mu = \frac{1}{8\pi}\int_{\text{vol}} d^3x \partial_\lambda Q^{\lambda 0\nu} = \frac{1}{8\pi}\int_{\text{vol}} d^3x \partial_i Q^{i0\nu}$$

$$= \frac{1}{8\pi}\int_{\partial \text{vol}} d^2s\ n_i Q^{i0\mu} \tag{5.79}$$

The expression (5.77) does not look symmetric in $\mu\nu$, however it actually is, as can be checked by computing the divergence.

The last expression involves *only* the linearized gravitational field, that too only the linear power. As such this could be expected to be valid far away from *compact sources or isolated bodies*, where we physically expect the space-time metric to be approximately Minkowskian. Provided the asymptotic behaviour of the h field is such that *finite and non-zero* integrals result, we could interpret the P^μ as the energy-momentum of the matter plus gravity system. We will return to the issue of energy-momentum of matter plus gravity system in the context of asymptotically flat space-times in chapter 7.

For the plane gravitational waves discussed earlier, which corresponds to exact solution to the linearized, source-free equation, $\tau^{\mu\nu} = 0$ and we expect the above energy-momentum integral to be zero. It is easy to check that the Q vanishes for the plane wave solution using the TT-gauge and the equation of motion. If we do *not* take $h_{\mu\nu}$ to be a solution of the exact linearized equation, then the Q is non-zero and we can obtain total energy, momentum of the space-time from the equation (5.79), provided the solution has the appropriate asymptotic behaviour.

There are a couple of points to note. The conserved stress tensor $\tau_{\mu\nu}$ defined above is not generally covariant and is not uniquely defined. It is a Lorentz tensor though. This is to be expected since the split of the metric into the flat background and a small deviation is not generally covariant. Secondly, for an extended object like a wave, we need to *average* over a region larger than the wave length and average over a time larger than the period, when computing the energy-momentum carried by the wave.

In the next subsection, we perform a scale dependent split of the metric into a 'background' and a 'ripple'. Using a suitable averaging, a satisfactory computational definition of a gravitational stress 'tensor' is arrived at using which we obtain the quadrupole formula for radiated power.

5.3.3 Radiated Energy and the Quadrupole Formula

In any given coordinate system, say a lab frame, we can look for the spatial or temporal variation of the metric. Consider those metrics which have two well separated scales which could be spatial or temporal. Let L_B and f_B denote a length scale and a frequency scale. For such metrics, we can write the metric as a sum of a *background* metric and a *ripple* [27],

$$g_{\mu\nu}(x) = \bar{g}_{\mu\nu} + h_{\mu\nu}, \quad \text{such that}$$

(a) length scale of variation of $\bar{g} \gg L_B$ while that of $h \ll L_B$ *or* (b) frequency of time variation of the background metric is much smaller while that of the ripple is much larger *and* $|h_{\mu\nu}| \ll |\bar{g}_{\mu\nu}|$. These conditions justify h being referred to as a ripple[6]. To filter out the fast or short length scales, we perform the corresponding temporal or spatial average over slower or longer length scales. In either case, the average will be denoted by sandwiching the expression between angle brackets. For definiteness, we will do temporal averages. The background geometry has a low frequency variation while the ripple is of high frequency and low frequency part of an expression is obtained by a temporal averaging[7] i.e. $[\ ]_{\mathrm{low}} \Leftrightarrow \langle [\ ] \rangle$.

Expanding Einstein equation in powers of h, we write,

$$\begin{aligned} R_{\mu\nu}(g) \ &:= \ \bar{R}_{\mu\nu}(\bar{g}) + R^{(1)}_{\mu\nu}(\bar{g}, h) + R^{(2)}_{\mu\nu}(\bar{g}, h) + \ldots \\ &= \ 8\pi(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T_{\alpha\beta} g^{\alpha\beta}) \end{aligned} \tag{5.80}$$

Hence,

$$\bar{R}_{\mu\nu} + \left[R^{(2)}_{\mu\nu}\right]_{\mathrm{low}} \ = \ 8\pi \left[T_{\mu\nu} - \ \frac{1}{2} g_{\mu\nu} T_{\alpha\beta} g^{\alpha\beta}\right]_{\mathrm{low}} \tag{5.81}$$

$$R^{(1)}_{\mu\nu} + \left[R^{(2)}_{\mu\nu}\right]_{\mathrm{high}} \ = \ 8\pi \left[T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T_{\alpha\beta} g^{\alpha\beta}\right]_{\mathrm{high}} \tag{5.82}$$

The first equation can be written as

$$\bar{R}_{\mu\nu} \ = \ -\langle R^{(2)}_{\mu\nu} \rangle + 8\pi \left(\bar{T}_{\mu} - \frac{1}{2} \bar{g}_{\mu\nu} \bar{T}\right)$$

where we have denoted the averaged stress tensor and its trace by the overbars. Define the averaged quantity,

$$t_{\mu\nu} \ := \ -\frac{1}{8\pi} \langle R^{(2)}_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} R^{(2)}_{\alpha\beta} \bar{g}^{\alpha\beta} \rangle \ \Rightarrow \ t \ := \ \bar{g}^{\mu\nu} t_{\mu\nu} \ = \ \frac{1}{8\pi} \langle R^{(2)} \rangle. \tag{5.83}$$

Therefore $-\langle R^{(2)}_{\mu\nu} \rangle = 8\pi(t_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} t)$ and putting back in the low frequency equation we rearrange the equation as,

$$\bar{R}_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} \bar{R} \ = \ 8\pi(\bar{T}_{\mu\nu} + t_{\mu\nu}) \ =: \ 8\pi \tau_{\mu\nu} \tag{5.84}$$

Evidently, the $t_{\mu\nu}$ defined in equation (5.83), is our candidate *gravitational*

[6]The gravitational waves that could be detected by Earth based instruments correspond to a reduced wavelengths ranging from about 50 km to 500 km which correspond to a frequency range from 100 Hz to 1000 Hz. The 'size' of Newtonian potential is of the order of about $10^{-9} \sim 2GM/(c^2 R)$ while expected wave amplitude is only about 10^{-21}. Thus, the metric corresponding to the Newtonian gravity will play the role of the background metric and since it is essentially static, we can always satisfy the condition $f \gg f_B$.

[7]For $f(t)$ possessing a Fourier transform, $f(\omega)$, let its average be defined by $\langle f_T \rangle(t) := \frac{1}{2T} \int_{-T}^{T} dt' f(t+t')$. Then, $\langle f_T \rangle(\omega) = \frac{\sin\omega T}{\omega T} f(\omega)$. Thus, for *lower frequencies*, $\omega \ll T^{-1}$, the Fourier transforms are preserved by the averaging while for higher frequencies the averaged function has vanishing Fourier components.

stress tensor. It serves as a source for the background metric due to the ripple and is quadratic in the ripple, h. Although the ripple has high frequency modes, the quadratic expression contains slow modes as well due to 'beating' and are picked-up under averaging. The combined tensor, $\tau_{\mu\nu}$ is *covariantly* conserved. It is *not a tensor* since the identification of the ripple itself is frame dependent and there is an averaging involved. Nevertheless, it provides an algorithm to construct a gravitational stress 'tensor' which is *quadratic in the ripple, covariantly conserved and serves as a source for the background.*

Far away from isolated sources, on physical grounds, we assume the existence of a frame in which the above identification can be applied with the background metric approximating the Minkowskian metric. Then $\tau^{\mu\nu}$ reduces to $t^{\mu\nu}$, the covariant derivative reduces to an ordinary derivative and we obtain $\partial_\mu t^{\mu\nu} = 0$. Far away from the sources, we get,

$$\begin{aligned} R^{(2)}_{\mu\nu} &\approx \frac{1}{2}h^{\alpha\beta}\left\{\partial^2_{\mu\nu}h_{\alpha\beta} - \partial^2_{\nu\alpha}h_{\mu\beta} - \partial^2_{\mu\beta}h_{\nu\alpha} + \partial^2_{\alpha\beta}h_{\mu\nu}\right\} \\ &\quad -\frac{1}{4}\left\{2\partial_\alpha h^\alpha{}_\beta - \partial_\beta h^\alpha{}_\alpha\right\}\left\{\partial_\mu h^\beta{}_\nu + \partial_\nu h^\beta{}_\mu - \partial^\beta h_{\mu\nu}\right\} \\ &\quad +\frac{1}{4}\left\{\partial_\alpha h_{\beta\nu} + \partial_\nu h_{\alpha\beta} - \partial_\beta h_{\nu\alpha}\right\}\left\{\partial^\alpha h^\beta{}_\mu + \partial_\mu h^{\alpha\beta} - \partial^\beta h^\alpha{}_\mu\right\} \end{aligned} \tag{5.85}$$

Since we are far away from the source, we *can* choose the transverse, traceless gauge. We can do spatial average over a length scale much larger than the relevant wavelength. Under the integral sign, we can drop the 'boundary terms' which are down by inverse powers of the length scale. Using this spatial averaging, the TT-gauge and equation of motion in the asymptotic region, we get,

$$\langle R^{(2)}_{\mu\nu}\rangle = -\frac{1}{4}\langle\partial_\mu h_{\alpha\beta}\partial_\nu h^{\alpha\beta}\rangle_{TT} \quad \Longrightarrow \quad t_{\mu\nu} = \frac{1}{32\pi}\langle\partial_\mu h_{\alpha\beta}\partial_\nu h^{\alpha\beta}\rangle_{TT} \tag{5.86}$$

The trace of $t_{\mu\nu}$ vanishes due to the equation of motion and the TT gauge.

Note that in the TT gauge we can use the residual gauge freedom to set $h_{00} = 0 = h_{0i}$ which allows us to replace $h_{\alpha\beta}$ by h_{ij} in the sum in the above equation. For the plane gravitational wave solution, we note that

$$t_{\mu\nu} = \frac{1}{16\pi}k_\mu k_\nu\left(|\Phi_{mm}|^2 + |\Phi_{\bar{m}\bar{m}}|^2\right).$$

Using the conservation equation for this gravitational stress tensor, we derive the flux formula. Taking two concentric shells in the asymptotic region, we see that,

$$\int_{\rm vol}\partial_0 t^{00} = -\int_{\rm vol}\partial_i t^{i0} = -\left[-\int_{S_1}d\Omega\ r_1^2\hat{r}_i t^{i0} + \int_{S_2}d\Omega\ r_2^2\hat{r}_i t^{i0}\right] \tag{5.87}$$

This suggests that we define *flux* of gravitational energy through a surface as (n_i is the outward normal),

$$\text{flux}_S := \int_S ds\ n_i\ t^{i0}. \tag{5.88}$$

We can quickly see that the flux of the plane gravitational waves (solution of the homogeneous equation), through a sphere at large r, *vanishes*. What about the flux of radiation (solution of the inhomogeneous equation), given in equation (5.62)?

To the leading approximation for r large compared to the source size ($r \gg r'$), in the TT-gauge, the solution is of the form, $h^{TT}_{ij}(t,r) \sim r^{-1} f_{ij}(t-r)$. Clearly $\partial_r f_{ij} = -\partial_t f_{ij}$ and this leads to $\partial_r h^{TT}_{ij} \approx -\partial_0 h^{TT}_{ij} - o(r^{-1})$. Using this in the flux expression (5.88), we get,

$$\begin{aligned}
\frac{r^2}{32\pi}\int d\Omega \langle \partial^0 h^{TT}_{ij} \partial^r h^{ij}_{TT}\rangle &= \frac{\partial E}{\partial t} \qquad (5.89)\\
&= -\frac{r^2}{32\pi}\int d\Omega \langle \partial^0 h^{TT}_{ij} \partial_0 h^{ij}_{TT}\rangle \\
\therefore \frac{1}{r^2}\frac{dE}{dt d\Omega} &= \frac{1}{32\pi}\langle \dot{h}^{TT}_{ij} \dot{h}^{ij}_{TT}\rangle \quad \text{(Radiated power)} \quad (5.90)
\end{aligned}$$

and this is non-zero as long as in the TT-gauge, the ripple has time-dependence.

In equation (5.71), we have obtained the leading order solution related to the quadrupole moment of the source without imposing the TT-gauge. The formula for radiated power however uses h field in the TT-gauge. We can extract the TT-gauge parts of $\tilde{h}_{ij}$ as follows. For a wave going in the $\hat{n}$ direction, introduce a projection operator, $P_{ij}(\hat{n}) := \delta_{ij} - n_i n_j$. This projects onto the transverse plane. Define, $\Lambda_{ij,kl}(\hat{n}) := P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}$. This is also a projection operator: $\Lambda_{ij,kl}\Lambda_{kl,mn} = \Lambda_{ij,mn}$ and contraction with n in any of the indices, gives zero. It is also traceless in each pair of indices (ij) and (kl). For any tensor $X_{mn}(\hat{n})$, we can get its TT-projection as, $X^{TT}_{ij}(\hat{n}) = \Lambda_{ij,kl}X_{kl}(\hat{n})$. Noting that in the TT-gauge, $\tilde{h}_{ij} = h_{ij}$, we get,

$$\begin{aligned}
h^{TT}_{ij}(t,r) &= \Lambda_{ij,kl}\int \frac{d\omega}{\sqrt{2\pi}} h_{kl}(\omega,r)e^{i\omega t} \\
&= \Lambda_{ij,kl}\int \frac{d\omega}{\sqrt{2\pi}}\left(-\frac{2}{r}e^{-i\omega r}\omega^2 q_{kl}(\omega)\right)e^{i\omega t)} \\
&= \frac{2}{r}\Lambda_{ij,kl}\int \frac{d\omega}{\sqrt{2\pi}}\frac{\partial^2}{\partial t^2}e^{i\omega(t-r)}q_{kl}(\omega) \\
\therefore h^{TT}_{ij}(t,r) &= \frac{2}{r}\Lambda_{ij,kl}\ddot{Q}_{kl}(t-r), \text{where,} \\
Q_{ij} &:= q_{ij} - \frac{1}{3}\delta_{ij}q_{mn}\delta^{mn} \qquad (5.91)
\end{aligned}$$

With these, we relate the radiated power to the source quadrupole (in the leading order in $1/r$) as,

$$\frac{d^2E}{d\Omega dt} = r^2\frac{1}{32\pi}\frac{4}{r^2}\Lambda_{ij,kl}\Lambda_{ij,mn}\langle \dddot{Q}_{kl}\dddot{Q}_{mn}\rangle \qquad \text{and}$$

$$P_{\text{total}}(t,r) = \left[\frac{1}{8\pi}\int d\Omega\Lambda_{kl,mn}(\hat{r})\right]\langle \dddot{Q}_{kl}\dddot{Q}_{mn}\rangle(t-r) \tag{5.92}$$
$$= \frac{1}{8\pi}\frac{2\pi}{15}\left(11\delta_{ik}\delta_{jl} - 4\delta_{ij}\delta_{kl} + \delta_{il}\delta_{jk}\right)\langle \dddot{Q}_{ij}\dddot{Q}_{ij}\rangle$$
$$\therefore P_{\text{quadrupole}} = \left(\frac{G}{c^5}\right)\frac{1}{5}\langle \dddot{Q}_{ij}\dddot{Q}_{kl}\rangle(t-r) \tag{5.93}$$

This is the famous Einstein *quadrupole formula.* We note, without proof, some special cases which provide a feeling for orders of magnitude and are practically useful.

Rotating rod: For a rod of mass M, length L and rotating about its mid-point with a angular frequency of Ω, the quadrupole formula becomes,

$$P_{\text{quadrupole}} = \frac{2}{45}\frac{G}{c^5}M^2\ L^4\ \Omega^6 \quad \sim \quad 1.2\times 10^{-54}\ M^2L^4\Omega^6 \quad \text{Watts} \tag{5.94}$$

Binary stars: For a binary system of stars or stellar mass black holes, $L \sim 10^{12}$ meters (about 10 Earth-Sun distance), $M \sim 10^{30}$ kg and angular speed from Kepler's law is $\Omega \sim 10^{-8}$ which gives the power to be about 10^{-4} Watts. A tighter binary with compact bodies could have $L \sim 10^9$ meters with angular speed of about 10^{-4} and radiate gravitational energy at 10^{18} watts. A millisecond pulsar would have $M \sim 10^{30},\ L \sim 10^4,\ \Omega \sim 10^4$ leading to power rate of 10^{46} watts. Actually, this is an over-estimate. Although the neutron star radius is about 10^4 meters, it also almost spherical and hence would have very small quadrupole moment e.g. due to some irregularity on the surface. Taking into account this factor, the effective length scale for a neutron star could be reduced by about a factor of 10^{-3} or even smaller. This would reduce the power by a factor of at least a million [28].

Indirect evidence: A binary system will shrink in size due to loss of energy and speed-up its orbital period which is detectable. The rate of decrease of radius per revolution can be deduced from the power.

$$\Delta R \sim -R\frac{\Delta E}{E} = -\frac{R}{E}\frac{2\pi P}{\Omega} \sim \frac{2R^2}{GM^2}\frac{P}{\Omega} \sim 10^{-43}R^6\Omega^5 \text{ meters.}$$

This rate of decrease of radius is correlated to decrease in the period by, $\Delta T/T = (3/2)(\Delta R/R)$. The relevant parameters for the Hulse–Taylor binary pulsar (PSR B1913+16) are: $R \sim 10^9,\ M \sim 10^{30},\ T \sim 10^4$ in MKS units. This gives $\Delta T \sim 10^{-14}$. There are many corrections that need to be applied [28] after which it increases by two orders of magnitude. The observation of the orbital decay of the Hulse–Taylor binary pulsar (PSR B1913+16) over a 30-year period (1975–2005) matches with the loss by the quadrupole formula to within 1/2 percent [19].

Direct detection efforts: Direct detection methods do not rely on the energy loss but on the amplitude of the waves which causes tidal distortions. There is a large scale world-wide effort. It is discussed briefly in the section 10.2.

5.4 Black Holes—Elementary Aspects

5.4.1 Static Black Holes

We return to the implication of general relativity in the context of compact objects. We will take a first look at the vacuum solutions representing compact objects, namely, the black hole solutions, study their basic properties and briefly look at their observational status. We will consider them in more generality in the part II. We will also consider 'stars' (non-vacuum solutions) to see what general relativity has to say about them.

5.4.1.1 Schwarzschild Black Hole

We looked at the exterior Schwarzschild solution in the previous sections. Imagine now that the gravitational collapse has proceeded so far that the 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 $R_\odot \le r \le R_S$. Here we meet the famous Schwarzschild singularity that caused confusion in the early history. Quite 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. How do we see this?

Consider for instance 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 $r > 0, 0 \le \theta < 2\pi$, 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 is a Killing vector field, then taking the parameter λ itself as one of the local coordinates ensures that the metric is 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 (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 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,

$$\begin{aligned} ds^2 &= \left(1 - \frac{R_S}{r}\right)\left\{-dt^2 + \left(1 - \frac{R_S}{r}\right)^{-2} dr^2\right\} + r^2 d\Omega^2 \\ &:= \left(1 - \frac{R_S}{r}\right)\left\{-dt^2 + dr_*^2\right\} - r^2 d\Omega^2 \end{aligned} \tag{5.95}$$

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

$$r_*(r) = r + R_S \ell n \left|\frac{r - R_S}{R_S}\right| \tag{5.96}$$

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

$$\begin{aligned} t &:= \frac{1}{2}(\epsilon_u u + \epsilon_v v), \quad r_* := \frac{1}{2}(-\epsilon_u u + \epsilon_v v), \quad \epsilon_u, \epsilon_v = \pm 1, \\ u &= \epsilon_u(t - r_*), \quad v = \epsilon_v(t + r_*) \end{aligned} \tag{5.97}$$

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 to 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 $\epsilon_u = -1$). Taking $\epsilon_u = 1$ for definiteness and substituting for r_* one sees that,

$$\left(1 - \frac{R_S}{r}\right) = \frac{R_S}{r} e^{-r/R_S} e^{(v-u)/(2R_S)} \tag{5.98}$$

$$\begin{aligned} ds^2 &= -\frac{R_S}{r} e^{-r/R_S}\left(e^{-u/(2R_S)} du\right)\left(e^{v/(2R_S)} dv\right) + r^2 d\Omega^2 \\ &= -\frac{4R_S^3}{r} e^{-r/R_S} dU dV + r^2 d\Omega^2, \quad \text{with} \end{aligned} \tag{5.99}$$

$$U := -e^{-u/(2R_S)} \qquad := \quad T - X$$

$$V \;:=\; e^{v/(2R_S)} \;:=\; T + X \tag{5.100}$$

$$-UV \;=\; \left(\frac{r}{R_S} - 1\right) e^{r/R_S} \;=\; X^2 - T^2 \tag{5.101}$$

The coordinates T, X defined in (5.100) are known as the Kruskal coordinates. Their relation to the Schwarzschild coordinates (t, r) is summarized below.

$$\begin{aligned} F(r) &= X^2 - T^2 \;:=\; \left(\frac{r}{R_S} - 1\right) e^{r/R_S} \\ \frac{t}{R_S} &= 2\; tanh^{-1}\left(\frac{T}{X}\right) \end{aligned} \tag{5.102}$$

$$\begin{aligned} X &= \pm\sqrt{|F(r)|}\; cosh\left(\frac{t}{R_S}\right) \\ T &= \pm\sqrt{|F(r)|}\; sinh\left(\frac{t}{R_S}\right) \end{aligned} \tag{5.103}$$

$$ds^2 = \frac{4\; R_S^3\; e^{-r/R_S}}{r}\left(-dT^2 + dX^2\right) + r^2(T, X)\; d\Omega^2 \tag{5.104}$$

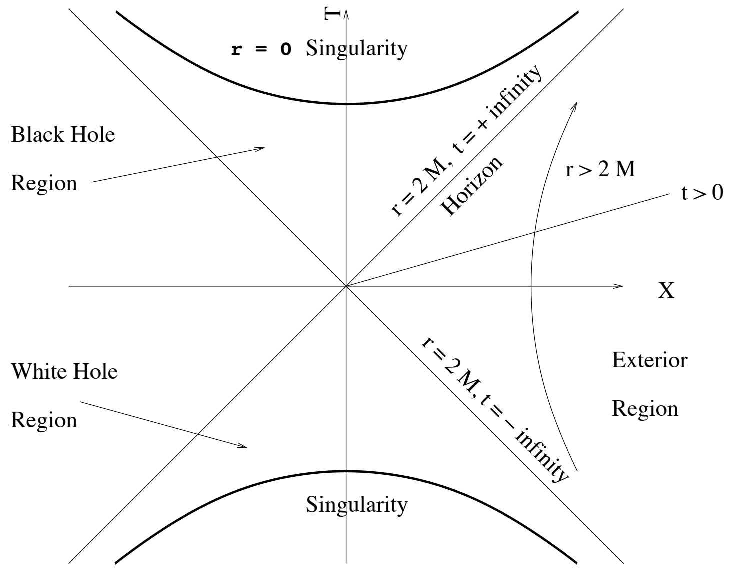


FIGURE 5.3: Kruskal diagram for the Schwarzschild space-time.

Looking at the figure (5.3) 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 labelled 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, there can be none *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 the existence of an event horizon is the distinguishing and (defining) property of a black hole. For the corresponding *Penrose Diagram*, see the figure 8.2.

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 a 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}Hz, to see it as yellow color light of frequency of about 3×10^{15}Hz, 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 these too are mathematically peculiar.

5.4.1.2 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 (2.24). In addition, we need an ansatz for the electromagnetic field. It is straightforward to show that spherical symmetry and staticity implies that the only non-vanishing components of $F_{\mu\nu}$ are,

$$F_{tr} = \xi(r), \quad F_{\theta\phi} = \eta(r)\sin\theta. \tag{5.105}$$

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 (4.13). For notational convenience, we divide by the extra factor of 4π to avoid factors of 8π in the metric.

$$T_{\mu\nu} = \frac{1}{4\pi}\left[F_{\mu\alpha}F_{\nu\beta}g^{\alpha\beta} - \frac{1}{4}g_{\mu\nu}\left(F_{\alpha\rho}F_{\beta\sigma}g^{\alpha\beta}g^{\rho\sigma}\right)\right]. \tag{5.106}$$

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

$$\begin{aligned} T_{tt} &= \frac{1}{8\pi}\frac{Q^2}{r^4}f(r), \quad Q^2 := Q_e^2 + Q_m^2 \\ T_{rr} &= -\frac{1}{8\pi}\frac{Q^2}{r^4}g(r) \\ T_{\theta\theta} &= \frac{1}{8\pi}\frac{Q^2}{r^2}, \quad T_{\phi\phi} = \sin^2\theta\, T_{\theta\theta} \end{aligned} \tag{5.107}$$

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 (5.3, 5.107), it is straight forward to obtain the Reissner–Nordstrom solution:

$$\begin{aligned} f(r) &= \frac{\Delta(r)}{r^2}, \quad g(r) = f^{-1}(r) \\ F_{tr} &= \frac{Q_e}{r}, \quad F_{\theta\phi} = Q_g \sin\theta \\ \Delta(r) &:= r^2 - 2Mr + Q^2 \ , \quad M,\ Q \text{ are constants,} \end{aligned} \tag{5.108}$$

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,

$$r_*(r) \ = \ r + \frac{r_+^2}{r_+ - r_-}\ell n\left|\frac{r - r_+}{r_+}\right| - \frac{r_-^2}{r_+ - r_-}\ell n\left|\frac{r - r_-}{r_-}\right|. \tag{5.109}$$

There are now *three* regions to be considered:

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

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' them at the chart boundaries $r_\pm$. The corresponding *Penrose Diagram* can be obtained by specializing the figure 8.1 and is discussed in section (8.1).

5.4.2 Stationary (Non-Static) Black Holes

5.4.2.1 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 title of '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 [17, 29]:

$$ds^2 \ = \ -\frac{\eta^2\Delta}{\Sigma^2}dt^2 + \frac{\Sigma^2 sin^2\theta}{\eta^2}(d\phi - \omega dt)^2 + \frac{\eta^2}{\Delta}dr^2 + \eta^2 d\theta^2 \qquad \text{where,} \tag{5.110}$$

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

$a = 0$	,	$Q = 0$	:	Schwarzschild solution
$a = 0$	,	$Q \neq 0$	:	Reissner–Nordstrom solution
$a \neq 0$	,	$Q = 0$	:	Kerr solution

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:

$$A_{r_+} := \int_{r_+} \sqrt{det(g_{ind})} d\theta d\phi \ = \ \sqrt{\Sigma^2} \int sin\theta d\theta d\phi \ = \ 4\pi(r_+^2 + a^2) \quad (5.111)$$

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, *for* $a = 0$,

$$\text{Surface Gravity, } \kappa \ := \ -\frac{1}{2}\frac{dg_{tt}}{dr}|_{r=r_+} \ = \ \frac{r_+ - M}{r_+^2} \ = \ \frac{r_+ - M}{2Mr_+ - Q^2} \quad (5.112)$$

Although for rotating black holes 'surface gravity' cannot be defined so simply, it turns out that when appropriately defined (see equation (8.8) it is still given by the last equality in the above expression.

There is one more quantity associated with the event horizon of a rotating black hole – the angular velocity of the horizon, Ω. 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 $(norms)^2$'s 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 *at the horizon* 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,

$$\Omega := \frac{a}{r_+^2 + a^2}. \quad (5.113)$$

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

$$\Phi := \frac{Q\, r_+}{r_+^2 + a^2} \quad (5.114)$$

The parameters of the solution are identified as: Q is charge and $J := Ma$ is the angular momentum.

Thus we have defined:

$$\begin{array}{lllllll} M & = & M & ; & r_+ & = & M + \sqrt{M - a^2 - Q^2} \\ A & = & 4\pi(r_+^2 \ + \ a^2) & ; & \kappa & = & \frac{r_+ \ - \ M}{2Mr_+ \ - \ Q^2} \\ J & = & Ma & ; & \Omega & = & \frac{a}{r_+^2 \ + \ a^2} \\ Q & = & Q & ; & \Phi & = & \frac{Qr+}{r_+^2 \ + \ a^2} \end{array} \quad (5.115)$$

Now one can verify explicitly that,

$$\delta M \ = \ \frac{\kappa}{8\pi}\delta A \ + \ \Omega\delta J \ + \ \Phi\delta Q \quad (5.116)$$

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. We will return to general definition of black holes and its generalization in section 8.2. In the next section we take a brief look at the observational status of black holes.

5.4.3 Observational Status

Theoretically, black hole *solutions* exist for all values of mass, angular momentum and charge subject to $M^2 > a^2 + Q^2$. Their physical realization however depends on astrophysical gravitational collapse. As per current understanding, beyond a neutron star, there does not seem to be a stable compact object i.e. if a neutron star crosses the Chandrasekhar limit by an accretion process, then a complete gravitational collapse is un-stoppable with an end-result most likely being a black hole (size smaller than its Schwarzschild radius, say) or perhaps a 'naked singularity'. It is this feature that lends support to physical realizations of black holes as well as the expectation that the astrophysical black holes have masses in excess of the solar mass.

There is also the possibility of that a black hole may keep growing by swallowing nearby stars or other black holes. It is also conceivable that a super-massive star may collapse directly to a super-massive black hole without passing through a supernova stage in which a lot of parent stellar material is exploded away. The star gobbling possibility is likely to occur in the central regions of most galaxies. These centres could then be super-massive black holes.

A black hole being black, is hard to detect. It is through its accretion disc which glows, that one looks for evidence for a black hole. A solar scale black hole is a few km in size and has strong tidal forces at the horizon. Any matter reaching nearby can be torn away producing a glowing accretion disc. Super-massive black holes by contrast, have a smaller curvature and weaker tidal disruption which cannot sustain a glowing accretion disc. The detection methods of these two types of black holes are different.

The stellar scale black holes are searched in binary systems with a visible star and its invisible companion. The optically invisible companion can be bright in the x-ray region produced by the heated matter drawn from the optically bright star. Depending upon the type of accretion disk, evidence for a black hole horizon is inferred from the luminosity of the 'reflected' energy from the compact object [30].

The super-massive black holes are indicated by the jets of matter and/or plasma, emanating from the central regions. Such jets can be sustained only by a powerful central engine, the best candidate for which are rotating, super-massive black holes. Another method of confirming the black holes in the centres of nearby galaxies, including our own is, direct, high resolution millimeter wavelength observation. For the Milky Way, Keplerian orbits of stars in the central region already indicate a 4 million solar mass object confined to a small region consistent with the horizon size. The Event Horizon telescope system has already provided evidence for the central black hole.

So far, there have been candidate black holes in the mass range of few tens to few hundreds of solar masses. There are also candidates of the super-massive class, with masses ranging from a million to a few billion solar masses. There are no candidates of intermediate mass range. Further details may be seen in [31].

5.5 Stars in GR

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:

$$\begin{array}{lllllll} e_0^\mu & := & \frac{1}{\sqrt{f}}(1,0,0,0) & , & e_1^\mu & := & \frac{1}{\sqrt{g}}(0,1,0,0) \\ e_2^\mu & := & \frac{1}{r}(0,0,1,0) & , & e_3^\mu & := & \frac{1}{rsin\theta}(0,0,0,1) \end{array} \qquad (5.117)$$

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, for 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 which is first order. Thus we can arrange our equations as 3 first order equations [32]:

$$r\frac{dg}{dr} = -g(g-1) + (8\pi\rho_0 r^2)g^2 \qquad (G_{00} = 8\pi T_{00}) \tag{5.118}$$

$$r\frac{df}{dr} = f(g-1) + (8\pi\rho_1 r^2)fg \qquad (G_{11} = 8\pi T_{11}) \tag{5.119}$$

$$r\frac{d\rho_1}{dr} = 2(\rho_2 - \rho_1) - \frac{\rho_0 + \rho_1}{2} r\frac{d\ell n f}{dr} \quad \text{(Conservation equation)} \tag{5.120}$$

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

$$m(r) - m(r_1) := 4\pi \int_{r_1}^{r} \rho_0(r')r'^2 dr', \qquad g(r) := \left(1 - \frac{2m(r)}{r}\right)^{-1} \tag{5.121}$$

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 this 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 $m(r_1) \equiv M$. The solution can be constructed easily and is also a black hole solution.)

With these we can write the final equations as:

$$m(r) := 4\pi \int_0^r \rho(r')r'^2 dr', \quad g(r) := \left(1 - \frac{2m(r)}{r}\right)^{-1} \tag{5.122}$$

$$\frac{dP(\rho(r))}{dr} = -\left[\frac{m(r)\rho}{r^2}\right]\left(1 + \frac{P(\rho)}{\rho}\right)\left[\frac{1 + \frac{4\pi r^3}{m(r)}P(\rho)}{1 - \frac{2m(r)}{r}}\right] \tag{5.123}$$

$$r\frac{d\ell n f}{dr} = 2\frac{m(r) + 4\pi P(\rho)r^3}{r - 2m(r)} \tag{5.124}$$

The middle equation (5.123) is the 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, the 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 a star quite non-trivial.

An instructive example which can be solved 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,

$$P(r) = \hat{\rho}\left[\frac{(1-2M/R)^{1/2} - (1-2Mr^2/R^3)^{1/2}}{(1-2Mr^2/R^3)^{1/2} - 3(1-2M/R)^{1/2}}\right] \tag{5.125}$$

$$P(0) = \hat{\rho}\left[\frac{(1-2M/R)^{1/2} - 1}{1-3(1-2M/R)^{1/2}}\right] \tag{5.126}$$

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 [17].

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 possibly 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). It turns out [2] that for the so called *Newtonian Polytropes* i.e. stars with equation of state of the form $P \propto \rho^\gamma$ and governed by the Newtonian equations for the hydrostatic equilibrium are stable for all values of the central density for $\gamma > 4/3$ and unstable for $\gamma < 4/3$. Applied to *white dwarfs* where the pressure is generated by the *electron degeneracy pressure*, the instability value of 4/3 is reached for an upper limit of mass, the *Chandrasekhar Limit* of about 1.4 times the solar mass. The corresponding radius of a white dwarf is about 4000 km. The general relativistic effects can continue to be neglected. Applied to neutron stars, with the pressure now being supplied by the *neutron degeneracy pressure*, there is a similar upper mass limit from stability and it goes up to about 2.5 solar mass with a size of about 10 km. At this stage, general relativistic effects begin to make contributions, but are still small. Thus, for real (stable) stars in the astrophysical context, general relativity does not play a significant role. It does provide the condition that physical radius of body must be larger than (9/8) times its Schwarzschild radius. If these upper limits on mass (or lower limits on size) are crossed, black hole (or a naked singularity) formation is unavoidable.

Part II

The Beyond

Chapter 6

The Space-Time Arena

We have seen several examples of space-times including those which are also solutions of the Einstein equation. Most of these were *local solutions* but we also saw *extended solutions* in the examples of the static black holes. The basic idea of an extension is to embed a given space-time, (M, g), into another one, $(\bar{M}, \bar{g})$, such that on $M \subset \bar{M}$, we have $g = \bar{g}$. The two space-times may be just smooth, or real analytic, or solutions of the Einstein equation. If no such extension is possible, the space-time is said to be *in-extendable*. The focus of this chapter is on in-extendable space-times. Eventually we would like such space-times as solutions of the Einstein equation with suitable matter stress tensor, but to begin with we just focus on *candidate* space-times.

Space-times are distinguished from Riemannian spaces by their Lorentzian nature which specifies a *local* notion of causality given by the local light cones. Causality and determinism are two basic features of predictability that a space-time allows us to formulate. It is therefore necessary that our candidate space-times should be *causally well behaved, deterministic and predictable.* These are loaded terms which need to be sharpened. We do this in stages.

The absolutely basic prerequisite for any notion of causality (as distinct from correlations) is a distinction between past and future. This is captured by the formulation of *time orientability.* The next feature is a possible mechanism of causation which requires the possibility of communication and is captured by positing that two events can be causally connected if there is a curve connecting them which is future (or past) directed and everywhere non-space-like. While locally, the division between time-like and space-like intervals is clear cut, at the global level it leaves open the possibility of *closed time-like or causal curves.* This is a causal pathology and is excluded by introducing the *chronology or causality condition.* While this condition excludes closed time-like or causal curves, it leaves open the possibility of a causal curve returning repeatedly arbitrarily close to a given event. This could jeopardise the identification of a cause for a given effect due to finite precision. This is prevented by defining *strong causality.* Causality is conditional on a given space-time. But the metric itself could be known to some precision and it is possible that a metric is strongly causal but the a nearby one has 'wider light cones' and therefore a closed curve that was not causal with respect to the first one can be causal, thus confusing cause and effect again. This is precluded by defining *stable causality.* This finally prevents all possible causal pathologies.

Next is an independent notion of determinism. Wave phenomena which involve partial differential equations with time evolution. Typically, the initial data for these is specified on some 'constant time' hypersurface and the equation is supposed to determine future data (and also the past) and this should be the *complete information at all space-time events*. Such space-times are distinguished as being *globally hyperbolic*. This is a pre-condition for an evolution equation to be deterministic.

Predictability is a further independent feature which is born *after* determinism is admissible and is really a property of particular deterministic evolution equations. This has to do with how sensitively 'future' data depends on 'initial' data and as such is independent of the particular globally hyperbolic space-time.

We will now take a closer look at various definitions and relevant theorems. This chapter is more mathematical in nature because of the generality of its scope and so is the presentation. The main references followed are the books by Hawking and Ellis [18] and by Wald [17]. For additional examples and information, please see [33, 34].

In the following, (M, g) is an in-extendable Lorentzian manifold of dimensions ≥ 2.

6.1 Preliminary Notions and Results

First, not all manifolds can admit a smooth Lorentzian signature metric!

Theorem 6.1 (Existence of Lorentzian Metric) *A manifold M, admits a Lorentzian metric iff either (i) M is non-compact, or (ii) if M is compact, then its Euler character is zero.*

The proof may be seen in [18, 35].

We now have the notion of 'time', how do we distinguish future and past? We know how to do so in the Minkowski space-time. We have already discussed the division of intervals into space-like, time-like and light-like. The linear structure of the Minkowski space-time allows us to treat coordinate intervals as vectors. From the properties of the Lorentz transformations, we also know that for a time-like or a light-like vector, the *sign of its time component* is unchanged by infinitesimal Lorentz transformations. On the set of time-like vectors, define a relation, $X \sim Y$ iff $X \cdot Y := \eta_{IJ} X^I Y^J < 0$. This is an equivalence relation and for members in the same equivalence class we have $X^0 Y^0 > 0$. There are precisely, *two* equivalence classes, which may be labelled as $[X]$ and $[-X]$, for some time-like vector X. We arbitrarily designate one of these as future and the other as past, of the origin $(X = 0)$ of the Minkowski space-time. We *extend* these classes to include light-like vectors by stipulating that a *light-like vector*, Y, belongs to the future (past) if $Y \cdot X < 0$ for some X

in the future (past). The future (past) *including the light-like vectors* is referred to as the future (past) *light cone* at $X = 0$. Thus in Minkowski space-time, we have a definition of future (past) in terms of the light cones. The tangent space T_p at any event p in a generic space-time, is naturally a Minkowski space-time and hence has a notion of future and past light cones of the origin in T_p. By a slight abuse of language we will refer to these light cones also as future (past) *light cones at* p.

We transport these notions to a manifold by means of *convex normal neighborhoods*. Recall that on a (pseudo)Riemannian manifold which has a built-in definition of geodesics, we can assign coordinates to points q in a neighborhood U_p of any point p, by finding a geodesic starting from p with a tangent vector X_q and connecting to the point q in a unit affine parameter distance. This defines the exponential map from a neighborhood V_0 of the origin in the tangent space T_p, to a neighborhood U_p in the manifold. The coordinates of q are just the components of the tangent vector X_q and are called *Riemann normal coordinates* while the neighborhood of p is called a Riemann normal neighborhood.

Definition 6.1 (Convex Normal neighborhood:) *A Riemann normal neighborhood is* convex *if for all points q, r in the neighborhood, there is a* unique geodesic *connecting them and lying entirely in the neighborhood.*

These are open sets in which, none of the geodesics have any intermediate points missing. It is a further result that *convex normal neighborhoods always exist on pseudo–Riemannian manifolds*. This result enables us to import the future/past structure from the Minkowski space-time to the general space-time, albeit only *locally*.

Definition 6.2 (Time-Orientability:) *A space-time is said to be time-orientable if the future/past assignment can be done consistently everywhere.*

Theorem 6.2 (M, g) *is time-orientable iff there exists a smooth time-like vector field.*

Such a vector field is not unique and choosing a particular one fixes a time-orientation. We will assume that a choice has been made.

Definition 6.3 (Time-Like/Causal Curves:) *A smooth curve is said to be* future (past) directed time-like (causal) curve *if at every point p on the curve, the tangent vector is future (past) directed time-like (causal) vector.*

This allows us to extend the notion of the Minkowski light cone to *subsets* of general time-oriented space-times.

Definition 6.4 (Chronological/Causal Future/Past:) Chronological future (past) of p *is defined as,*

$I^{\pm}(p) := \{q \in M \ / \ \exists$ *a future (past) directed time-like curve from* p *to* $q\}$.
Likewise Causal future (past) of p *is defined as,*
$J^{\pm}(p) := \{q \in M \ / \ \exists$ *a future (past) directed causal curve from* p *to* $q\}$.

In the above, trivial curves ($\lambda(t) = p \ \forall \ t$) are excluded, they have no time-like/light-like/space-like attribute.

The extension to future/past of any subset $S \subset M$ is done by taking union of future/pasts of points of the subset S e.g. $I^{\pm}(S) := \cup_{p \in S} I^{\pm}(p)$.

The Minkowski light cones have further properties e.g., $I^{+}(p)$ consists of *exactly all* the points which lie on future directed time-like *geodesics* and the *boundary*[1], $\dot{I}^{+}(p)$, of $I^{+}(p)$ is generated by *null geodesics.* This means that $J^{+}(p) = I^{+}(p) \cup \dot{I}^{+}$. These properties are *not* true in general space-times. For example, in a space-time obtained from the Minkowski space-time by removing a point r from the light cone based on p, any point q 'beyond' r will no longer belong to $J^{+}(p)$ although it will still belong to the boundary $\dot{I}^{+}$. Clearly, $\dot{I}^{+}(p) \neq J^{+}(p) - I^{+}(p)$. However, locally, *in a convex normal neighborhood*, the Minkowski properties hold. Let N_p be a convex normal neighborhood of p. Then,

Theorem 6.3 (in a Convex Normal neighborhood)

1. $I^{+}(p) \cap N_p$ *is the set of points* $q \in M$ *such that* q *can be connected to* p *by a future directed time-like geodesic contained within* N_p*, and*
2. $\dot{I}^{+}(p) \cap N_p$ *is generated by future directed null geodesics from* p.

This leads to the corollaries,

1. $I^{+}(p)$ is open. In contrast, $J^{+}(p)$ is *not* open.
2. If $q \in N_p$ satisfies $q \in J^{+}(p)$, $q \notin I^{+}(p)$, then q lies on a null geodesic from p i.e. $J^{+}(p) \cap N_p = \overline{I^{+}(p)} \cap N_p$ which is equivalent to saying that within N_p, $J^{+}(p) = I^{+}(p) \cup \dot{I}^{+}(p)$.
3. If, $q \in J^{+}(p) - I^{+}(p)$, then any causal curve connecting p to q must be null geodesic.

The first corollary is equivalent to the assertion that the tip q, of a time-like curve can be deformed to range over a neighborhood, $q' \in u_q$, without changing its time-like character. This also implies that if $q \in I^{+}(p)$ then $I^{+}(q) \subset I^{+}(p)$.

Theorem 6.4 (Properties of Future/Past of a Subset:)
For $S \subset M$*,*

[1]Boundary of a subset $A \subset X$ is the set of all points p of X such that every neighborhood of p has a non-empty intersection with Ai, and with $X - A$.

1. $I^+(I^+(S)) = I^+(S)$;
2. $I^+(\bar{S}) = I^+(S)$;
3. $J^+(S) \subset \overline{I^+(S)}$;
4. $I^+(S) = int[J^+(S)]$;
5. $\dot{I}^+(S) = \dot{J}^+(S)$.

The first result really expresses the 'transitivity' - $r \in I^-(q), q \in I^-(p) \Rightarrow r \in I^-(p)$.

Definition 6.5 (Achronal Sets) *A subset $S \subset M$ is said to be* achronal *if for every $p, q \in S$, $p \notin I^+(q)$, and $q \notin I^+(p)$; equivalently, $I^+(S) \cap S = \emptyset$;*

Note that a space-like hypersurface is achronal but the converse is not true. For instance, if S is the 'surface' of a Minkowski light-cone in two-dimensions (i.e. the 45 degree pair of null lines), its chronological future is the interior of the light-cone and clearly $I^+(S) \cap S = \emptyset$. An achronal set in general is made up of portions which are space-like, or null, or isolated boundary points. We have,

Theorem 6.5 *For any $S \subset M$,* either $\dot{I}^+(S) = \emptyset$ or *is an achronal, three-dimensional, embedded, C^0 submanifold of M.*

This establishes that the boundary of the future of any subset, if non-empty, is 'well behaved' at least in a C^0 sense.

It turns out that we need to be able to extend the notions of time-like or causal curves also to curves which are only *continuous*. Since a continuous curve is does not have a tangent, its causal attribute is assigned indirectly.

Definition 6.6 (Continuous Time-Like/Causal Curves)

A continuous curve $\lambda(t)$ is said to be future directed time-like (causal) if for every $p \in \lambda(t)$, $\exists$ a convex normal neighborhood, N_p such that if $\lambda(t_1),\ \lambda(t_2) \in N_p$ with $t_1 < t_2$, then $\exists$ a future directed differentiable time-like (causal) curve from $\lambda(t_1)$ to $\lambda(t_2)$.

We will also need the notion of *extendibility* for curves. For this we have to introduce,

Definition 6.7 (End Point) *$p \in M$ is said to be a future end point of $\lambda(t)$ if for every neighborhood, u_p, $\exists\ t_0$ such that $\lambda(t) \in u_p\ \forall\ t > t_0$;*

$\lambda(t)$ is future in-extendible if it does not have a future end point.

Note that end points (future and/or past) need not exist. However if they exist, they are unique. It is also possible that an end point exists but does not belong to the curve i.e. $\nexists$ a t' such that $p = \lambda(t')$. If however, an end point exists and belongs to the curve, then it is possible to *extend* the curve by *adjoining* another time-like (causal) appropriately directed curve. This however can only be guaranteed to be a continuous extension and this is an instance requiring the inclusion of continuous curves.

Consider now a sequence $\{\lambda_n(t)\}$ of causal curves. We have the natural definitions:

Definition 6.8 (Convergence and Limit Points)

- *$p \in M$ is a convergence point of the sequence if for every u_p , $\exists\ N > 0$ such that $\lambda_n \cap u_p \neq \emptyset\ \forall\ n > N$;*
- *$p \in M$ is a limit point of the sequence if for every u_p, $\lambda_n \cap u_p \neq \emptyset$ for infinitely many n.*

  A convergence point is a limit point, but not conversely.
- *A curve $\lambda(t)$ is said to be a convergence curve of the sequence if every $p \in \lambda$ is a convergence point of $\{\lambda_n(t)\}$. And finally,*
- *$\lambda(t)$ is a limit curve of the sequence, if $\exists$ a* sub-sequence *$\{\lambda'_n\}$ such that λ is a convergence curve of this subsequence.*

We have the result: *If λ is a limit curve, then every $p \in \lambda$ is a limit point.* The converse is not necessarily true i.e. a curve whose every point is a limit point is not necessarily a limit curve. It could happen that there may be *no common subsequence* for which the curve is a convergence curve.

Note that a given sequence of causal curves, may or may not have (a) any convergence point; (b) any limit points; (c) any convergence curve or (d) any limit curves or combinations of these. However, we do have a theorem:

Theorem 6.6 *If $\{\lambda_n\}$ be a sequence of future directed, in-extendible, causal curves having p as a limit point, then there exists a future in-extendible causal curve through p which is a limit curve of the sequence.*

Thus, existence of even a single limit point, is enough to imply existence of a causal limiting curve. The proof *constructs the curve as a limit curve.* In the comment below the definition (6.8), a curve of limit points was *given* and this is not guaranteed to be a limit curve. Combining this with the properties of the boundary of the chronological future of subsets, we have the theorem:

Theorem 6.7 *Let $S \subset M$ be a closed subset. Then for every $p \in \dot{I}^+(S)$, $p \notin S$, the point p lies on a null geodesics which is entirely in $\dot{I}^+(S)$ and is either past-in-extendible or has an end point on S.*

Every curve is either extendible or not (has an end point or not). The non-trivial statement is that the null geodesic can have a past end point *only* on S. That S is closed is important in the proof [17].

After these preliminaries, we turn to analysis of causality proper.

6.2 Causality

The most obvious threat to the notion of causality is the existence of closed time-like or causal curves. While locally (i.e. in a convex normal neighborhood) this cannot happen, globally, we can have such curves. A simple example is to take a two-dimensional Minkowski space-time and identify the $t = 0$ and $t = 1$ (say) lines. The $x = 0$ is a closed time-like curve. Such space-times have to be explicitly excluded. This is achieved by,

Definition 6.9 (Chronology/Causality Condition)

(M, g) *is said to satisfy* chronology (causality) condition *if* $\nexists$ *a closed time-like (causal) curve in* M *i.e.* $\forall\ p \in M,\ \ p \notin I^{\pm}(p)$ *(respectively* $p \notin J^{\pm}(p)$*).*

The causality condition implies the chronology condition, but the converse is not true. There may be no closed time-like curves, but a non-space-like closed curve may have portions which are light-like and therefore would violate causality condition. How commonly are these conditions violated?

Theorem 6.8 (Compact Space-Time) *If the space-time manifold is compact, then both chronology and causality conditions are violated.*

We have already noted that a Lorentzian metric is not admissible on a compact manifold unless its Euler character is zero. Now we see that compact space-times, even with vanishing Euler character, are in-appropriate for a reliable notion of causality. Physically admissible space-times should not be compact.

Note that on a non-compact manifold too, either or both conditions can be violated as the two-dimensional example above shows.

There is another type of causal pathology that can arise - the space-time interval between two neighboring events can be both time-like and space-like. This can happen if a time-like curve starting from p has a point q which is arbitrarily close to p, which will then imply that two nearby points in a convex normal neighborhood of p, which are space-like separated, are nevertheless connected by a time-like curve, going out of the neighborhood and re-entering it. This is excluded by the notion of *strong causality.*

Definition 6.10 (Strong Causality) (M, g) *is strongly causal if for every* p *and a* u_p , $\exists\ u'_p \subset u_p$ *such that no causal curve intersects* u'_p *more than once i.e.* $\lambda(t) \cap u'_p \neq \emptyset \ \Rightarrow\ t \in (a, b)$*, in contradistinction from* t *in multiple, disjoint intervals.*

At any point and any neighborhood, there will always be one segment of a causal curve, namely, a causal curve through the point itself. Strong causality precludes another t-interval of the same causal curve within some neighborhood of the point.

Evidently, strong causality implies causality. An example where causality holds but strong causality fails is shown in the figure (6.1) [18].

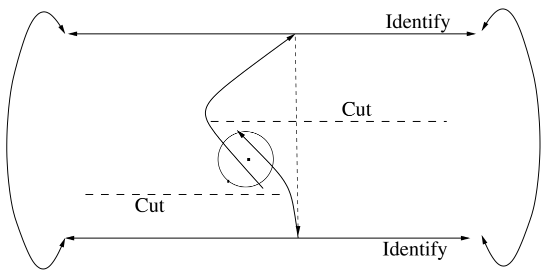


FIGURE 6.1: A two-dimensional example. Every neighborhood of the point has a causal curve entering it twice.

In a strongly causal space-time, we have a result:

In a strongly causal space-time, if we have a causal curve confined to a compact *subset of the space-time, then it must be both future and past extendible (have future and past end points).*

Notice that the compact subset may be a compact submanifold. Then by the previous theorem 6.8, there would be a violation of causality, but this cannot be so due to strong causality. The only alternative is that the causal curve must 'try to exit' the compact submanifold i.e. must have end points.

Now we come to the notion of 'stable causality' which means that even if the metric is perturbed slightly, the strong causality/causality condition continues to hold. This would obviously be true if the perturbed metric has light-cones which are 'narrower' than the original metric which trivially preserve causal character of a curve. If however, the new metric has light-cones which are 'wider', then curves which were space-like earlier could now become causal. So how does one 'widen' a metric?

Let t^μ be a time-like vector field with respect to a metric $g_{\mu\nu}$ and define $\bar{g}_{\mu\nu} := g_{\mu\nu} - t_\mu t_\nu$. Its inverse is given by, $\bar{g}^{\mu\nu} = g^{\mu\nu} + t^\mu t^\nu/(1 - t \cdot t)$. Then $\bar{g}(X,X) = g(X,X) - (g(t,X))^2$. Therefore, $g(X,X) \leq 0 \Rightarrow \bar{g}(X,X) < 0$ and $\bar{g}(X,X) \geq 0 \Rightarrow g(X,X) > (t \cdot X)^2 > 0$. Thus the light cone with respect to g is a subset of that with respect to $\bar{g}$ or $\bar{g}$ is *wider* than g.

Definition 6.11 (Stable Causality) (M, g) *is stably causal, if* $\exists$ *a continuous, non-vanishing time-like vector field* t^μ *such that* $(M, \bar{g})$ *satisfies* chronology condition.

We have a convenient characterization of stably causal space-times:

Theorem 6.9 (Stable Causality) (M, g) *is stably causal iff* $\exists$ *a differentiable function* $f : M \to \mathbb{R}$ *such that* $\partial^\mu f$ *is a future directed, time-like vector field. The function is called a global time function.*

It leads to the corollary: *Stable causality implies strong causality.* Stable causality ensures there are no causal pathologies.

Having identified conditions for a causally well behaved space-time, we seek the property that physically acceptable space-times should afford a *deterministic dynamics for all matter entities.*

6.3 Determinism and Global Hyperbolicity

The notion of determinism is tied with the ability to *predict* i.e. having given sufficient information at some 'instance', we can predict what the information will be at a later instance and also retrodict what the information was at an earlier instance which lead to the information 'now'. These intuitive ideas from a Newtonian view of the dynamics are formulated in a relativistic view, in terms of domains of dependence of a suitable submanifold. As we will see, the 'now' surface (submanifold) can be generalized to an achronal submanifold (which can have light-like portions which are absent in the Newtonian causality).

Let S be a *closed, achronal set.*

Definition 6.12 (Edge Set) *$p \in S$ is an* edge point *of S if for every neighborhood $u_p, \exists\ q, r \in u_p$ such that $q \in I^+(p),\ r \in I^-(p)$ and there is a time-like curve λ from r to q which does not intersect S.*

The set of all edge points of S is called the edge *of S, denoted edge(S).*

Note that S is only a set, no smoothness properties are implied. The x-axis of usual Minkowski space-time is clearly closed, achronal and has each of its points as its edge point since we can always skirt around by a time-like curve, hence in this case edge$(S) = S$. The x-y plane in three-dimensional Minkowski space-time however has no edge points - edge$(S) = \emptyset$. In fact we have the theorem,

Theorem 6.10 *If a non-empty, closed, achronal set S has no edge points, then S is a three-dimensional, embedded, C^0 submanifold of M.*

Such an edge-less S is called a slice.

For a generic closed, achronal set S, we define its domains of dependences.

Definition 6.13 (Future/Past Domains of Dependence)

$D^+(S)\ :=\ \{p \in M/$ *Every, past in-extendible causal curve through p intersects S* $\}$;

$D^-(S)\ :=\ \{p \in M/$ *Every, future in-extendible causal curve through p intersects S* $\}$;

$D^{\pm}(S)$ *are called future/past domain of dependence of S.*

Clearly, $S \subset D^{\pm}(S) \subset J^{\pm}(S)$. However, since S is achronal, $D^{+}(S) \cap I^{-}(S) = \emptyset = D^{-}(S) \cap I^{+}(S)$. The qualifier *every* is important since it implies that the future domain of dependence precisely consists of *only* those events whose causes have been registered on S and have no other causes un-registered on S. Likewise, the past domain of dependence consists of only those 'causes' whose 'effects' have to be registered on S. Therefore, $D(S) := D^{+}(S) \cup D^{-}(S)$, the *Domain of Dependence* of S, is the set of events at which all physical properties should be *completely determined* by the properties at events on S. Evidently for a space-time supporting predictability, we would like existence of an achronal set whose domain of dependence is the full space-time! This leads to the central definition of this section:

Definition 6.14 (Cauchy Surface and Global Hyperbolicity)
If S is an achronal, closed subset of M such that $D(S) = M$, then S is called a Cauchy Surface while the space-time is said to be Globally Hyperbolic.

It follows immediately that,

Theorem 6.11 *A Cauchy surface is edge-less i.e. a slice which is of course an embedded three-dimensional, C^0 submanifold.*

The proof is simple. If edge$(S) \neq \emptyset$, then $\exists\, p \in S$ such that every neighborhood u_p containing $q \in I^{+}(p),\ r \in I^{-}(p)$ and a time-like curve λ connecting the two without intersecting S. Hence $q,\ r$ do not belong to the domain of dependence of S. But this contradicts global hyperbolicity, hence S must be edge-less.

To consider the converse, let M be a globally hyperbolic space-time so that it *can* admit a Cauchy surface. Consider a three-dimensional, achronal, closed, edge-less submanifold, $S \subset M$. Under what conditions can such an S be a Cauchy surface?

Theorem 6.12 *S is a Cauchy surface iff every, in-extendible null geodesic intersects S and enters $I^{\pm}(S)$.*

Edge-less, achronal, closed submanifolds which are *not* intersected by all in-extendible null geodesics are called *Partial Cauchy Surfaces.* The domain of dependence of a partial Cauchy surface, while clearly not all of the space-time, *is* by itself a globally hyperbolic space-time with the surface being its Cauchy surface.

When the domain of dependence of a closed, achronal set S does not coincide with the space-time, we have the notion of a *Cauchy Horizon.* To define it, Let us note some properties of the domains of dependences for a generic closed, achronal set S. Neither the $D^{\pm}(S)$ not their closures $\overline{D^{\pm}(S)}$ coincide with the full space-time. Then the following are true:

Theorem 6.13 (Properties of Domains of Dependence)

1. *An event* $p \in \overline{D^+(S)}$ *iff every past in-extendible time-like curve from* p *intersects* S*;*

2. *It follows that,*

$$\mathrm{Int}[D^+(S)] = I^-[D^+(S)] \cap I^+(S);$$
$$\mathrm{Int}[D(S)] = I^-[D^+(S)] \cap I^+[D^-(S)];$$

3. *Define,*

$$H^{\pm}(S) := \overline{D^{\pm}(S)} - I^{\mp}[D^{\pm}(S)] \qquad \textit{(Future/Past Cauchy Horizons)}\ ,$$
$$H(S) := H^+(S) \cup H^-(S) \qquad \textit{(Cauchy Horizon)}\ ;$$

4. *Every* $p \in H^+(S)$ *lies on a null geodesic* λ *contained within* $H^+(S)$*; it is either past in-extendible or has an end point on* S*;*

5. *Cauchy horizon is the boundary of the domain of dependence:* $H(S) = \dot{D}(S)$*. As a corollary, it follows that for a connected space-time,* S *is a Cauchy surface iff its Cauchy Horizon is empty,* $H(S) = \emptyset$*;*

6. *If* Σ *is a Cauchy surface, every in-extendible, causal curve,* λ *intersects* $\Sigma, I^+(\Sigma)$ *and* $I^-(\Sigma)$*.*

Here are some simple examples. In two-dimensional Minkowski space-time, the *boundary* of the 'future' light cone (i.e. the two 45 degree lines emanating into the *future* from some point, including that point), is a closed, achronal edge-less submanifold. Its future domain of dependence is the full future light cone; its past domain of dependence is just the vertex of the light cone. This set is *not* a Cauchy surface for the Minkowski space-time - there are several null geodesics which do not intersect the light cone. Including the past light cone does not help either. The full light cone is also not a Cauchy surface.

Now we note some of the main properties of globally hyperbolic space-times. These properties refer to the absence of causal pathologies as well as implications for the topology of the space-time itself. There are additional implications related to properties of spaces of curves. These are needed in the proofs of singularity theorems and will be discussed there.

Theorem 6.14 (Well Behaviour of Causality) *A globally hyperbolic space-time satisfies the chronology condition and is strongly causal. Furthermore, it is stably causal.*

The first assertion is easy to see. If the chronology condition is violated, there there exist a closed time-like curve. If it intersects the Cauchy surface, it violates the achronality of the Cauchy surface and if it does not intersect, then it violates global hyperbolicity. If strong causality is violated, then there exists an event p such that every $u_p \supset u'_p$ which is visited more than once at least by one causal curve. The previous logic applies again to this curve. The proof of the second assertion is by construction [17].

Theorem 6.15 (Topological Implications:)

If $\Sigma,\ \Sigma'$ are two Cauchy surfaces in (M, g), then they are homeomorphic;

A globally hyperbolic space-time admits a global time function such that its level sets are Cauchy surfaces;

Thus M can be foliated by Cauchy surfaces and topologically, $M \sim \mathbb{R} \times \Sigma$, topology of Σ being arbitrary.

This concludes our discussion of determinism and global hyperbolicity. We will see their role in the singularity theorems as well as in the initial value formulations.

6.4 Geodesics and Congruences

We have seen the role of time-like or causal curves in the discussion of causality and determinism. Time-like or causal *geodesics* also play an important role in revealing the structure of a space-time. Physically they represent test particles - massive or massless and thus serve as probes. Bundles of geodesics or more technically *congruences* of geodesics serve as probes of curvature through the geodesic deviation equation and describe the 'tidal effects' or 'real gravity'. In this section we discuss their basic properties.

The idea of a congruence is that it is a family of curves which fill out an open set. Since we want congruence of geodesics, we have to consider appropriately smooth curves. The notion of *continuous* causal/time-like curves introduced above, is not adequate in this context since we will need many properties of geodesics such as preservation of inner products etc., which cannot be mimicked by 'continuous geodesics'. Hence the curves will be smooth. A family of neighborhood filling smooth curves can be equivalently represented by a vector field. Smoothness property of the congruence can be defined in terms of smoothness of the vector field. Thus we define,

Definition 6.15 (Smooth Geodesic Congruence)

A smooth congruence of time-like/null geodesics is defined by a smooth vector field whose integral curves are time-like/null geodesics.

We consider time-like and null congruences separately. For definiteness, we take the geodesics to be *future directed.*

Let ξ denote a congruence of time-like geodesics. Since each of the integral curves is time-like, we can (and do) normalize it so that the curve parameter is the proper time, $\xi \cdot \xi = -1$. Thus ξ satisfies two properties: $\xi \cdot \nabla \xi^\mu = 0$ and $\xi^\mu \nabla_\nu \xi_\mu = \frac{1}{2}\nabla(\xi^2) = 0$. Let $B_{\mu\nu} := \nabla_\mu \xi_\nu$. It follows immediately that $\xi^\mu B_{\mu\nu} = 0 = \xi^\nu B_{\mu\nu}$. Since ξ is time-like, the tensor $B_{\mu\nu}$ is 'purely spatial'.

Define the projection operator: $h^\mu{}_\nu \ := \ \delta^\mu{}_\nu + \xi^\mu \xi_\nu$. Further we define,

$$\begin{aligned}
\sigma_{\mu\nu} &:= \frac{1}{2}(B_{\mu\nu} + B_{\nu\mu}) - \frac{\theta}{3} h_{\mu\nu} && \text{Shear} \\
\omega_{\mu\nu} &:= \frac{1}{2}(B_{\mu\nu} - B_{\nu\mu}) && \text{Twist} \\
\theta &:= h^{\mu\nu} B_{\mu\nu} && \text{Expansion}
\end{aligned} \tag{6.1}$$

Thus, $B_{\mu\nu} = \sigma_{\mu\nu} + \omega_{\mu\nu} + \frac{\theta}{3} h_{\mu\nu}$. We will study the evolution of these quantities along any geodesic of the congruence.

$$\begin{aligned}
\xi \cdot \nabla B_{\mu\nu} &= \xi^\lambda \nabla_\lambda \nabla_\mu \xi_\nu \ = \ \xi^\lambda [\nabla_\lambda \ \nabla_\mu] \xi_\nu + \xi^\lambda \nabla_\mu \nabla_\lambda \xi_\nu \\
&= \xi^\lambda [\nabla_\lambda \ \nabla_\mu] \xi_\nu + \nabla_\mu \left(\xi \cdot \nabla \xi_\nu\right) - (\nabla_\mu \xi^\lambda)(\lambda_\nabla \xi_\nu) \\
&= R_{\alpha\nu\mu\beta} \xi^\alpha \xi^\beta - B_\mu{}^\alpha B_{\alpha\nu}
\end{aligned} \tag{6.2}$$

The curvature term signifies the dependence of evolutions along geodesics on the geometry of the space-time, the remaining terms involve only the congruence (and metric). From this equation, we get the equations for the shear, twist and expansion. Taking the trace of the equation with $h^{\mu\nu}$ leads to the *Raychaudhuri equation*,

$$\frac{d\ \theta}{d\tau} = \xi \cdot \nabla \theta \ = \ - \frac{\theta^2}{3} - \sigma^{\alpha\beta} \sigma_{\alpha\beta} + \omega^{\alpha\beta} \omega_{\alpha\beta} - R_{\alpha\beta} \xi^\alpha \xi^\beta \tag{6.3}$$

The first three terms come from $B_{\alpha\beta} B^{\beta\alpha}$ which explains the relative minus sign of the ω^2 term. Both the σ^2 and ω^2 terms are positive because $\sigma_{\mu\nu}, \ \omega_{\mu\nu}$ are purely spatial since $B_{\mu\nu}$ is.

Taking the antisymmetric part of the (6.2) we get,

$$\xi \cdot \nabla \omega_{\mu\nu} \ = \ - \sigma_\mu{}^\alpha \omega_{\alpha\nu} + \sigma_\nu{}^\alpha \omega_{\alpha\mu} - \frac{2}{3} \theta \omega_{\mu\nu} \tag{6.4}$$

In deriving this we have used the cyclic identity which shows that the curvature term drops out. The $\sigma^2, \omega^2, \theta^2$ terms are all symmetric in $\mu \leftrightarrow \nu$ and hence drop out too. Important point is that the twist equation is *linear* and *homogeneous* in twist. Consequently, if $\omega = 0$ at some point along a geodesic, it will remain zero along the entire geodesic.

Noting that $\xi \cdot \nabla h_{\mu\nu} = 0$ we can obtain the equation for shear from the above three equations (6.2, 6.3, 6.4):

$$\begin{aligned}
\xi \cdot \nabla \sigma_{\mu\nu} &= -\frac{2}{3} \theta \sigma_{\mu\nu} - \ \{\sigma_\mu{}^\alpha \sigma_{\alpha\nu} + \omega_\mu{}^\alpha \omega_{\alpha\nu}\} + \frac{1}{3} h_{\mu\nu} (\sigma_{\alpha\beta} \sigma^{\alpha\beta} - \omega_{\alpha\beta} \omega^{\alpha\beta}) \\
&\quad + \xi^\alpha \xi^\beta (R_{\alpha\mu\nu\beta} + \frac{1}{3} h_{\mu\nu} R_{\alpha\beta})
\end{aligned} \tag{6.5}$$

Note that the last term vanishes when contracted with $h^{\mu\nu}$.

In order to appreciate the decomposition of $B_{\mu\nu}$, consider a vector field

Z^μ which satisfies $\mathcal{L}_\xi Z = 0$. Such a vector field always exists and satisfies, $\xi \cdot \nabla Z^\mu = Z \cdot \nabla \xi^\mu$ (It is not parallelly transported). Clearly,

$$\xi_\mu \xi \cdot \nabla Z^\mu = \xi_\mu Z \cdot \nabla \xi^\mu = Z \cdot \nabla \xi^2/2 = 0 = \xi \cdot \nabla (Z \cdot \xi) - Z^\mu \xi \cdot \nabla \xi_\mu = \xi \cdot \nabla (Z \cdot \xi)$$

Hence, $Z \cdot \xi$ is constant along each geodesic. Let us take $Z \cdot \xi = 0$ at some point along a geodesic, so that Z is orthogonal to ξ along the entire geodesic. It is spatial and Its norm varies along the geodesic as,

$$\frac{1}{2}\xi \cdot \nabla Z^2 \;=\; Z^\mu \xi^\nu \nabla_\nu Z_\mu \;=\; Z^\mu Z^\nu B_{\nu\mu} \;=\; \sigma_{\mu\nu} Z^\mu Z^\nu + \frac{\theta}{3} h_{\mu\nu} Z^\mu Z^\nu$$

If the shear of the congruence is zero, then $\xi \cdot \nabla Z^2 = (2/3)\theta Z^2$. This shows that if $\theta > 0$ the norm of a vector orthogonal to the geodesic *increases* along the geodesic which justifies θ being called 'expansion'. Like wise, if expansion vanishes, $\xi \cdot \nabla Z^2 = 2\sigma_{\alpha\beta} Z^\alpha Z^\beta$. Since σ is symmetric[2], norm of Z will vary by different amounts along different spatial directions. Hence σ is appropriately called 'shear'. If both shear and expansion vanish, then the norm is preserved. Furthermore, $\xi \cdot \nabla Z^\mu = Z^\nu B_\nu{}^\mu = Z^\nu \omega_\nu{}^\mu$ implies that $Z^\mu(\epsilon) \approx (\delta_\nu{}^\mu + \epsilon \omega_\nu{}^\mu) Z^\nu$ and antisymmetry of ω implies that Z 'rotates' as it evolves along the geodesic which justifies ω being called 'twist'. Twist is also related to 'hypersurface orthogonality' of the congruence.

Theorem 6.16

A time-like congruence is hypersurface orthogonal iff its twist is zero.

By definition, ξ is hypersurface orthogonal if $T_{\mu\nu\lambda} := [\xi_\mu(\nabla_\nu \xi_\lambda - \nabla_\lambda \xi_{nu}) + cyclic] = 0$, while $\omega_{\mu\nu} = \frac{1}{2}(\nabla_\mu \xi_\nu - \nabla_\nu \xi_\mu)$. Thus vanishing of twist immediately implies hypersurface orthogonality. Conversely, if $T_{\mu\nu\lambda} = 0$, then dotting with ξ^λ and using the facts that $\xi^2 = -1$ and ξ satisfies geodesic equation, it follows that the twist vanishes[3].

Consider now the case of a hypersurface orthogonal time-like congruence so that at the surface of orthogonality, the twist is zero and by the twist equation (6.4), it is zero along any of the geodesics and this is independent of the curvature. If the space-time is Riemann flat and we choose the expansion and shear also to be zero initially, then they will be zero identically. If we have only Ricci-flatness, then initially zero shear could evolve into a non-zero shear thanks to the non-zero Weyl curvature. This in turn forces the expansion to be non-zero as well and it decreases monotonically as seen from equation (6.3). This shows that in a non-trivial *vacuum* solution of Einstein equation (Ricci-flatness), gravity pulls together nearby freely falling particles. The same conclusion follows for non-vacuum solution, provided $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0 \longleftrightarrow T_{\mu\nu}\xi^\mu\xi^\nu \geq -\frac{T}{2}\xi \cdot \xi$ i.e. the matter stress tensor satisfies the strong energy condition. Since $T_{\mu\nu}\xi^\mu\xi^\nu$ is just the local energy density measured in

[2]The shear tensor is spatial and metric on the spatial subspace is positive definite.

[3]For null geodesic congruence, the converse does *not* hold.

the rest frame of a freely falling frame, we infer that *gravity (= tidal forces) is attractive.* In this case we can say more:

Theorem 6.17

In a space-time solving the Einstein equation with stress tensor satisfying the strong energy condition, an initially converging, hypersurface orthogonal, time-like, geodesic congruence attains $\theta \to -\infty$ *in finite proper time along a geodesic.*

Expansion becoming $-\infty$ signals a singularity of the congruence and indicates development of *focussing* of nearby geodesics.

We will also focus on some specific time-like (and later on light-like) geodesic which may or may not be a member of a congruence. It is convenient to choose a pseudo-orthonormal basis, $\xi^\mu, e_i^\mu,\ \xi\cdot\xi = -1, e_i\cdot e_j = \delta_{ij}\ i = 1,2,3$ at a point along a geodesic and define it along the geodesic by parallel transport i.e. $\xi\cdot\nabla e_i^\mu = 0$. Since $B_{\mu\nu}$ is spatial, its components can be written as: $B_{ij} := e_i^\mu e_j^\nu B_{\mu\nu}$. The i, j indices will be raised/lowered by δ^{ij}/δ_{ij}.

As an example, let us construct a hypersurface orthogonal congruence of time-like geodesics, with expansion going to $-\infty$ at some point along a geodesic.

Fix a point p in the space-time and consider the set of all geodesics defined by future directed, time-like tangent vectors at p. These define a congruence, away from the point p, of future directed, time-like geodesics which *emanate from* p into its future. or *focus into* p from its past. We can make these explicit by choosing Riemann normal coordinates, x^μ, at p: $x^\mu(p) := 0$. Each such geodesic will be given locally as $x^\mu(\epsilon) = \epsilon v^\mu, v\in T_p(M), v\cdot v = -1$ and future directed. For positive ϵ the geodesics emanate from p and for negative ϵ they focus into p. Clearly, $x\cdot x = -\epsilon^2$ defines space-like hypersurfaces in a convex normal neighborhood - the future and past hyperboloids. Their normals are given by, $n_\mu := \partial_\mu(x\cdot x+\epsilon^2) = 2\eta_{\mu\nu}x^\nu = 2\epsilon v_\mu$. Thus, the vector field defined by v_μ is proportional to the normal to a space-like hypersurface and hence the geodesics congruence is hypersurface orthogonal. Furthermore, we also get $B_\mu{}^\nu := \nabla_\mu v^\nu = \epsilon^{-1}\partial_\mu x^\nu = \delta_\mu{}^\nu/\epsilon$ which implies $B_i{}^j = \epsilon^{-1}\delta_i{}^j$. The expansion of the congruence, $h^\mu{}_\nu B_\mu{}^\nu$ is $3/\epsilon$. As we approach $p, \epsilon\to 0_\mp$, the expansion $\theta\to\mp\infty$. This congruence will be useful below.

Let us return to a vector field Z which satisfies $\mathcal{L}_\xi Z = 0\ ,\ Z\cdot\xi = 0$. Recall from chapter (14), section 14.6, that this is a *deviation vector* with $\xi\cdot\nabla Z^\mu =: v^\mu$ as the *relative velocity* and $\xi\cdot\nabla v^\mu =: a^\mu$ as the *relative acceleration* which satisfies the *geodesic deviation equation:*

$$(\xi\cdot\nabla)^2 Z^\mu \ = \ R^\mu{}_{\alpha\beta\nu}\xi^\alpha\xi^\beta Z^\nu\ .$$

Referring to the orthonormal basis along a geodesic, any deviation vector, being spatial, can be expressed as, $Z^\mu = z^i e_i^\mu$.

This allows us to write the defining equation for a deviation vector as,

$$(\xi\cdot\nabla z^i)e_i^\mu = \xi\cdot\nabla Z^\mu \ = \ Z\cdot\nabla\xi^\mu = z^j(e_j\cdot\nabla\xi^\mu) \quad \Rightarrow$$

$$\frac{dz^i}{d\tau} = B_j{}^i z^j \tag{6.6}$$

Likewise, the deviation equation gets expressed as,

$$\frac{d^2 z^i}{d\tau^2} = R^i{}_{00j} z^j \qquad \text{(Jacobi equation)} \tag{6.7}$$

Both the two equations above are linear, ordinary differential equations of 1st and 2nd order respectively and can be viewed as evolution equations along any given geodesic. The defining equation for a deviation vector explicitly refers to the congruence through the B matrix while the Jacobi equation has no reference to a congruence and explicitly depends on the curvature. Notice that solution of (6.6) is a solution of (6.7), but not conversely. To emphasize this distinction, we will denote solutions of the first order equation by $z^i(\tau)$ and those of the Jacobi equation by $\eta^i(\tau)$. Both equations being linear, their solutions can be constructed from corresponding matrix equations.

Define $A^i{}_j(\tau)$ by, $\dot{A} = B(\tau)A(\tau)$, $A(0) = \mathbb{1}$. Then any non-trivial solution of the defining equation can be constructed from an initial deviation vector $z(0) \neq 0$ as, $z^i(\tau) = A^i{}_j(\tau)z^j(0)$. For the Jacobi equation, we have two initial conditions and to allow arbitrary choices of these, we define two matrices, $I(\tau), J(\tau)$ by the equations,

$$\begin{aligned} \ddot{I}^i{}_j(\tau) &= R^i{}_{00k}(\tau)I^k{}_j(\tau) \ , \ I(0) = \mathbb{1} \ , \ \dot{I}(0) = \mathbb{0} \\ \ddot{J}^i{}_j(\tau) &= R^i{}_{00k}(\tau)J^k{}_j(\tau) \ , \ J(0) = \mathbb{0} \ , \ \dot{J}(0) = \mathbb{1} \end{aligned}$$

A general Jacobi field is then given by, $\eta^i(\tau) = I^i{}_j(\tau)\eta^j(0) + J^i{}_j(\tau)\dot{\eta}^j(0)$.

Clearly, if $\vec{z}(0) = 0$ then the deviation vector is identically zero. However, it may happen that a non-trivial deviation vector can still vanish at some points along the geodesic. For this to happen, we must have $detA$ vanish at these points. When is this possible? The defining equation gives, $B(\tau) = \dot{A}A^{-1}$ which implies $\theta(\tau) = Tr(B) = Tr(\dot{A}A^{-1}) = d_\tau(Tr \ln A) = (detA)^{-1}d_\tau(detA)$. Therefore, if $detA$ is to vanish as $(\tau - \tau_*) \to 0_-$, then we must have $\theta \to -\infty$. Thus, a non-trivial deviation vector can vanish at some point in the future, provided the expansion of the congruence diverges as that point is approached. That such a choice of initial deviation vector exists is seen from the example discussed above.

Next, suppose we are given a non-trivial Jacobi field vanishing at some point p, when can we find a deviation vector which matches with the Jacobi field at least when both are non-zero? Since deviation vectors depend on a congruence, it is enough to find *some* congruence with at least *one* deviation vector matching.

Let $p = \gamma(0)$ be a point at which a Jacobi field $\eta(0) = 0$. For $\epsilon > 0$, $\eta(\epsilon) \approx \eta(0) + \epsilon\dot{\eta}(0) + (\epsilon^2/2)\ddot{\eta}(0)\ldots$. The first and the last terms vanish because of the initial condition and the Jacobi equation. Let if possible, $z(\tau)$ be a deviation vector which matches $\eta(\tau)$ for all $\tau \geq \epsilon$. Let $\hat{z} := \eta(\epsilon) = \epsilon\dot{\eta}(0)$. Define $z(\tau)$ as a solution of (6.6) with $\hat{z}$ as initial condition. This will be identical to $\eta(\tau)$ if at some point, say $\gamma(\epsilon)$, we have $z(\epsilon) = \eta(\epsilon)$ *and* $\dot{z}(\epsilon) = \dot{\eta}(\epsilon)$. We have already matched the values. Requiring that the derivatives match, gives: $\dot{z}^i(\epsilon) = \dot{\eta}(\epsilon) \Rightarrow B_j{}^i(\epsilon)\hat{z}^j \approx \dot{\eta}^i(0) + \epsilon\ddot{\eta}^i(0) = \dot{\eta}^i(0)$, since the Jacobi equation implies that the $\ddot{\eta}(0)$ vanishes. Hence, $\dot{\eta}^i(0) = B_j{}^i\hat{z}^j = B_j{}^i\epsilon\dot{\eta}^j(0)$ which implies, $[\epsilon B_i{}^j(\epsilon) - \delta_i{}^j]\dot{\eta}^j(0) = 0$. Thus, a Jacobi field vanishing at p will match with a deviation vector for all $\tau \geq \epsilon$ provided $det(B - \epsilon^{-1}\mathbb{1}) = 0$. Note that this is a condition on the congruence *and* the particular Jacobi field. We could have a matching deviation vector for *every* Jacobi field vanishing at p, by choosing a congruence with $B = \epsilon^{-1}\mathbb{1}$. And we *can* choose such congruences as shown by the example above. We conclude that,

Theorem 6.18

If η is a non-trivial Jacobi field vanishing at a point $p = \gamma(0)$, then there exists a deviation vector z such that $z(\tau) = \eta(\tau)$, $\forall\ \tau \geq \epsilon$. Such a deviation vector can be chosen for the congruence of geodesics emanating from (focusing into) p.

Thus, if along any given a geodesic, there are several points at which a Jacobi field vanishes, that at each of these, we can find hypersurface orthogonal geodesic congruences with a non-trivial deviation vector matching with the Jacobi field. Given that 'gravity is attractive', does it follow that there *will be* another point on the given geodesic at which the Jacobi field will vanish?

Definition 6.16 (Conjugate Points)

p and q on a geodesic γ are said to be conjugate points if there is a Jacobi field vanishing at both the points.

We have a theorem:

Theorem 6.19 (Existence of Conjugate Points)

Let (M, g), be a space-time such that $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0\ \forall$ time-like vectors ξ^μ. For a time-like geodesic γ and a point p on it, consider a Jacobi field vanishing at p and the geodesic congruence emanating from p. Let $r \in \gamma$ be such that the expansion is negative *at r. Then within $\Delta\tau \leq 3/|\theta|_r$ from r, there exist a point $q \in \gamma$ conjugate to p, assuming that the geodesic extends that far.*

The congruence of emanating geodesics is hypersurface orthogonal and hence twist free. Each term on the right-hand side of the Raychaudhuri equation, (6.3), is negative and hence implies that there exists a q at which the expansion goes to $-\infty$. To show that this implies that q is conjugate to p, we have to show a Jacobi field vanishing at both the points.

Consider the matrix equation, $\ddot{J}^i{}_j(\tau) = R^i{}_{00k}(\tau)J^k{}_j(\tau)$, $J(0) = \mathbb{0}, \dot{J}(0) = \mathbb{1}$. Then $\eta(\tau) := J(\tau)\dot{\eta}(0)$ is a Jacobi field vanishing at $p = \gamma(0)$, for *every choice* of $\dot{\eta}(0)$. By the previous theorem, every such Jacobi field matches with a deviation vector $z(\tau)$ satisfying $\dot{z}(\tau) = B(\tau)z(\tau)\ \forall\ \tau \geq \epsilon$. It follows that $\dot{J}(\tau) = B(\tau)J(\tau)$. Now $\theta = Tr(B) = d_\tau(\ln det(J)) \to -\infty$ at $q = \gamma(\tau_*)$, implies $detJ \to 0$. Hence, there is a choice of $\dot{\eta}(0)$ such that $\eta(\tau_*) = 0$ (namely, the eigenvector of J with zero eigenvalue). We have thus found a Jacobi field vanishing at p and q i.e. q is conjugate to p.

Note that existence of a conjugate point q is conditional on existence of point r after p at which expansion of the emanating congruence is *negative.* There is a stronger version of the theorem regarding existence of conjugate points, namely,

Theorem 6.20 (Existence of Conjugate Points)
Let γ be a geodesic. Let $p_1 := \gamma(\tau_1)$ be such that $R_{\mu\nu}\xi^\mu\xi^\nu(\tau_1) \neq 0$. Let $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0$ all along the geodesic, then $\exists\ \tau_0 < \tau_1 < \tau_2$ such that $p := \gamma(\tau_0)$ and $q := \gamma(\tau_2)$ are conjugate points provided the geodesic extends that far.

One also has a notion of a point conjugate to a spatial hypersurface.

Let Σ be a spatial hypersurface and ξ be a geodesic congruence orthogonal to Σ. Let p be a point on a geodesic γ in this congruence. *p is said to be conjugate to Σ along γ if $\exists$ a non-trivial deviation vector $z \neq 0$ on Σ and vanishing at p.*

The corresponding existence theorem states that *If the space-time satisfies the condition $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0$ and $\theta|_\Sigma < 0$, then there exists a point p conjugate to Σ along a geodesic.*

Conjugate points are important because they invalidate the property of geodesics being curves of (locally) maximum 'length' among the time-like curves connecting two given points. This is sharpened as follow.

Fix p and a $q \in I^+(p)$ in M. Let $\lambda(\alpha, t)$ denote a smooth family of time-like curves so that for each α, we have a time-like curve from p to q with the parameter $t \in [a, b]$. Smoothness means that $\lambda(\alpha, t)$ constitute an embedded two-dimensional surface in M, $[\partial_t, \partial_\alpha] = 0$. Denote $T^\mu\partial_\mu := \partial_t\ ,\ X^\mu\partial_\mu := \partial_\alpha$. This T is a time-like vector (not normalized to -1) and X is called a *deviation vector* (not a geodesic deviation vector) which vanishes at the endpoints. Define,

$$\tau(\alpha) \ := \ \int_a^b dt f(\alpha, t) \quad , \quad f(\alpha, t) := \sqrt{-T^\mu T^\nu g_{\mu\nu}(t)}. \tag{6.8}$$

Clearly $\tau(\alpha)$ is positive and is called the *length function.* The following results hold [17]:

$$\begin{aligned} \frac{d\tau(\alpha)}{d\alpha} &= \int_a^b dt\,[X^\mu T^\nu \nabla_\nu\,(T_\mu/f)] \\ \therefore\ \left.\frac{d\tau(\alpha)}{d\alpha}\right|_{\alpha_0} &= 0\ \forall\ X \quad \Rightarrow \quad \lambda(\alpha_0, t) \ \text{ is a geodesic.} \qquad (6.9) \\ \left.\frac{d^2\tau(\alpha)}{d\alpha^2}\right|_{\alpha_0} &= \int_a^b dt\ X^\mu \left\{g_{\mu\nu}(T\cdot\nabla)^2 - R_{\mu\rho\sigma\nu}T^\rho T^\sigma\right\} X^\nu. \qquad (6.10) \end{aligned}$$

In getting the final simplified expression for the second variation, the extremal curve is taken to be a geodesic which is affinely parametrised ($f = 1$ along the geodesic) and the deviation vector is taken to be orthogonal to the geodesic (thus X is space-like). The expression in the braces is just the operator appearing in the geodesic deviation equation. *If* it is *negative definite*, then the second variation is negative and the geodesic, $\lambda(\alpha_0)$ is a local *maximum* of the length function. It also means that there are no conjugate points between p, q. The details of these calculations may be seen in [18].

As an implication, we have the two theorems,

Theorem 6.21

Let γ be a time-like geodesic between two points p, q. The necessary and sufficient condition for γ to maximize the proper time function over smooth, one parameter variations is that γ has no *conjugate points between p, q.*

Theorem 6.22

Let γ be a smooth, time-like curve connecting a point $p \in M$, to some point q on a space-like hypersurface Σ. γ locally maximizes the proper time function between p and Σ iff γ is a geodesic, orthogonal to Σ with no point conjugate to Σ between p and Σ.

In essence we have seen that for space-times with $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0 \; \forall \; \xi\cdot\xi < 0$, conjugate points always exist on sufficiently 'long' time-like geodesics thanks to the Raychaudhuri equation and whenever they exist, the geodesic segment containing a pair of conjugate points cannot locally maximize proper times.

There is another property of physically acceptable maximally extended space-times which forces geodesics to attain maxima of proper time function and this produces a contradiction of the singularity theorems discussed later.

Analogous notions and results also hold for null geodesic congruences. We note below the distinctive features of null congruences and the results.

Null Geodesic Congruence is defined by a smooth everywhere light-like vector field k^μ , $k \cdot k = 0$ satisfying the geodesic equation, $k \cdot \nabla k^\mu = 0$. Define as before, $B_{\mu\nu} := \nabla_\mu k_\nu$. It follows that

$$k^\mu B_{\mu\nu} \;=\; 0 \;=\; B_{\mu\nu}k^\nu \quad , \quad (k\cdot\nabla)B_{\mu\nu} = R_{\alpha\nu\mu\beta}k^\alpha k^\beta - B_\mu{}^\alpha B_{\alpha\nu}$$

We cannot define $h_{\mu\nu}$ though thanks to $k^2 = 0$. Hence, definition of shear and expansion is not straightforward. Likewise, while defining deviation vectors, we cannot use the condition $Z \cdot k = 0$. Consider first the deviation vectors. Let Z^μ be defined by $\mathcal{L}_k Z^\mu = 0$ or $k \cdot \nabla Z^\mu = Z \cdot \nabla k^\mu$. Along a given geodesic, choose an 'orthonormal basis' by making a choice at a point p and parallel transporting it along the geodesic: $k\cdot\nabla e^\mu_{\dots} = 0$. At p we choose: $e^\mu_0 := k^\mu$, $e^\mu_1 := l^\mu$, e^μ_a, $a = 1, 2$ such that,

$$l^2 = 0 \;,\; k\cdot l = -1 \;,\; e_a \cdot e_b \;=\; \delta_{ab} \;,\; k\cdot e_a \;=\; 0 \;=\; l \cdot e_a \;.$$

Note that neither l nor the 2-plane spanned by e^μ_a is uniquely determined.

Referring to such a basis, we write $Z^\mu := z^0 k^\mu + z^1 l^\mu + z^a e_a{}^\mu$. Now,

$$0 = k_\mu Z \cdot \nabla k^\mu = k_\mu k \cdot \nabla Z^\mu = k \cdot \nabla (k \cdot Z) = -\frac{dz^1}{d\lambda}$$

Hence, it is possible to choose $z^1 = 0$ consistently along the geodesic and we will do so. Substitution of the expansion of Z, we get,

$$\frac{dz^a}{d\lambda} = B_b{}^a z^b := e_a^\mu e_b^\nu B_{\mu\nu} \quad , \quad \frac{dz^0}{d\lambda} = -\left(l^\mu e_a^\nu B_{\nu\mu}\right) z^a$$

Notice that equation for z^a does not depend on z^0 and in turn determines z^0. The 'orthogonality' condition $Z \cdot k = 0$ precisely leaves the z^0 component ambiguous. Hence, we *define* the equivalence class of Z relative to the equivalence relation , $Z' \sim Z \Leftrightarrow Z' = Z + \alpha k$, as a deviation vector for the null congruence.

The deviation equation for B_{ij} is derived as,

$$\begin{aligned} k \cdot \nabla B_{ab} &= e_a^\mu e_b^\nu R_{\alpha\nu\mu\beta} k^\alpha k^\beta - (e_a^\mu B_\mu{}^\alpha)(B_{\alpha\nu} e_b{}^\nu) \\ &= R_{\alpha ba\beta} k^\alpha k^\beta - B_{a\alpha} B_{\beta b} g^{\alpha\beta} \quad \text{and} \because B_a{}^\alpha = B_a{}^0 k^\alpha + B_a{}^c e_c^\alpha, \\ \therefore \frac{dB_{ab}}{d\lambda} &= -R_{a\alpha\beta b} k^\alpha k^\beta - B_a{}^c B_{cb} \end{aligned} \tag{6.11}$$

The shear, twist and expansion can now be defined and computed as,

$$B_{ab} := \sigma_{ab} + \omega_{ab} + \frac{\theta}{2} h_{ab} \quad , \quad h_{ab} = \delta_{ab}$$

Using this equation and the defining equation for a deviation vector, we deduce the geodesic deviation equation (or Jacobi equation) as,

$$\begin{aligned} \frac{d^2 z^a}{d\lambda^2} &= \frac{dz^b}{d\lambda} B_b{}^a + z^b \frac{dB_b{}^a}{d\lambda} = z^c B_c{}^b B_b{}^a - z^b \left(R_{b\alpha\beta}{}^a k^\alpha k^\beta + B_b{}^c B_c{}^a\right) \\ \therefore \ddot{z}^a &= -R^a{}_{00b} z^b \end{aligned} \tag{6.12}$$

We have obtained the basic defining equation for deviation vector as well as Jacobi equation in the same form as before except the indices take two values instead of three. Subsequent analysis of conjugate points and their existence is similar to the time-like case except: (a) hypersurface orthogonality of the congruence implies vanishing twist but the *converse is not true*; (b) the proper time is replaced by affine parameter; (c) the condition on the Ricci tensor, for the Raychaudhuri equation to force a conjugate point is $R_{\mu\nu} k^\mu k^\nu \geq 0 \ \forall \ k^2 = 0$; (d) the deviation vector is an equivalence class and therefore its vanishing at a point is also defined to within an arbitrary component along k^μ.

Since there is no proper time for light-like geodesics, the implications of conjugate points is differently formulated. The relevant theorem is,

Theorem 6.23 (Consequence of Conjugate Points)

Let γ be a smooth causal curve connecting p and $q \in J^+(p)$. Then, it cannot be deformed to a time-like curve via a smooth, 1-parameter variation over causal curves iff γ is a null geodesic with no conjugate points between p and q.

The notion of point conjugate to a hypersurface also changes. Let Σ be a *two-dimensional* space-like submanifold. At every point, the complement of its tangent space has *two null basis* vectors which we label as 'out-going' and 'in-coming'. For an orientable Σ a consistent choice of this labelling can be made over Σ. We can construct two null geodesic congruences emanating from any point on Σ, by selecting the sets of all out-going (in-coming) future directed light-like vectors. Both of these will have their twists equal to zero.

Let μ be one of the out-going (in-coming) null geodesic orthogonal to Σ. *$p \in \mu$ is said to be conjugate to Σ if $\exists$ a non-trivial deviation vector which is non-zero at Σ but vanishes at p.*

With these definitions we have the obvious analogues of the theorems for the time-like case. One additional result, specific to null geodesic congruences is,

Theorem 6.24 *Let the space-time be globally hyperbolic and let K be a compact, two-dimensional, orientable, space-like submanifold. Then every point $p \in \dot{I}^+(K)$ lies on a future directed null geodesic orthogonal to K and with conjugate point in-between.*

We will note more results relevant in the context of singularity theorems.

6.5 Singularity Theorems

We have so far noted the conditions under which conjugate points do exist and some of their implications. A particularly crucial property is the failure of local maximization of the proper time function when conjugate points exist as noted in the theorems (6.21, 6.22). Now we note conditions for attaining the global maximum of the function.

For this, let us introduce the space $C(p,q)$, which is the space of all continuous, future directed, causal curves connecting points p and $q \in J^+(p)$. Likewise, for a smooth, achronal hypersurface Σ and $q \in J^+(\Sigma)$, we denote by $C(\Sigma, q)$ the space of all continuous causal curves from some point $p \in \Sigma$ to q.

Let (M,g) be strongly causal. For u an *open* subset of M, define the subset $O(u) \subset C(p,q)$ consisting of those causal curves which are contained entirely in u (here curve means the set of points). Note that $O(u)$ can be empty for some u. Taking these subsets together with their arbitrary unions and finite intersections as open sets on $C(p,q)$, define a topology on $C(p,q)$. This topology is controlled by the causality properties of (M,g). For instance, strong causality on (M,g) is required for this topology to be 'nice' (e.g. Hausdorff) and have the same notion of convergence of sequence of continuous causal curves as defined before. Furthermore, *if (M,g) is globally hyperbolic, then $C(p,q)$ is compact* [17].

Analogous definition is given for the set of all continuous, causal, future directed curves from a Cauchy surface Σ to a point q it its future. The space $C(\Sigma,q)$ is then also compact. These compactness properties lead to:

Theorem 6.25

Let (M,g) be strongly causal. Consider the length function, $\tau(\gamma)$ (see eqn. 6.8) defined for $\gamma \in C(p,q)$.

If τ attains its maximum value for γ_0, then γ_0 is a geodesic with no conjugate points.

Likewise, for τ defined over $C(\Sigma,q)$, if τ attains its maximum at γ_0, then γ_0 is a geodesic, orthogonal to Σ and with no conjugate points.

Note that only *strong causality* is needed. Also, τ need *not* attain its maximum! *However,*

Theorem 6.26

If (M,g) is globally hyperbolic, then there exists a $\gamma_0 \in C(p,q)$ on which the length function does attain its maximum.

Likewise, in a globally hyperbolic space-time, if Σ is a Cauchy surface, then there exist a causal curve $\gamma_0 \in C(\Sigma,q)$ on which the maximum is *attained.*

By the previous theorem, γ_0 is a geodesic with no conjugate points between p and q.

As yet, there is no contradiction between local existence of conjugate points along a finite geodesic and global existence of also a finite a geodesic with no conjugate points on it.

With a further input of certain 'physical' conditions, a contradiction does arise and these are formulated as the *Singularity Theorems.* There are four main versions - two pertaining to cosmological context and two to compact bodies. The formulation is as given in [17].

Theorem 6.27 (Singularity Theorem 1)

Let (M,g) be globally hyperbolic with $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0\ \forall$ time-like vectors ξ (equivalently for a solution of Einstein equation with stress tensor satisfying the strong energy condition). Suppose there exists a space-like Cauchy surface, Σ, for which the trace of the extrinsic curvature, $k := g^{\mu\nu}\nabla_\mu n_\nu \ \leq C < 0$ everywhere on Σ , then no past directed, time-like curve from Σ can have a length greater than $3/|C|$. In particular, all past directed time-like geodesics are incomplete.

The proof is short, so let us go over it. Let if possible λ be a past directed time-like curve with length greater than $3/|C|$. Let p be a point on λ beyond the length $3/|C|$. By theorem (6.25), since strong causality is implied by global hyperbolicity, λ is a geodesic with no conjugate points. However, since the expansion of the geodesic congruence emanating from p is negative and bounded away from zero, by theorem (6.19), a conjugate point must exist within length $\leq 3/|C|$. We reach a contradiction and hence the presumed curved *cannot* exist.

The inputs that have gone in the conditions of the theorem are: two generic conditions - good causal behaviour and space-time being a solution of Einstein equation with a physically reasonable stress tensor (which roughly says that the gravity is attractive) *and* a condition stipulating a special context roughly saying that the universe is expanding everywhere with a rate bounded away from zero. May be the special condition is too special. Here is a second version.

Theorem 6.28 (Singularity Theorem 2)

Let (M, g) be strongly causal satisfying $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0 \ \forall$ time-like ξ everywhere on M. Suppose $\exists$ $\underline{a}$ compact, edge-less, achronal, smooth space-like hypersurface S such that for past directed normal geodesic congruence from S, its expansion is everywhere negative and bounded away from zero, then <u>at least one</u> past directed time-like geodesic from S has length $\leq 3/|C|$.

This proof too is by *reductio ad absurdum.* Suppose *all* past directed, inextendible, time-like geodesics have length greater than $3/|C|$. The portion $\tilde{M} := int[D(S)] \subset M$, is 'globally hyperbolic' and hence by the previous theorem all the above geodesics must be incomplete. Thus they must exit $\tilde{M}$ and hence must intersect the boundary $H^-(S)$ of $\tilde{M}$ and this must happen *before* $3/|C|$ i.e. the boundary is non-empty. Now one shows that the past Cauchy horizon is *compact.*

However, S being edge-less implies that the past Cauchy horizon contains a future inextendible null geodesic while its compactness in a strongly causal space-time makes this impossible (see the result near the figure 6.1). This invalidates the initial assumption thereby proving the theorem.

Both these theorems correspond to the cosmological context as it uses the condition of everywhere non-zero expansion and deduces past-incomplete time-like curves. Notice that the theorem does *not* assert that the universe *must be* everywhere expanding. This happens to be an observational input for our universe.

There are two more theorems which correspond to gravitational collapse. For this we first need a definition,

Definition 6.17 (Trapped Surface)

A compact, two-dimensional, space-like submanifold is said to a trapped surface if the expansions $\theta_\pm$ for both the orthogonally out-going and in-coming, null geodesics is strictly negative at all points of the surface. If zero expansion *is allowed for one of the null geodesics congruences, then one has* outer/inner marginally trapped surface.

Theorem 6.29 (Singularity Theorem 3)

Let (M, g) be globally hyperbolic with a non-compact *Cauchy surface. Suppose $R_{\mu\nu}k^\mu k^\nu \geq 0 \; \forall \; k^2 = 0$ (equivalently, solution of Einstein equation with stress tensor satisfying the weak or the strong energy condition). Suppose the space-time contains a trapped surface, S, with $\theta_0 < 0$ being the maximum value for both $\theta_\pm$. Then, at least one, future directed, in-extendible, orthogonal null geodesic emanating from S, has an affine length $\leq 2/|\theta_0|$.*

This context corresponds to collapse because non-compact Cauchy surfaces arise in asymptotically flat space-times which correspond to space-times appropriate for isolated bodies. The incomplete null geodesic is future directed. This proof derives a contradiction between the trapped surface being compact and the Cauchy surface being non-compact. The existence of trapped surface signals that the collapse has progressed *far enough to be irreversible.* This theorem too does *not* assert that collapse must advance enough to form a trapped surface. That this is likely to happen in an astrophysically realizable context, is dependent on further dynamical properties of collapsing matter.

Finally we have the version,

Theorem 6.30 (Singularity Theorem 4)

Suppose a space-time satisfies the following conditions:

(a) $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0 \; \forall$ time-like and null ξ^μ (translates into strong energy condition on the stress tensor);

(b) Every time-like geodesic has at least one point with $R_{\mu\nu\alpha\beta}\xi^\mu\xi^\beta \neq 0$ (time-like genericness condition) and every null geodesic has a point at which either $R \cdot k \cdot k > 0$ or $k_{[\rho}C_{\mu]\nu\alpha[\beta}k_{\sigma]}k^\nu k^\alpha \neq 0$ (null genericness condition);

(c) No closed time-like curves exist; and

(d) At least one of the following holds: (i) (M, g) has a compact, edgeless achronal set ('closed universe'), (ii) (M, g) has a trapped surface, or (iii) there exists a point $p \in M$ such that expansion of the future (or past) directed null geodesics emanating from p becomes negative along each geodesic in this congruence; then, there exist at least one incomplete time-like or null geodesic.

This version has significantly weakened the causal properties and has also replaced the expanding universe hypothesis by the closed universe hypothesis. The conclusion too is correspondingly weaker - only existence of *one* causal curve is inferred with no further information.

We should emphasize that apart from general conditions such as good causal behaviour and attractiveness of gravity (energy conditions), there is always a condition identifying a *special physical input* such as existence of trapped surface or everywhere expanding universe with expansion bounded away from zero, and *only with this additional input*, geodesic incompleteness is deduced. Without such an input, there are perfectly non-singular solutions in valid physical situations.

There have been further developments beyond these classic singularity theorems. Notable among these is the singularity theorem applicable to the context of inflationary cosmology. Recall the singularity theorems which apply to the cosmological context. They use the strong energy condition. *If* the universe is non-singular (and other causality conditions hold together with the expansion being bounded away from zero everywhere), then the strong energy condition must be violated. For perfect fluid in FLRW cosmology, this means $\rho + 3P < 0$ and the Raychaudhuri equation immediately implies existence of an *accelerated* phase. Note that the converse need not be true i.e. violation of energy condition does *not guarantee* non-singularity. Borde and Vilenkin considered the new context of inflation which does violate strong energy condition but nevertheless proved that under the assumption of *future eternal inflation*, the space-time is past geodesically incomplete [36].

The space-times in the context of singularity theorems are all in-extendible (without which incompleteness of geodesics can be trivially arranged). These are usually constructed by extending/patching different solutions with certain degrees of smoothness for the metric as well as stress tensors assumed. These could be relaxed and new versions of singularity theorems could be explored. The characterization of singularity as geodesic incompleteness itself can also be replaced by another suitable criteria and in fact has been done [37].

For an extensive review of singularity theorems, please see [38].

One could certainly question the validity of Einstein equations themselves in the context when singularities are suppose to occur especially if the geodesic incompleteness is accompanied by diverging curvature invariants. Energy conditions, which enter through the Einstein equation could also be violated if quantum effects are significant, though this by itself does not guarantee non-singularity as pointed out above. What then is a *physical* import of these theorems?

In brief, these theorems demarcate the conditions under which it is impossible to have a physically well behaved, self consistent, *classical* model of relativistic space-time with matter. These inadequacies of classical models perhaps hint at the need to go beyond and also give a hint as to in which physical context one may seek extensions—the early universe and an un-stoppable gravitational collapse.

Chapter 7

Asymptotic Structure

The issue of asymptotic structure of space-times is tied with the question: what is the appropriate space-time which corresponds to an 'isolated body' or a source of gravitation confined to a compact region? This means that we expect there are regions in the manifold where the matter stress tensor vanishes and hence the metric satisfies the vacuum Einstein equation. Allowing for the possibility of a *cosmological constant*, there are precisely three matter-free space-times which are simplest in the sense that they have maximum possible symmetry. These are: the Minkowski space-time ($\Lambda = 0$), De Sitter ($\Lambda > 0$) and the anti-De Sitter ($\Lambda < 0$). The space-times exterior to the matter sources are expected to be approaching these special solutions as one goes 'sufficiently far away' from the sources. Thus we need to understand what the 'sufficiently far away' (or infinity) from the 'origin' (a point interior to a compact region) means for these special solutions. This is non-trivial because the Lorentzian signature implies there are different ways to approach *an* infinity. For instance, in a Euclidean space, we can 'go to large r' in different directions. For Lorentzian space-times, there is a further possibility of 'going to large r' with *different speeds* along different directions. In particular the speeds are delineated by the speed of light. These asymptotic approaches can be understood in terms of taking (say) affine parameters to their asymptotic values along time-like/light-like/space-like geodesics. The solutions of basic wave equation also shows possibilities of different asymptotic behaviours. Secondly, without any preferred coordinate system, specification of asymptotic fall-off behaviours of *metric* is at best ambiguous. It would conceivably be easier if we could bring the 'infinity' - region of infinite coordinate values - to region of *finite* coordinate values. Let us see how this could work. We will focus on $\Lambda = 0$ case first and comment on the other cases at the end.

Introduce the standard spherical coordinates in the Minkowski space-time, M, so that the line element is given by,

$$ds^2 \;=\; -dt^2 + dr^2 + r^2 d\omega^2 \quad , \quad d\omega^2 := d\theta^2 + \sin^2\theta d\phi^2$$

Define, $u := t - r,\ v := t + r \ \leftrightarrow\ t = (u+v)/2,\ r = (v-u)/2$. The line element then becomes, $ds^2 = -dudv + \frac{1}{4}(v-u)d\omega^2$. Suppress the angular part for notational convenience. Both $u, v \in \mathbb{R}$ with the restriction, $v - u \geq 0$. We can bring the infinite range of u, v to a finite range by using new coordinates: $U := \tan^{-1}u$, $V := \tan^{-1}v$, both ranging over $(-\pi/2, \pi/2)$. Further introduce, $T := V + U$, $R := V - U \ \leftrightarrow\ 2V = T + R$, $2U = T - R$

both $\in (-\pi, \pi)$ and $v - u \geq 0$ implies $R \geq 0$ as well. It follows that,

$$\begin{aligned}
dudv &= [sec^2U sec^2V]dUdV \quad , \\
(v-u)^2 &= (\tan V - \tan U)^2 = \tan^2(V-U)\,(1+\tan U \tan V)^2 \\
&= tan^2(V-U)(sinU sinV + cosU cosV)^2 \ sec^2U \ sec^2V \\
&= [sec^2U sec^2V]tan^2(V-U)cos^2(V-U) \\
&= [sec^2U sec^2V]sin^2R \qquad (7.1) \\
\therefore ds^2 &= \left[\frac{sec^2U sec^2V}{4}\right]\left(-dT^2 + dR^2 + sin^2Rd\omega^2\right) \qquad (7.2) \\
&:= [\Omega^{-2}](d\tilde{s}^2)
\end{aligned}$$

By changing the coordinates, we have transformed the infinite extent (u, v) chart on the Minkowski space-time (angular part suppressed), into the finite extent chart (U, V) or (T, R) and getting in the process a new metric $d\tilde{s}^2$ which is *conformal* to the Minkowski metric. Notice that the conformal factor vanishes as $U, V \to \pm\pi/2$ or $T \pm R \to \pm\pi$ and these are exactly the points that correspond to going-to-infinity along various directions in the original chart. In terms of (U, V), we can *extend* the chart to include these points since the $d\tilde{s}^2$ metric is well behaved and thus obtain a *conformal extension* of the Minkowski space-time. The spatial metric is the canonical metric on S^3. Once $R = 0, \pi$ points are added during the extension, we get the extended space-time, $\tilde{M}$ to be $\mathbb{R} \times S^3$. What is the 'boundary' of the original Minkowski space-time in the extended space-time?

The boundary of M in $\tilde{M}$ is precisely defined by the the set of points where $\Omega = 0$. These are given by $U = \pm\pi/2$, *or* $V = \pm\pi/2$, *or both* $U, V = \pm\pi/2$ such that $R = V - U \geq 0$. The hypersurface $R = 0$ ($\leftrightarrow r = 0$) is already part of the Minkowski space-time. Additionally considering radial geodesics (straight lines) and taking their affine parameter to infinity, we can discover the boundary points reached. This leads to figure 7.1.

Future Null Infinity:	$\mathcal{J}^+$	$(U \in (-\pi/2, \pi/2), V = \pi/2) \times S^2$
Past Null Infinity:	$\mathcal{J}^-$	$(U = -\pi/2, V \in (-\pi/2, \pi/2)) \times S^2$
Future Time-like Infinity:	i^+	$(U = \pi/2, V = \pi/2)$
Past Time-like Infinity:	i^-	$(U = -\pi/2, V = -\pi/2)$
Spatial Infinity:	i^0	$(U = -\pi/2, V = \pi/2)$

The i^+, i^-, i^0 are single points since they have $R = 0, \pi$ at which the S^3 degenerates to a point. The u, v axes are oriented so that light propagates along 45^0 lines. The labels of the different components of the 'infinity' indicate the boundaries asymptotically reached by all time-like (future/past directed), light-like (future/past directed) and space-like curves respectively[1].

[1]Note that the Null infinities, $\mathcal{J}^{\pm}$, which by definition are boundary points of null geodesics, are themselves null hypersurfaces. This however is not always the case. In the De Sitter space-time, the null infinity hypersurfaces are *space-like*.

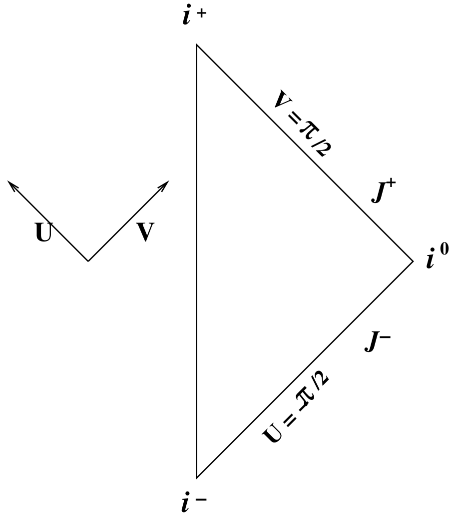


FIGURE 7.1: Conformal diagram of Minkowski space-time.

Observe that Minkowski space-time is geodesically complete and hence is *inextendible* as a vacuum solution with $\Lambda = 0$. The extension we have seen above is a conformal extension. The boundary of M in $\tilde{M}$ has the property that *every geodesic has its end-points on the boundary of* M *in* $\tilde{M}$*, none are missed.* One can easily verify this by considering radial geodesics. We have obtained a satisfactory description that whatever attempts to exit from a compact region ends up at the boundary.

The idea now is to *define* an asymptotically flat space-time to be a solution of Einstein equation which *admits an asymptotic structure* similar to that of the Minkowski space-time. The notion of asymptotic structure is understood to mean an embedding of (M, g) into $(\tilde{M}, \tilde{g})$ so that the boundary of the embedded M has specified components and on the embedded M, the two metrics are conformal to each other. A judicious choice of these allows us to provide a suitable class of coordinates near the boundary and stipulate specified fall off behaviours for the metric which is the eventual goal of this construction. It will turn out that the choices made also allow formulation of conservation laws and corresponding definitions of conserved quantities such as mass/energy, angular momentum etc. We will discuss the definitions for only *asymptotically* empty and flat solutions of Einstein equation, based on conformal completion [17,39,40] and comment briefly on more recent alternatives in the context of spatial infinity [41]. For a recent review of conformal techniques, please see [42].

Definition 7.1 *A space-time* (M, g)*, called* physical space-time*, is said to be asymptotically empty and flat at null and* spatial infinity *if there exists an-*

other, unphysical*, space-time* $(\tilde{M}, \tilde{g})$ *and an embedding* $\phi : M \to \tilde{M}$ *satisfying following conditions:*

1. *The metric,* $\tilde{g}$ *is smooth everywhere except possibly at a point* i^0*, called the spatial infinity, where* $\tilde{M}$ *is* $C^{>1}$ *while* $\tilde{g}$ *is* $C^{>0}$ *(defined below).*

   Let $J(i^0) := J^+(i^0) \cup J^-(i^0)$ *denote the causal future and past of the spatial infinity and let* $\mathcal{J}^{\pm} := \dot{J}^{\pm}(i^0) - i^0$*. Denote the future and past null infinities.*

2. *The complement of* $\phi(M)$ *in* $\tilde{M}$ *is the closure of the causal future and past of the spatial infinity i.e.* $\tilde{M} - \phi(M) = \overline{J(i^0)} := \overline{J^+(i^0)} \cup \overline{J^-(i^0)}$*.*

3. *There exists a neighborhood* V *of* $\{i^0\} \cup \mathcal{J}^+ \cup \mathcal{J}^-$ *in* $\tilde{M}$ *such that* $(V, \tilde{g})$ *is strongly causal and in* $M \cap V$*, the physical metric satisfies the vacuum Einstein equation,* $R_{\mu\nu} = 0$*.*

4. *There exist a function* Ω *on* $\tilde{M}$ *which is* C^2 *at* i^0 *and* C^∞ *everywhere such that on* $\phi(M)$*,* $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$*.*

   At null infinity: $\Omega|_{\mathcal{J}^{\pm}} = 0 \ , \quad \tilde{\nabla}_\mu \Omega\Big|_{\mathcal{J}^{\pm}} \neq 0.$

   At the spatial infinity: $\Omega(i^0) = 0, \ \lim_{i^0} \tilde{\nabla}_\mu \Omega = 0 \ , \ \lim_{i^0} \tilde{\nabla}_\mu \tilde{\nabla}_\nu \Omega = 2\tilde{g}_{\mu\nu}$*.*

5. *(a) The space of integral curves of* $\tilde{n}^\mu := \tilde{g}^{\mu\nu}\partial_\nu \Omega$ *on* $\mathcal{J}^{\pm}$ *is naturally diffeomorphic to the space of null directions at* i^0*.*

   (b) Given any smooth function ω *on* $\tilde{M} - i^0$ *which is strictly positive on* $M \cup \mathcal{J}^+ \cup \mathcal{J}^-$ *and* $\tilde{\nabla}_\mu(\omega^4 \tilde{n}^\mu)$ *vanishes on the null infinity, the vector field* $\omega^{-1}\tilde{n}^\mu$ *is complete on the null infinity*[2]*.*

An immediate and important observation which is easily checked is that the *conformal factor satisfying the conditions is not unique.* Any $\Omega' = \omega\Omega$ will satisfy all the conditions for all $\omega > 0$ which are smooth everywhere except possibly at i^0 where it can be $C^{>0}$ with $\omega(i^0) = 1$.

These are quite a few stipulations and elaborations are in order. The definition begins with an extension of the space-time of interest - the physical space-time - and stipulates the distinctive regions that must be present in the *conformally extended* (un-physical) space-time, namely spatial infinity and null infinity. The spatial infinity is postulated to be a *single point* which is distinguished by the possibility of modification of the smooth manifold structure of $\tilde{M}$ as well as smoothness properties of the metric $\tilde{g}$. There is no requirement of existence of the time-like infinities $i^{\pm}$ seen in the Minkowski space-time. One reason is that we may have a body existing in the infinite past and/or future, which will not be 'empty' at time-like infinities.

The un-usual $C^{>k}$ structures mean the following. The 'k' denotes that the

[2]In the following, we adopt the notation of putting a ˜ on quantities referring to the un-physical metric and/or quantities defined on the null infinity. Thus $\tilde{n}^\mu := \tilde{g}^{\mu\nu} n_\nu$, $\tilde{\xi}_\mu := \tilde{g}_{\mu\nu}\xi^\nu$ etc.

quantity such as coordinate transformations in overlapping charts or tensor fields, are k-times differentiable with the k^{th} partial derivatives being continuous. The '>' stipulates that the k^{th} partial derivatives are more than continuous in that the $(k+1)^{th}$ partial derivatives do have limits as i^0 is approached, but the limiting values depend on the direction along which the limit is taken. Nevertheless, these different limiting values which are thus functions of 'angular coordinates' labelling the directions, are smooth. Thus such quantities are more than continuous at the k^{th} level but less than full differentiability at the $(k+1)^{th}$ level [17, 39, 40]. Having allowed the metric to be $C^{>0}$, there are many possibilities for modified differential structures and this is restricted to be the $C^{>1}$.

Why such a complicated stipulation? This is essentially because the two-dimensional boundary of a three-dimensional Cauchy hypersurface is being 'compactified' into a single point. Explicit example of the Coulomb solution of Maxwell field in Minkowski space-time, exhibits this property. This is discussed further in the discussion of spatial infinity. Similarly, for the Schwarzschild solution and indeed any space-time associated with a massive body, one expects such a behaviour. This is the price to be paid for the 'one-point-compactification' of spatial infinity[3].

Condition (2) implies that what is 'added' in the extension is just the light cone at the spatial infinity together with its interior. It also means that the spatial infinity is related to all points of $\phi(M) \subset \tilde{M}$ in a space-like manner and the boundary of $\phi(M)$ in $\tilde{M}$ is precisely $i^0 \ \cup \mathscr{J}^+ \ \cup \mathscr{J}^-$.

Condition (3) incorporates the feature that *asymptotically* the vacuum Einstein equation holds - any non-zero matter stress tensor must vanish suitably in the vicinity of the 'added infinity'. This vicinity is also free of any closed or 'almost closed' causal curves.

Condition (4) stipulates that the extension of the physical space-time is a *conformal extension.* This in particular mean that the light cones of the two space-times agree on the image, $\phi(M)$. The conformal factor vanishes on the boundary of $\phi(M)$ as in the case of the Minkowski space-time. The vanishing of the conformal factor implies that physical intervals get *infinitely* stretched as the boundary is approached since the un-physical intervals remain finite. The stipulation on the derivatives, restricts the fall-off behaviours of the physical metric. We will see this in the next section.

The last conditions, (5), tie-up the spatial infinity and the null infinity in such a way that the boundary has the basic features of a light-cone of a point in Minkowski space-time *even though i^0 is not a regular point of a smooth manifold.* The first part of the condition ensures that all null geodesics emanating from i^0 span the null infinity. This would have been automatic if i^0 were a regular point, for which the null generators of its causal future/past are indeed null geodesics emanating from it. This implies that $\mathscr{J}^{\pm} \sim \mathbb{R} \times S^2$ [39,40].

[3]The [41] work gives a different definition of spatial infinity which is a two-dimensional manifold which does not have these differentiability conditions, but it looses the link with the null infinity.

The second part of the condition stipulates, that there are no 'missing points' in $\mathscr{J}^{\pm}$. This also has implications for the symmetries of the asymptotically flat space-times.

In summary, this definition of asymptotically empty and flat space-times, extends the physical space-times of this class by attaching a single point together with its causal past and future such that its light-cone is the boundary. The technical conditions such as a nonstandard differential structure, are needed to accommodate physically relevant solutions e.g. massive physical bodies, and imply a controlled form of asymptotic fall-off for the metric. The vanishing conformal factor accounts for infinite stretching and the 'completeness of the boundary' ensures that whatever escapes from the compact region, ends up on the boundary.

For use in subsequent discussion, we note that under a conformal scaling of the metric, the Ricci tensors of the two metrics are related as,

$$\begin{aligned} R_{\mu\nu} &= \tilde{R}_{\mu\nu} + 2\Omega^{-1}\tilde{\nabla}_\mu\tilde{\nabla}_\nu\Omega \\ &\quad + \tilde{g}_{\mu\nu}\,\tilde{g}^{\alpha\beta}\left(\Omega^{-1}\tilde{\nabla}_\alpha\tilde{\nabla}_\beta\Omega - 3\Omega^{-2}\tilde{\nabla}_\alpha\Omega\tilde{\nabla}_\beta\Omega\right) \qquad (7.3) \\ \Omega^{-2}R &= \tilde{R} + 6\Omega^{-1}\tilde{g}^{\mu\nu}\tilde{\nabla}_\mu\tilde{\nabla}_\nu\Omega - 12\Omega^{-2}\tilde{g}^{\mu\nu}\tilde{\nabla}_\mu\Omega\tilde{\nabla}_\nu\Omega \qquad (7.4) \end{aligned}$$

Let us see some of the implications of this definition.

7.1 Vicinity of the Null Infinity

Bondi gauge: As a first illustration, we see how the asymptotic analysis carried out by Bondi, Van der Burg and Metzner [43] can be derived from the abstract definition.

In V, we have $R_{\mu\nu} = 0$ and tilde quantities are well defined on the null infinity. Multiplying equations (7.3, 7.4) by Ω we see that the Ω^{-2} terms in both the equations must have a smooth limit to the null infinity. If we denote $\tilde{n}^\mu := \tilde{g}^{\mu\nu}\tilde{\nabla}_\nu\Omega$ then the last term implies that $\tilde{n}\cdot n = 0$ on the null infinity. $\tilde{n}^\mu$ is clearly normal to the $\Omega = 0$ surface i.e. to $\mathscr{J}^{\pm}$ which is light-like. Since we defined $\mathscr{J}^{\pm}$ as parts of the light-cone at i^0, this is a consistent consequence. Had we not defined $\mathscr{J}^{\pm}$ in this manner, but defined it only as a three-dimensional boundary where all null geodesics reach asymptotically in their affine parameter[4], we would have deduced that this boundary is necessarily a null surface. We can do more. Using the freedom to re-scale Ω by a positive ω, we can arrange $\Omega^{-1}\tilde{g}^{\alpha\beta}\tilde{\nabla}_\alpha\Omega\tilde{\nabla}_\beta\Omega = 0$. The required ω satisfies,

$$\tilde{n}\cdot\nabla\ln\omega = -\frac{1}{2}\Omega^{-1}\tilde{g}^{\alpha\beta}\tilde{\nabla}_\alpha\Omega\tilde{\nabla}_\beta\Omega \quad\Rightarrow\quad \tilde{\nabla}_\mu\tilde{\nabla}_\nu\Omega\Big|_{\mathscr{J}^{\pm}} = 0 \qquad (7.5)$$

[4]This is customary when asymptotic flatness is defined separately at null and spatial infinities.

The last implication follows because the second and the third terms on the right-hand side of $\Omega\times$(7.4) vanish. From this it follows immediately that null vectors $\tilde{n}^\mu$ satisfy the geodesic equation. The null geodesic congruence defined by these, trivially has zero shear, twist and expansion since $B_{\mu\nu} := \tilde{\nabla}_\mu\tilde{\nabla}_\nu\Omega = 0$. Vanishing twist implies that n^μ is hypersurface orthogonal (orthogonal to the $\Omega = 0$ hypersurface) and being null, it is also tangential to $\mathcal{J}^\pm$.

The condition (7.5) still allows further re-scalings by an ω that is *constant* along each of the null geodesics. Consider any two-dimensional submanifold of the future (resp. past) null infinity which is intersected by the null geodesics exactly once. Given the topology of the null infinity, such a cross-section must be a 2-sphere. The metric $\tilde{g}$ on $\mathcal{J}^\pm$ induces a Riemannian metric on the cross-section which is conformally related to the standard metric on S^2. The residual ω can now be chosen to take the metric on the sphere to have unit radius. The expansion and shear of the null generators being zero, completeness of the generators imply that the same metric can be transported to *all* the cross-sections.

The conformal factor satisfying $\tilde{\nabla}_\mu\tilde{\nabla}_\nu\Omega = 0$ on $\mathcal{J}^\pm$ and having unit sphere metric on the cross-sections, is called the *Bondi Gauge* or *Bondi conformal frame.*

The Bondi gauge, leads us to choose a coordinate system in $V \cap \mathcal{J}^\pm$. Consider $\mathcal{J}^+$ for definiteness. $\tilde{\nabla}\Omega \neq 0$ implies that Ω itself can be taken as a coordinate. Choosing any cross-section, introduce the standard angular coordinates θ, ϕ . Along each of the null geodesic generator of $\mathcal{J}^+$, assign the same angular coordinates and use an affine parameter u, of the geodesic as the third coordinate. The affine parameter may be initialized on the chosen cross-section and normalized as $\tilde{n}\cdot\tilde{\nabla}u = 1$. Each constant u cross-section of $\mathcal{J}^+$ is a sphere and has another null vector orthogonal to it at each of its points. Along null geodesics generated by these vectors, assign the same coordinates (u, θ, ϕ), the fourth coordinate being Ω ($\Omega = 0$ is $\mathcal{J}^+$). With this choice. the unphysical metric *on* $\mathcal{J}^+$ takes the form, $d\tilde{s}^2 = 2d\Omega du + d\theta^2 + \sin^2\theta d\phi^2$. The condition (5-b) implies that the coordinate u ranges over the full $\mathbb{R}$.

The gauge condition (7.5) on Ω implies that the metric components $\tilde{g}_{uu}, \tilde{g}_{u\theta}, \tilde{g}_{u\phi}$ all vanish with Ω as Ω^2. Hence, the asymptotic behaviour of the physical metric is determined from $g_{\mu\nu} = \Omega^{-2}\tilde{g}_{\mu\nu}$. With further coordinate transformations, it is possible to put the metric in the form postulated by Bondi et al. [43] in their analysis of gravitational radiation [17].

The Peeling Property: We have seen above the form of the physical metric as the null infinity is approached. In V, we already have the Ricci tensor to be zero. Geroch showed [44] that the Weyl tensor of the unphysical metric too vanishes on the null infinity. So the full Riemann tensor vanishes as the null infinity is approached. In V the two Weyl tensors are equal. Any null geodesic in $\tilde{M}$ landing on the null infinity with a bounded affine parameter corresponds to a null geodesic in M with an unbounded affine parameter $\lambda \to \infty$ as the null infinity is approached. Using the Bianchi identity, it is shown that the

Weyl curvature must behave as,

$$C_{\mu\nu\alpha\beta} = \frac{C^4_{\mu\nu\alpha\beta}}{\lambda} + \frac{C^3_{\mu\nu\alpha\beta}}{\lambda^2} + \frac{C^2_{\mu\nu\alpha\beta}}{\lambda^3} + \frac{C^1_{\mu\nu\alpha\beta}}{\lambda^4} + o(\lambda^{-5}) \ . \tag{7.6}$$

In the above, the $C^i_{\mu\nu\alpha\beta}$ are bounded and independent of λ and have the Petrov types IV (= N), III, II (or D) with a single, repeated principle null vector - the geodesic tangent - and the last one has type I with the geodesic tangent being *one of the* principle null directions. This property follows as a special case (spin 2) of the behaviour of spin-s massless fields in Minkowski space-time [45]. It played a role in the analysis of characterization of gravitational radiation. Petrov classification is discussed in section (14.8).

Asymptotic Symmetries at Null Infinity: The notion of asymptotic symmetries is naturally the set of diffeomorphism of the physical space-time which preserves the *asymptotic form of the metric.* The asymptotic form has been specified in terms of the definition of $\mathcal{J}^{\pm}$ and the conformal factor in the Bondi frame. Thus, symmetries at null infinity are the diffeopmorphisms of $\tilde{M}$ which induce a conformal diffeomorphism of $\mathcal{J}^{\pm}$ on the null infinity with a conformal factor constant along each null generator[5]. These symmetry transformations form a group known as the *Bondi-Metzner-Sachs* (BMS) group [47]. The infinitesimal generators of this group are conformal Killing vectors on the null infinity. The idea is to identify vector fields on the physical space-time which when extended to the null infinity boundary, become conformal Killing vectors there. There could be many ways of extending vector fields to the boundary. These are restricted by the demand that the asymptotic form in the neighborhood be preserved. This is done as follows.

We begin by noting a simple identity: For any vector field ξ^μ on $\tilde{M}$, on $\phi(M) \subset \tilde{M}$ we have,

$$\Omega^2 \mathcal{L}_\xi g_{\mu\nu} = \mathcal{L}_\xi \tilde{g}_{\mu\nu} - 2\Omega\xi \cdot \nabla(\Omega)\ g_{\mu\nu} \ = \ \mathcal{L}_\xi \tilde{g}_{\mu\nu} - 2\Omega^{-1}(\xi \cdot n)\tilde{g}_{\mu\nu} \tag{7.7}$$

where, $n_\mu := \tilde{\nabla}\Omega = \partial_\mu\Omega$. Observe that the right-hand side would be well defined on $\mathcal{J}$ *if* the vector field minimally satisfied $\xi^\alpha n_\alpha = \Omega\tilde{K}$ where $\tilde{K}$ is a smooth function on $\tilde{M}$ (at least on a neighborhood of the null infinity). Such a vector field, on the null infinity will be non-zero and tangential to $\mathcal{J}$ and thus will generate diffeomorphisms of $\mathcal{J}$. It would generate *conformal diffeomorphisms on the null infinity*, provided we demand that the right-hand side *vanishes* on the null infinity. Note that the physical metric is not defined on

[5]It is possible to identify symmetries in terms of an intrinsic description of the null infinity which is defined as 3 manifold with topology of $\mathbb{R} \times S^2$, a collection (q_{ij}, n^i) of a degenerate metric q_{ij} and a nowhere vanishing vector field n^i together with all the pairs of the form $(\omega^2 q_{ij}, \omega^{-1} n^i)$, $\omega > 0$ everywhere and satisfying: (i) n^i is the only degenerate direction of q_{ij}, (ii) $\mathcal{L}_n q_{ij} = 0$, (iii) the vector field n^i is complete with its orbit space diffeomorphic to S^2. Diffeomorphisms of the 3 manifold which preserve the collection of these pairs, are symmetries of the null infinity. Note that in this definition, n^i is not related to any gradient of any conformal factor. This group of symmetries is precisely the BMS group [46].

the conformal boundary and hence $\mathcal{L}_\xi g_{\mu\nu}$ is not well defined on the boundary either. It is nonetheless demanded that *after* multiplying by Ω^2, the Lie derivative be extendable to the boundary and should vanish there for ξ to qualify as a BMS generator. Thus, we define a vector field ξ^μ on $\tilde{M}$ to be a generator of an asymptotic symmetry if [48],

$$\text{(i)} \quad \xi^\alpha n_\alpha = \Omega\tilde{K} \quad \text{and} \quad \text{(ii)} \quad \mathcal{L}_\xi \tilde{g}_{\mu\nu} - 2\tilde{K}\tilde{g}_{\mu\nu} = \Omega\tilde{X}_{\mu\nu} \quad , \tag{7.8}$$

where $\tilde{K}$ and $\tilde{X}_{\mu\nu}$ are smooth fields on a neighborhood of the null infinity.

If ξ^μ is an isometry of the physical space-time, then the left-hand side of the identity (7.7) is exactly zero even off-$\mathcal{J}$ and therefore $\tilde{X}_{\mu\nu} = 0$ as well. $\tilde{K}$ may or may not be zero and thus ξ^μ generates an isometry ($\tilde{K} = 0$) or a conformal isometry ($\tilde{K} \neq 0$) of the un-physical space-time.

As an example, consider a vector field

$$\xi^\mu \ := \ \alpha\tilde{n}^\mu \ = \ \alpha\tilde{g}^{\mu\nu}n_\nu \ , \ \ n_\nu \ := \ \nabla_\nu\Omega \quad , \ \ \tilde{n}\cdot\nabla\alpha|_{\mathcal{J}} \ = \ 0 \ . \tag{7.9}$$

On $\mathcal{J}$, we also have $\tilde{g}^{\mu\nu}n_\mu n_\nu = 0$. In the vicinity of $\mathcal{J}$, We want to compute $\mathcal{L}_\xi\tilde{g}_{\mu\nu} = \tilde{\nabla}_\mu\tilde{\xi}_\nu + \tilde{\nabla}_\nu\tilde{\xi}_\mu$, $\tilde{\xi}_\mu := \tilde{g}_{\mu\nu}\xi^\nu$ and identify $\tilde{K}, \tilde{X}_{\mu\nu}$ to check if ξ is a BMS generator. Since $n\cdot n$ and $n\cdot\nabla\alpha$ vanish on $\mathcal{J}$, away from it they must have appropriate factors of Ω. To determine these factors, recall the equations (7.3, 7.4) and use $R_{\mu\nu} = 0$ which is valid in the vicinity. This leads to,

$$\tilde{R}_{\mu\nu} - \frac{\tilde{R}}{6}\tilde{g}_{\mu\nu} \ = \ - \ \Omega^{-1}(\tilde{\nabla}_\mu n_\nu + \tilde{\nabla}_\nu n_\mu) + \Omega^{-2}\tilde{g}_{\mu\nu}\tilde{n}\cdot n \tag{7.10}$$

Since the left-hand side has a smooth limit to $\mathcal{J}$, we must have $\tilde{\nabla}_\mu n_\nu + \tilde{\nabla}_\nu n_\mu := \Omega\tilde{Y}_{\mu\nu}$ and $n\cdot n := \Omega^2\tilde{\rho}$ for some tensors $\tilde{Y}_{\mu\nu}$, $\tilde{\rho}$ which have smooth limit on $\mathcal{J}$. For the gradient of α, it suffices to have $\tilde{n}^\mu\tilde{\nabla}_\mu\alpha := \Omega\tilde{\beta}$.

Substituting for ξ, we get, $\mathcal{L}_\xi\tilde{g}_{\mu\nu} = \tilde{\nabla}_\mu(\alpha n_\nu) + \mu \leftrightarrow \nu \ = \ \Omega\alpha\tilde{Y}_{\mu\nu} + n_\mu\tilde{\nabla}_\nu\alpha + n_\nu\tilde{\nabla}_\mu\alpha$. While the first term has a smooth limit to the null infinity, the last two terms do not! Thus the candidate vector field is *not* a BMS generator. We can get rid of the offending terms by subtracting $\Omega\tilde{g}^{\mu\nu}\tilde{\nabla}_\nu\alpha$ from the ξ. This leads to, (now $\tilde{\xi}_\mu := \alpha n_\mu - \Omega\tilde{\nabla}_\mu\alpha$)

$$\begin{aligned} \xi^\mu n_\mu &= \Omega^2\tilde{\rho}\alpha - \Omega^2\tilde{\beta} &\Rightarrow \quad \tilde{K} &= \Omega(\alpha\tilde{\rho} - \tilde{\beta}) \qquad (7.11) \\ \mathcal{L}_\xi\tilde{g}_{\mu\nu} &= \Omega\alpha\tilde{Y}_{\mu\nu} - 2\Omega\tilde{\nabla}_\mu\tilde{\nabla}_\nu\alpha &\Rightarrow \quad \tilde{X}_{\mu\nu} &= \alpha\tilde{Y}_{\mu\nu} - 2\tilde{\nabla}_\mu\tilde{\nabla}_\nu\alpha - 2(\alpha\tilde{\rho} + \tilde{\beta}) \end{aligned}$$

$\tilde{Y}_{\mu\nu}$ may be substituted in terms of curvature from (7.10).

Notice that it is possible to have two different vector fields on $\tilde{M}$ which are *equal* on the null infinity (e.g. the above example) in which case they generate the same infinitesimal conformal transformation on $\mathcal{J}$ and should be identified as the same generator. Hence we define that *two BMS generators are equivalent if they are equal on* $\mathcal{J}$. How is this freedom characterized?

The difference of two BMS generators vanishes on $\mathcal{J}$ and hence must be of the form: $\xi_1^\mu - \xi_2^\mu = \Omega\zeta^\mu$. Substitution in the identity (7.7) and evaluating it on the null infinity, leads to $\zeta^\mu|_{\mathcal{J}} = 0$ i.e. $\zeta^\mu = \Omega u^\mu$ or difference of two BMS generators is proportional to $\Omega^2 u^\mu$ for some vector field u^μ in a neighborhood of null infinity. The identity then implies, $(\delta\xi^\mu := \Omega^2 u^\mu)$

$$\delta\tilde{K} = \Omega u \cdot n \quad , \quad \delta\tilde{X}_{\mu\nu} = n_\mu u_\nu + n_\nu u_\mu - \tilde{n} \cdot \tilde{u}\ \tilde{g}_{\mu\nu} + \frac{\Omega}{2}\left(\tilde{\nabla}_\mu \tilde{u}_\nu + \tilde{\nabla}_\nu \tilde{u}_\mu\right) \ .$$

All the above equations refer to the un-physical metric explicitly. What happens when this is changed by the allowed scaling by ω^2? The same vector field is again a BMS generator but now $\tilde{K}$ and $\tilde{X}_{\mu\nu}$ change as,

$$\tilde{K}' = \tilde{K} + \omega^{-1}\xi^\mu \tilde{\nabla}_\mu \omega \quad , \quad \tilde{X}'_{\mu\nu} = \omega \tilde{X}_{\mu\nu} \ .$$

This completes the definition of infinitesimal generators of the BMS group together with the two sets of gauge ambiguities.

The BMS generators are tangential to the null infinity. Hence there is a subclass of them so that on $\mathcal{J}$, $\xi^\mu = \alpha n^\mu$, $n \cdot \tilde{\nabla}\alpha = 0$. Clearly since α is constant along the integral curves, these generate *isometries* (of the induced metric) on $\mathcal{J}$. These are termed *super-translations*, they translate along the null generators in an angle dependent way since α is a smooth function of the angles. It turns out that these form an infinite-dimensional, Abelian, normal subgroup of the BMS group and the quotient group is isomorphic to the *Lorentz group*[6].

Conservation Laws at Null Infinity: Do these asymptotic symmetries also lead to any corresponding 'conserved quantities'? What would such a notion mean (at this stage, we are not using any action formulation to appeal to Noether theorem for a conservation law nor are the symmetries defined as invariance of any action)? To appreciate it, consider first the case of physical space-time admitting a Killing vector, ξ^μ.

If we have non-vanishing matter stress tensor with, $\nabla_\mu T^{\mu\nu} = 0$, then we define $J^\mu := T^{\mu\nu}\xi_\nu$ which is also covariantly conserved and as discussed before (see 4.10), we can define for a 3-manifold Σ, a 'conserved charge', $Q_\xi(\Sigma) := \int_\Sigma J \cdot n\ ds$. The conservation aspect follows by noting that if Σ_1, Σ_2 are two hypersurfaces which bound a 4-region $\mathcal{V}$, then $Q_\xi(\Sigma_1) = Q_\xi(\Sigma_2)$. For a hypersurface orthogonal, time-like Killing vector, Σ_i being surfaces of orthogonality are the natural choice. This is conserved 'energy' associated with *matter*. What happens in the absence of matter, as in the case of asymptotically empty space-time? Here we note a few mathematical relations, (please see section 14.6). Using $\xi := g_{\mu\nu}\xi^\nu dx^\mu$,

$$\begin{aligned} \alpha \ &:= \ *d\xi \ \Rightarrow \ d\alpha \ = \ *\,\delta d\xi \ ; \ \text{and for a Killing 1-form,} \\ (\delta d\xi)_\mu \ &= \ \nabla^\nu\left(\nabla_\mu \xi_\nu - \nabla_\nu \xi_\mu\right) \ = \ 2R_{\mu\nu}\xi^\nu \end{aligned} \tag{7.12}$$

[6]There is a further unique, four-dimensional subgroup of the super-translations, namely that generated by super-translations with $\alpha(\theta,\phi)$ spanned by the four $Y_{lm}(\theta,\phi)$ with $l = 0, 1$, which is also a *normal subgroup of the BMS group*. For the Minkowski space-time, it corresponds to translations [17]. This plays a role in the definition of the *Bondi Energy-Momentum* discussed below.

In the last equality, we have used the Killing equation.

Notice that α is a 2-form and $d\alpha$ is a 3-form which vanishes in the region where the Ricci tensor vanishes. Therefore, if S_1, S_2 are two 2-surfaces without boundary, which bound a 3-region where the Ricci tensor vanishes, then $Q_S := \int_S \alpha$, has the same value for S_1, S_2. For vacuum solutions with isometries, these 'charges' are the *Komar Integrals* [49]. Specifically, for stationary and/or axisymmetric vacuum solutions, these define the mass and the angular momentum *of the space-time* [17]:

$$M_S := \frac{1}{8\pi G}\int_S *d\xi = -\frac{1}{8\pi G}\int_S ds^{\mu\nu}\epsilon_{\mu\nu\alpha\beta}\nabla^\alpha\xi^\beta \ ; \qquad (7.13)$$

$$J_S := -\frac{1}{16\pi G}\int_S *d\xi = +\frac{1}{16\pi G}\int_S ds^{\mu\nu}\epsilon_{\mu\nu\alpha\beta}\nabla^\alpha\psi^\beta \ . \qquad (7.14)$$

Here, S is a space-like topological 2-sphere surrounding matter sources, lying in the asymptotic region where the Ricci tensor vanishes and the integrals are independent of the sphere S. The ξ is a time-like Killing vector, normalized as $\xi \cdot \xi = -1$ at infinity while ψ is the space-like Killing vector whose orbits are closed curves, normalized so that the Killing parameter ranges over $[0, 2\pi]$.

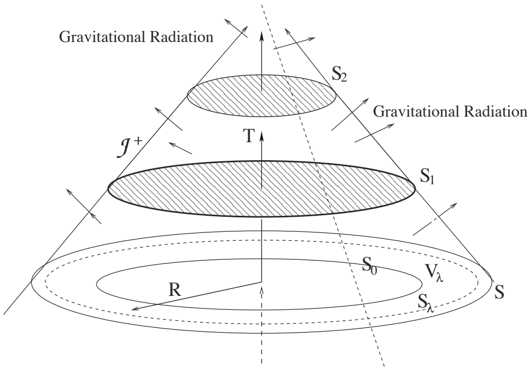


FIGURE 7.2: In the lowest slice, S is the limiting member of a family of S_λ used in defining the conserved charge associated with a BMS generator. The circles S_1, S_2 are intersections of the shaded slices with $\mathcal{J}^+$ and bound the region $\mathcal{V}$ of $\mathcal{J}^+$ over which the flux is integrated.

Thus we see that for an asymptotically Ricci flat space-time with isometries, we can define conserved quantities (the Komar integrals) which enable us to evaluate them 'on a sphere at infinity' in a limiting sense. The question now is whether analogous quantities can be defined when we have only asymptotic symmetries of the BMS group.

Recall that a vector field ξ^μ on the un-physical space-time which is tangential on $\mathcal{J}$ and generates conformal transformation on $\mathcal{J}$ is a representative of a BMS generator and there are two types of gauge freedoms. Consider the integrands in the Komar integrals but express it in terms of the un-physical metric $\tilde{g}_{\mu\nu}$.

$$\begin{aligned}
\epsilon_{\mu\nu\alpha\beta}\nabla^{\alpha}\xi^{\beta} &= \epsilon_{\mu\nu\alpha\beta}g^{\alpha\rho}g^{\beta\sigma}\nabla_{\rho}\xi_{\sigma} \;=\; \mathcal{E}_{\mu\nu\alpha\beta}\sqrt{g}g^{\alpha\rho}g^{\beta\sigma}\partial_{\rho}\xi_{\sigma} \\
&= \mathcal{E}_{\mu\nu\alpha\beta}\Omega^{-4}\Omega^{4}\sqrt{\tilde{g}}\,\tilde{g}^{\alpha\rho}\,\tilde{g}^{\beta\sigma}\partial_{\rho}\Omega^{-2}\tilde{\xi}_{\sigma} \;=\; \tilde{\epsilon}_{\mu\nu\alpha\beta}\tilde{g}^{\alpha\rho}\tilde{g}^{\beta\sigma}\tilde{\nabla}_{\rho}\Omega^{-2}\tilde{\xi}_{\sigma} \\
&= \tilde{\epsilon}_{\mu\nu\alpha\beta}\tilde{\nabla}^{\alpha}\Omega^{-2}\tilde{\xi}^{\beta} \quad , \quad \tilde{\xi}^{\beta} = \xi^{\beta} \qquad (7.15)
\end{aligned}$$

If we were to attempt to adopt the Komar integral on *a cross-section, S, of* $\mathcal{J}$ as a candidate conserved quantity, we immediately face a problem with the Ω^{-2}. However, we can consider a cross-section, S, being approached as a limiting sphere from a family, S_λ, of topological 2-spheres from a neighborhood of the null infinity. Being away from $\mathcal{J}$, Ω is non-zero on each member of the family. Since ξ is not a Killing vector, $d * d\xi$ is non-zero even in the Ricci flat region. Nevertheless, we can still use the Stoke's theorem. To do this, consider a 3-region, V_λ, bounded by a fixed 2-sphere, S_0, in the physical space-time and a member S_λ. The Stokes' theorem then gives,

$$\int_{S_\lambda} ds^{\mu\nu}\tilde{\epsilon}_{\mu\nu\alpha\beta}\tilde{\nabla}^{\alpha}(\Omega^{-2}\tilde{\xi}^{\beta}) \;=\; \int_{S_0} ds^{\mu\nu}\epsilon_{\mu\nu\alpha\beta}\nabla^{\alpha}\xi^{\beta} + \int_{V_\lambda} *\delta d\xi \qquad (7.16)$$

$$\begin{aligned}
\text{where} \quad (\delta d\xi)_{\mu} &= \tilde{\nabla}^{\nu}\left(\tilde{\nabla}_{\mu}\Omega^{-2}\tilde{\xi}_{\nu} - \tilde{\nabla}_{\nu}\Omega^{-2}\tilde{\xi}_{\mu}\right) \qquad (7.17) \\
&= 2\Omega^{-1}\left(-\tilde{\nabla}^{\nu}\tilde{X}_{\mu\nu} + \tilde{\nabla}_{\mu}\tilde{X}^{\nu}{}_{\nu} + 3\Omega^{-1}\tilde{X}_{\mu\nu}\tilde{n}^{\nu}\right)
\end{aligned}$$

Here we have used, $\Omega^{-2}\tilde{\xi}_{\mu} = \Omega^{-2}\tilde{g}_{\mu\nu}\xi^{\nu} = \xi_{\mu}$ and in the second line we have used equation (7.12). The last equality follows by noting the general identity,

$$\nabla^{b}\left(\nabla_{a}v_{b} - \nabla_{b}v_{a}\right) = 2R_{ab}v^{b} + 2\nabla_{a}(\nabla\cdot v) - \nabla^{b}\left(\nabla_{a}v_{b} + \nabla_{b}v_{a}\right) \;,$$

and applying it to $v_{\mu} = \omega^{-2}\tilde{\xi}_{\mu}$. The curvature is eliminated using (7.10) and $\tilde{X}_{\mu\nu}$ enters from the last two term, using the conditions (7.8) for ξ^{μ} to be a BMS generator.

The integral in the first term of eq. (7.16) is well defined and fixed. *If* the *integrand* of the second term remains finite in the limit $S_\lambda \to S$, then the integral too will remain finite. It is shown in [48] that this is indeed so and therefore the integral on the cross-section is well defined. Furthermore, it is also independent of family of spheres approaching the same cross-section S since the integral over S_0 and the interpolating region V are independent of the family chosen. This integral has been termed as a '*linkage*' [48].

Now to address the gauge ambiguities. Under the conformal scaling by ω^2, one can see explicitly that the integral is unchanged. However, as the representative of the BMS generator is changed, the integral changes! This is remedied by stipulating that the representative be required to satisfy a further condition, namely, $\tilde{\nabla}_{\mu}\tilde{\xi}^{\mu} = 0$ [17,48]. Thus we have achieved a goal of associating a gauge invariant quantity with each BMS generator with linear dependence. *What is 'conserved' about it?* It is a conserved quantity in the sense that it can be again be evaluated on a 'sphere at infinity' in a limiting sense, independent of the way the limiting family is chosen, just as in the case of the exact symmetries. For exact symmetry, the value of the integral remains the *same* for all members of the family while this is not so for the asymptotic symmetries. For asymptotic symmetries, the value does depend on the limiting sphere on $\mathcal{J}$.

Is it possible to relate the values at two different cross-sections? The answer turns out to be yes.

Recall the mathematical relations (7.12). The Stokes' theorem already gives us $\int_V *\delta d\xi = \int_{\partial V} *d\xi$. Consider a region V in the vicinity of $\mathcal{J}^+$ (say) which is bounded by two 2-spheres S_0, S_0'. These bounding spheres will be taken to two cross-sections S, S' of $\mathcal{J}^+$ respectively. We have already defined a conserved quantity (linkage integral) associated with a BMS generator and a cross-section: $Q_S(\xi) := \int_S \tilde{\epsilon}_{\mu\nu\alpha\beta}\tilde{\nabla}^\alpha(\Omega^{-2}\tilde{\xi}^\beta)$. The volume integral in the Stoke's theorem can be written as,

$$\begin{aligned}
\int_V *\delta d\xi &= \int_V \frac{dx^\mu \wedge dx^\nu \wedge dx^\alpha}{3!}\tilde{\epsilon}_{\mu\nu\alpha}{}^{\beta}\tilde{\nabla}^\rho\left(\tilde{\nabla}_\beta\Omega^{-2}\tilde{\xi}_\rho - \tilde{\nabla}_\rho\Omega^{-2}\tilde{\xi}_\beta\right) \\
&= \frac{1}{3!}\int_V d^3x\mathcal{E}^{\mu\nu\alpha\sigma}n_\sigma\tilde{\epsilon}_{\mu\nu\alpha}{}^{\beta}\tilde{\nabla}^\rho\left(\tilde{\nabla}_\beta\Omega^{-2}\tilde{\xi}_\rho - \tilde{\nabla}_\rho\Omega^{-2}\tilde{\xi}_\beta\right) \\
&= \int_V d^3x\,\sqrt{\tilde{h}}\tilde{n}^\sigma\tilde{\nabla}^\rho\left(\tilde{\nabla}_\sigma\Omega^{-2}\tilde{\xi}_\rho - \tilde{\nabla}_\rho\Omega^{-2}\tilde{\xi}_\sigma\right) \\
\therefore \int_V *\delta d\xi &:= \int_V d^3x\sqrt{\tilde{h}}F
\end{aligned} \tag{7.18}$$

In the above, we have used $\mathcal{E}^{\mu\nu\alpha} := \mathcal{E}^{\mu\nu\alpha\sigma}n_\sigma$, where n_σ is the normal to the region V regarded as a hypersurface in $\tilde{M}$ and $\tilde{h}$ is the determinant of the induced metric on this hypersurface. Note that away from $\mathcal{J}$, $\tilde{n}\cdot n \neq 0$ and the induced metric is non-degenerate too.

We have already simplified part of the expression for the *flux F* defined above, in eqn. (7.17,7.18). Using $\tilde{n}^\sigma = \tilde{\nabla}^\sigma\Omega$, keeping only the terms that survive in the limit of going to $\mathcal{J}^+$ and using the gauge condition $\tilde{\nabla}\cdot\tilde{\xi} = 0$ on the representative of BMS generator, a convenient and gauge invariant expression of flux is [48]:

$$F = -\tilde{\nabla}^\mu\tilde{\nabla}^\nu\tilde{X}_{\mu\nu} + 3\tilde{\nabla}^\mu\tilde{X}_\mu + \frac{3}{4}\tilde{\nabla}^2\tilde{X} + \frac{1}{24}\tilde{R}\tilde{X} \quad , \quad \tilde{X}_\mu := \Omega^{-1}\tilde{X}_{\mu\nu}\tilde{n}^\nu \tag{7.19}$$

As an illustration, the flux corresponding to the *super-translation* (7.9) is given by [48],

$$\begin{aligned}
F &= \frac{1}{4}\alpha^{-1}\left(\tilde{\nabla}^2 - 2\Omega^{-1}\tilde{n}^\mu\tilde{\nabla}_\mu\right)\tilde{H} - 2\alpha^{-1}\tilde{X}^{\mu\nu}\tilde{X}_{\mu\nu} + \alpha^{-1}\tilde{X}^2 \\
\tilde{H} &:= \alpha\tilde{\nabla}^2\alpha + \alpha^2\left(\frac{1}{6}\tilde{R} + \tilde{\rho}\right) - \tilde{\nabla}^\mu\alpha\,\tilde{\nabla}_\mu\alpha - 2\alpha\tilde{\beta}
\end{aligned} \tag{7.20}$$

A further specialization to *translations*, where α is restricted to linear combination of only $l = 0, 1$ spherical harmonics, $\tilde{H}$ is a constant and $\tilde{X}_{\mu\nu} - \frac{1}{2}\tilde{X}\tilde{h}_{\mu\nu} =: \alpha\tilde{N}_{\mu\nu}$ where $\tilde{h}_{\mu\nu}$ is the induced, degenerate metric on $\mathscr{J}$ and $\tilde{N}_{\mu\nu}$ is the Bondi news function. The flux then reduces to

$$F = \alpha\tilde{N}^{\mu\nu}\tilde{N}_{\mu\nu} \ .$$

7.2 Vicinity of the Spatial Infinity

The discussion of this section is based on [39, 40].

This region is more complicated to analyse because approaching infinity along all spatial directions is squeezed into a single point i^0 at which the differentiability conditions are also complicated. That such a squeezing necessarily implies non-smooth behaviour of physical fields can be seen as follows. Consider the electric field of a charge moving uniformly in an otherwise empty, flat Euclidean space or equivalently following a time-like geodesic in the Minkowski space-time. Any spatial hyperplane together with i^0 is topologically a 3-sphere (3) in $\tilde{M}$ and therefore the total charge on this compact manifold without boundary which equals the divergence of the electric field integrated over the 3-manifold must vanish. But we have a charge inside the volume. So at i^0 there much be an effective image charge if Maxwell equations are to hold for appropriately scaled Maxwell fields on $\tilde{M}$. But this also means that the Maxwell fields must diverge in a direction dependent manner at i^0. Ashtekar and Hansen give a detailed discussion of the appropriateness of the chosen non-standard smoothness requirements at i^0.

To have an analogue of the Bondi coordinate system and for discussion of asymptotic symmetries and conserved charges, a 'blown-up' model of i^0 is needed. This means i^0 is to to be understood as another manifold together with some additional structures chosen such that the physical fields will be smooth on this manifold with smoothness properties corresponding to the differentiable structure at i^0. To appreciate this, let us note a few points.

The idea of spatial infinity is to characterize the different ways in which one may go far away from localized sources, in a space-like manner. This may be done by selecting spatial hypersurfaces and then going to infinity in any direction *or* by simply following space-like curves directed away from the sources. The curves should be in-extendible to capture the sense of 'reaching to infinity'. The later is in spirit similar to the null infinity being understood as 'end-points' of null geodesics. The space-like curves, in the physical space-time, are smooth (C^∞) however in the conformally extended space-time the available smoothness at i^0 is only $C^{>1}$. This means that in extending a space-like curve to spatial infinity, it is meaningless to demand higher degree of smoothness and thus it appropriate to define *equivalence classes of curves* which agree only upto 'second order'. The *set* S *of such equivalence classes* can be given a (smooth) manifold structure and this manifold serves as a blown-up model for the i^0 - the smooth fields on this manifold correspond to the fields on the conformal infinity with direction dependent limits at i^0. The specific details and discussion of this construction are best seen in Ashtekar and Hansen. The upshot is that the spatial infinity can be equivalently described as a *four-dimensional manifold S, called* Spi *for spatial infinity, which is a principal fibre bundle with the base manifold $\mathcal{K}$ given as the unit, time-like hyperboloid in the tangent space at i^0 and the additive group of reals as the structure group.* Spi inherits two tensor fields: A degenerate metric h_{ab} which is the pull-back of the natural metric on $\mathcal{K}$ ($\pi : S \to \mathcal{K}$) and a vertical vector field v^a generating the one parameter diffeomorphisms induced by the action of the structure group[7]. The appendix C of the Ashtekar-Hansen paper also gives the explicit form of Kerr metric near spatial infinity as well as gives a general form of the physical metric which can be conformally extended such that it will satisfy the local condition at i^0.

Having gotten the Spi structure in a form similar to that of null infinity, the notions of asymptotic symmetries and conserved charges proceeds similar to that in the case of the null infinity. The infinitesimal diffeomorphisms which preserve the structure of Spi turn out to be characterized by vector fields ξ on S which satisfy, (a) $\mathcal{L}_{\bar{\xi}}\bar{g}_{ab} = 0$ on $\mathcal{K}$ and (b) $\mathcal{L}_{\xi}v^a = 0$ on S. Here $\bar{g}$ is the natural metric on the base manifold $\mathcal{K}$ and $\bar{\xi}$ is the projection of ξ on it. The Lie algebra of these vector fields is the Lie algebra of Spi. The special case where the projection $\bar{\xi}$ vanishes i.e. ξ^a itself is a vertical field proportional to v^a, turn out to form an invariant Abelian sub-algebra and constitute the *Spi super-translations.* The further special case wherein the proportionality function f in $\xi^a = fv^a$, is linear in the position vectors, η^a on $\mathcal{K}$ i.e. $4f = k_a\eta^a$ for some co-vector k_a on the hyperboloid, constitute the *Spi translations.* The quotient of the Spi Lie algebra by the super-translations, is isomorphic to the isometries of the unit time-like hyperboloid $\mathcal{K}$ which is the Lorentz algebra, *Spi/supertranslations ∼ Lorentz*. This is very similar to the BMS symmetries. These structures extend to the finite diffeomorphisms to i.e. to groups.

The conserved charges corresponding to the Spi generators are constructed analogously and depend on the asymptotic forms of various fields. The spatial infinity is completely decoupled from the dynamics (no causal relations) and hence the conserved quantities are constants characterising the asymptotically flat space-times. Furthermore, there are only 4 non-trivial conserved quantities—the 4-momentum which corresponds to the *translations* and the charges corresponding to the other super-translations vanish. Likewise, if Maxwell fields are included, the non-trivial charges are the electric and the magnetic charge only. These results follow from the detailed asymptotic forms of the gravitational and the Maxwell fields. To define angular momentum

[7]The null infinities $\mathcal{J}^{\pm}$ too have a similar structure a degenerate metric and the null normal fields. The base manifold is however an s^2 and the null infinities themselves are 3-manifolds.

though, further restrictions on the asymptotic form of the Weyl tensor need to be imposed.

This concludes our discussion of the asymptotically flat space-times, their symmetries and associated conserved quantities.

The case of non-zero cosmological constant has not been analysed as extensively [50]. The asymptotic structure of the de Sitter (dS) and the Anti de Sitter (AdS) space-times is well known [18]. With the cosmological models favouring $\Lambda > 0$ (De Sitter) space time, there is some motivation to consider the asymptotically De Sitter space-times. The AdS case has been analysed in more details thanks to the theoretical interest in the AdS/CFT correspondence.

Chapter 8

Black Holes

Among the solutions of space-times with compact sources are the Black Hole solutions. The very first Schwarzschild solution provided the initial model of space-time near the Sun with which general relativity passed its first tests. Its *mathematical extension* (decreasing the radial coordinate, first below the physical radius of the star and then below the Schwarzschild radius) already revealed the exotic nature of the space-time of a point mass. We have seen the Kruskal extension of the Schwarzschild solution and indicated the similar one for the Reissner–Nordstrom solution. These and the Kerr–Newmann family of solutions are all asymptotically flat. The portion of space-time connected to the asymptotic region is the *exterior region.* The mathematical extension refers to extension away from the asymptotic region, in the *interior region.* The different regions are signalled in terms of coordinates where some of the metric components vanish or diverge and are demarcated by the zeros of the function $\Delta(r) = (r - r_+)(r - r_-)$ where $r_\pm$ are constants. As noted before, although some of the metric components vanish or diverge, the Riemann tensor - which encodes the physical effects of gravity - is perfectly well behaved. The geodesics across the $r = r_\pm$ surfaces are well behaved too and in fact signal how an extension may be sought. Fundamentally, an extension across a chart boundary is sought by changing the local coordinates, obtaining the corresponding metric and continuing the same metric form to a larger neighborhood till the next chart boundary where the extended metric or the curvature may develop singularities.

We now illustrate the method for the Kerr–Newman family and then discuss more general black holes and briefly touch upon their further generalization to isolated and dynamical horizons.

8.1 Examples of Extended Black Hole Solutions

Let us recall the metric of the Kerr–Newman solution in two different forms $(a \neq 0, Q \neq 0\ , M^2 > a^2 + Q^2)$,

$$ds^2 = -\frac{\eta^2\Delta}{\Sigma^2}dt^2 + \frac{\Sigma^2 sin^2\theta}{\eta^2}(d\phi - \omega dt)^2 + \frac{\eta^2}{\Delta}dr^2 + \eta^2 d\theta^2 \qquad (8.1)$$

$$= -\frac{\Delta}{\eta^2}\left\{dt - asin^2\theta d\phi\right\}^2 + \frac{sin^2\theta}{\eta^2}\left\{(r^2+a^2)d\phi - adt\right\}^2$$
$$+ \frac{\eta^2}{\Delta}dr^2 + \eta^2 d\theta^2 \quad \text{where,} \tag{8.2}$$
$$\Delta := r^2 + a^2 - 2Mr + Q^2 \quad , \quad \eta^2 := r^2 + a^2 cos^2\theta$$
$$\Sigma^2 := \left(r^2+a^2\right)^2 - a^2 sin^2\theta\Delta \quad , \quad \omega := \frac{a(2Mr - Q^2)}{\Sigma^2} \tag{8.3}$$

As mentioned before, there are coordinate singularities at the zeros of the $\Delta(r)$ function while at $r = 0$ there is a curvature singularity. The two roots of $\Delta(r) = 0$ are: $r_\pm = M \pm \sqrt{M - a^2 - Q^2}$ which split the range of r into three segments

$$(A): \quad -\infty < r < r_- \quad , \quad (B): \quad r_- < r < r_+ \quad , \quad (C): \quad r_+ < r < \infty \ .$$

Note that for the Kerr–Newman family, r is *not* the areal radial coordinate and is not required to be positive. The curvature singularity occurs when $\eta^2 = 0$ which in turn happens at $r = 0$ *and* $\theta = \pi/2$. Since $r = 0$ is a curvature singularity, one may suspect that negative r is excluded. This is not the case since the curvature blows up along a 'ring' in the equatorial plane $\theta = \pi/2$. This is most readily seen in the so-called *Kerr-Schild* form of the metric. It is therefore possible to continue through the '$r = 0$' singular space-time cylinder [17, 29]. For contrast, the '$r = 0$' singularity in the spherically symmetric Schwarzschild and Reissner–Nordstrom solutions is a sphere of radius zero (or a line in the space-time).

Observe that along $\theta = 0, \pi$ submanifolds, the metric is same as that of the spherically symmetric Reissner–Nordstrom solution ($a = 0$). Hence the extension across the three regions can be done in the same manner. We have already given the tortoise coordinate r_* defined by $dr_* := \frac{r^2}{(r-r_+)(r-r_-)}dr$ which leads to,

$$r_*(r) = r + \frac{r_+^2}{r_+ - r_-}\ln\left|\frac{r}{r_+} - 1\right| - \frac{r_-^2}{r_+ - r_-}\ln\left|\frac{r}{r_-} - 1\right| \tag{8.4}$$

Here we have chosen $r_*(0) = 0$ arbitrarily. In terms of this coordinate, the two-dimensional metric is conformal to the two-dimensional Minkowski metric: $ds^2 = \frac{\Delta}{r^2}(-dt^2 + dr_*^2)$. The radial null geodesics are given by $t = \pm r_*$. In the three regions we have $(r_+, \infty) \leftrightarrow r_* \in (-\infty, \infty)$, $(r_-, r_+) \leftrightarrow r_* \in (\infty, -\infty)$ and $(-\infty, r_-) \leftrightarrow r_* \in (-\infty, \infty)$. Introduce $u := \epsilon_u(t - r_*)$, $v := \epsilon_v(t + r_*)$, $\epsilon_{u,v} = \pm 1$ so that $dt^2 = -(\Delta/r^2)\epsilon_u\epsilon_v dudv$. In regions C and A, $\Delta > 0$ hence the signature of the metric requires that $\epsilon_u = \epsilon_v = \pm 1$ while in region B, Δ being negative requires $\epsilon_u = -\epsilon_v = \pm 1$. We have thus 6 possible choices labelled as $A_\pm, B_\pm$ and $C_\pm$ which are detailed in the equation (8.5) below. In each of the six blocks, the u, v coordinates range over $(-\infty, \infty)$. These ranges can be brought to $(-\pi/2, 0)$, $(0, \pi/2)$ by introducing new coordinates $U(u), V(v)$ suitably in each of the blocks. These are to be chosen so that the metric takes the same form and an extension is obtained by matching the individual chart boundaries. The following definitions – which are little different from [29] – achieve this. Following [29], the diagram is first constructed for $\theta = 0, \pi$ and then extended to other values of θ. Across different chart boundaries, different definitions of ϕ are needed. The final resulting Penrose diagram is shown in figure (8.1).

$$
\begin{array}{llll}
A_+ & : & u = t - r_* & , \quad tanU := e^{-\alpha u} \\
 & : & v = t + r_* & , \quad tanV := e^{\alpha v} \\
A_- & : & u = -t + r_* & , \quad tanU := e^{\alpha u} \\
 & : & v = -t - r_* & , \quad tanV := e^{-\alpha v} \\
B_+ & : & u = t - r_* & , \quad tanU := -e^{-\alpha u} \\
 & : & v = -t - r_* & , \quad tanV := -e^{\alpha v} \\
B_- & : & u = -t + r_* & , \quad tanU := e^{\alpha u} \\
 & : & v = t + r_* & , \quad tanV := e^{\alpha v} \\
C_+ & : & u = t - r_* & , \quad tanU := -e^{-\alpha u} \\
 & : & v = t + r_* & , \quad tanV := e^{\alpha v} \\
C_- & : & u = -t + r_* & , \quad tanU := e^{\alpha u} \\
 & : & v = -t - r_* & , \quad tanV := -e^{-\alpha v}
\end{array}
\tag{8.5}
$$

For the special case of Reissner–Nordstrom, the $r = 0$ is a curvature singularity and the portions B' and B in the right portion of figure (8.1) are absent. For the special case of Schwarzschild, the two roots of $\Delta(r)$ coincide and the

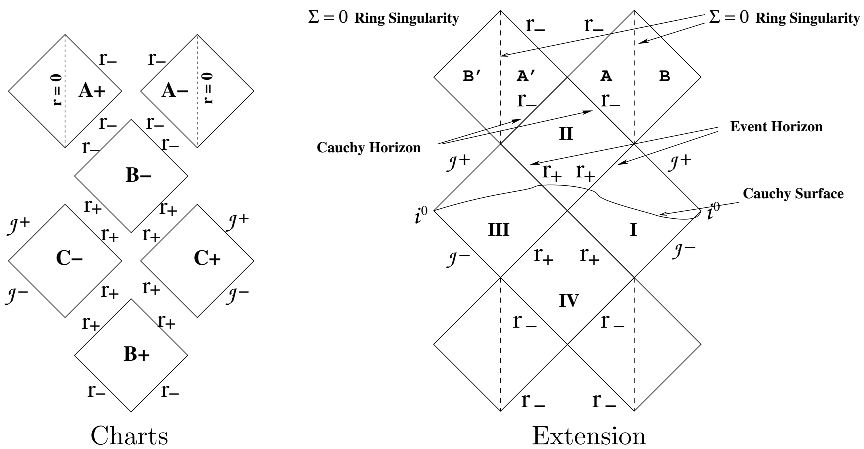


FIGURE 8.1: The t-axis goes from bottom to top, the r_*-axis goes from left to right and the metric is conformal to the Minkowski metric thereby having the same causal structure.

entire portions $A_\pm$ are absent. Furthermore, the r_* reaches a finite value, say zero when the curvature singularity at $r = 0$ is reached. This is space-like and therefore the top half of B_- and bottom half of B_+ are also absent leading to the maximally extended Schwarzschild space-time in figure (8.2).

The various null surfaces such as the *event horizon* ($r = r_+$) and the *Cauchy horizon* ($r = r_-$) are also identified together with the portions of the asymptotic infinities, $\mathcal{J}^{\pm}$, i_0. These will be defined in the more general context of black holes in asymptotically flat space-times in the next section.

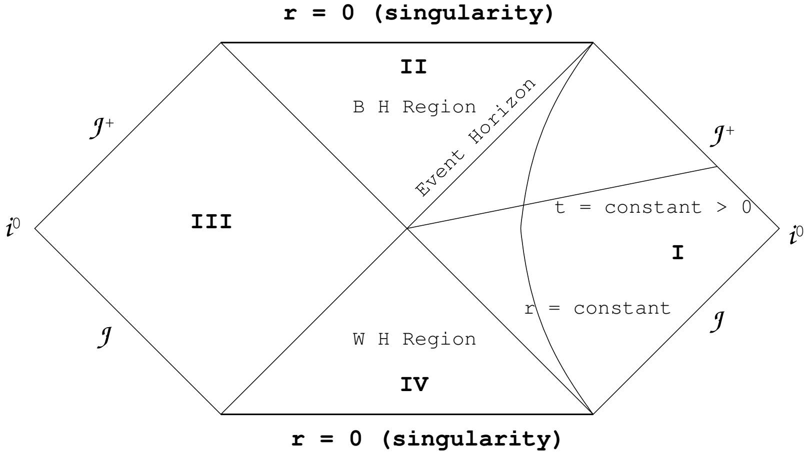


FIGURE 8.2: Maximally extended Schwarzschild space-time.

The Kerr–Newman family presents another novel feature apart from the two horizons of the Reissner–Nordstrom and the 'ring singularity' when the rotation parameter $a \neq 0$ - the *ergospheres.*

The stationary Killing vector has its norm given by (see equation 8.2),

$$\begin{aligned} \xi \cdot \xi &= g_{tt}(r,\theta) = \frac{\Delta - a^2 sin^2\theta}{\eta^2} = \frac{r^2 - 2Mr + a^2 + Q^2 - a^2 sin^2\theta}{\eta^2} \\ &= 0 \quad \text{for} \quad R_{\pm}(\theta) = M \pm \sqrt{(M^2 - a^2 - Q^2) + a^2 sin^2\theta} \end{aligned} \tag{8.6}$$

The hypersurface defined by $\xi^2 = 0 \;\leftrightarrow\; r = R_+(\theta)$ is called *infinite red-shift surface* since the light received at infinity from this surface will be infinitely red-shifted. The regions $r_+ < r < R_+(\theta)$ and $R_-(\theta) < r < r_-$ are called outer (inner) ergospheres, respectively. In the region between the inner and the outer ergospheres, the Killing vector ξ^μ is *space-like.* See figure (8.3).

Physically, it is the outer ergosphere that is of interest since it is accessible to far away observers. In this region, any time-like vector field, $u^\mu := \frac{dx^\mu}{d\tau}$, $u \cdot u = -1$, implies that $\frac{d\phi}{d\tau} > 0$ i.e. an observer within the ergosphere has to co-rotate with the Kerr–Newmann black hole. This is an extreme form of *frame dragging.*

This is also the region where the *Penrose Process* for extracting the rotational energy of the rotating black hole takes place. This is based on the following observation. When a space-time has Killing vectors, the time-like geodesics have corresponding constants of motion e.g. $\mathcal{E} := -u \cdot \xi$ which has

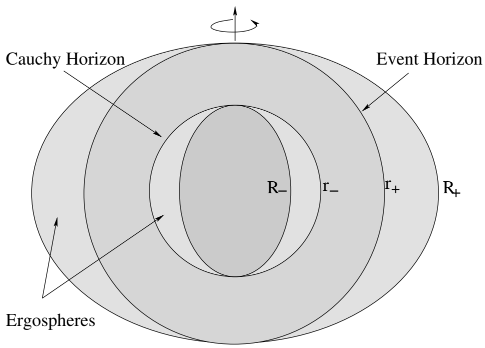


FIGURE 8.3: Ergospheres in Kerr–Newmann family.

the interpretation of energy per unit mass of the body following the geodesic. For a future directed ($u^t > 0$), time-like geodesic, $\mathcal{E}$ is positive in the exterior region where ξ^μ is time-like while it can have either sign if the geodesic is inside the ergosphere where ξ^μ is space-like. This, in particular means that a positive energy body can enter the ergosphere but a negative energy body *cannot* enter/exist in the exterior region. Furthermore, not only are negative energies possible only inside the ergosphere, the other constant of motion, angular momentum per unit mass, $L_z := u \cdot \psi$, must also be *negative*. Here $\psi^\mu = \partial_\phi$ is the Killing vector of axisymmetry and the z-component of the angular momentum of the black hole is defined as positive. The Penrose process now consists of sending in a body with initial (positive) energy, $E_i = m\mathcal{E}_i$, arranging it to separate it into two bodies A, B such that the body A (say) is put in an orbit with $E_A < 0$ and the body B is arranged to head to the exterior region. Energy conservation at separation implies that $E_B > E_i$ and therefore, we have extracted some energy from the black hole. The body A has to fall inside and it reduces both the mass and the angular momentum of the black hole. In this process, the ergosphere also shrinks a little. Repeated extraction of energy by this process will eventually halt the rotation of the black hole which also removes the ergoregion. The changes in the mass and angular momentum of the black hole by this process are not independent though.

In the discussion above equation (5.113), we noted that the Killing vector $\chi := \xi + \Omega\psi$, is *time-like* in the region exterior to the event horizon, $r > r_+$ and becomes light-like at the horizon. Hence, $u \cdot \chi \le 0$. This implies, $-\mathcal{E} + \Omega L_z \le 0$. For the body A which crosses the horizon, both $\mathcal{E}$ and L_z are negative and result in decrease of the mass and angular momentum of the black hole. Setting $\delta M = \mathcal{E}$ and $\delta J = L_z$, we get $\delta M \ge \Omega \delta J = \frac{a}{2Mr_+}(a\delta M + M\delta a)$ which translates into: $r_+^2 \delta M \ge Ma\delta a$. Substituting for $r_+^2 = M^2 + (M^2 - a^2) + 2M\sqrt{M^2 - a^2}$, the inequality can be expressed as,

$$r_+^2 \delta M - Ma\delta a \ge 0 \iff \delta\left(\frac{M^2 + M\sqrt{M^2 - a^2}}{2}\right) =: \delta M_{irr}^2 \ge 0 \, . \quad (8.7)$$

The condition on decreasing the mass and angular momentum of a Kerr black hole by the Penrose process, is conveniently expressed as stipulating that the *irreducible mass* [51], $M_{irr}^2(M,a) := (M^2 + M\sqrt{M^2 - a^2})/2$ must not decrease. Therefore, from a given black hole of mass M and angular momentum aM, the maximum amount of energy that can be extracted via the Penrose process is $E_{max}(M,a) = M(1 - \frac{1}{\sqrt{2}}\sqrt{1 + \sqrt{1 - a^2/M^2}})$ which in turn is the maximum possible for the maximally spinning black hole, $a = M$ and this is about 29% of the hole's initial mass.

There is also a counterpart of this mechanism in the scattering of waves off a Kerr black hole. For a massless wave equation for integer spin, the reflection coefficient is *larger* than 1 i.e. there is an enhancement of energy in the reflected wave. This is known as the *super-radiance* phenomena. It is absent for Dirac and Weyl wave equations.

This completes our discussion of the Kerr–Newman family of black holes.

8.2 General Black Holes and Uniqueness Theorems

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 modelled 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 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 *null infinity*, $\mathcal{J}^+$, of an asymptotically flat space-time. Events from its causal past can send signals to it. A black hole region must be excluded from this.

Thus, an *asymptotically flat* space-time, M, contains a *Black Hole region* $\mathcal{B}$ if $B := M - J^-(\mathcal{J}^+) \neq \emptyset$. Its boundary (three-dimensional) is called the *Event Horizon*, $H := M \cap \dot{J}^-(\mathcal{J}^+)$. It is a always a null hypersurface. However, such a definition is too general for proving useful statements. For many results discussed below, some further stipulations are required. As discussed in [17] for instance, the notion of *strong asymptotic predictability* suffices. It stipulates that the asymptotically flat space-time be such that in the unphysical space-time $(\tilde{M}, \tilde{g})$, there is a region $\tilde{V} \subset \tilde{M}$ containing the closure of $M \cap J^-(\mathcal{J}^+)$

such that $(\tilde{V}, \tilde{g})$ is globally hyperbolic. The 'closure' part implies that spatial infinity is included in $\tilde{V}$. The global hyperbolicity of $\tilde{V}$ further implies that $(M \cap \tilde{V}, g)$ is globally hyperbolic too. This in particular means that if from some regular conditions on a Cauchy surface (e.g. collapsing matter), a singularity develops, then it *cannot* be visible from the future null infinity. The condition of strong asymptotic predictability thus *excludes* the possibility of having naked singularity in such space-times[1].

The global hyperbolicity of the physical region guarantees existence of Cauchy slices and the intersection of the event horizon with a Cauchy slice Σ, is a two-dimensional, possibly disconnected submanifold. Each of its connected component is identified as an *instantaneous black hole.*

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 into more black holes (no bifurcation theorem). This result depends only on the definition of black holes and topology and is independent of the field equations. It stipulates that while black holes can merge and/or grow, they *cannot* 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 using the induced metric. Interestingly, the *Hawking area theorem* proves that *area of an instantaneous black hole may never decrease.* The proof uses Einstein equation together with the stress tensor satisfying the *null energy condition.* This result, known as the *second law of black hole mechanics*, 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 which is a stronger statement.

Indeed one can imagine processes involving black holes wherein a black hole does change its properties (e.g. 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.* These are black hole space-times with a time-like Killing vector. For these a lot more is known. Some of these results are summarized below [17, 54, 55].

1. *For stationary black hole, the event horizon is a Killing horizon i.e. a hypersurface on which the norm of some Killing vector ξ vanishes.*

2. *The instantaneous black holes have spherical topology. The area A of*

[1] Whether from the complete collapse of reasonable physical matter a strongly asymptotically predictable space-time *always* results or not is an open issue articulated as the *Cosmic Censorship Conjecture* [52, 53].

these black holes is therefore finite and of course constant due to stationarity. This is true for black holes in asymptotically flat space-times.

3. 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.*

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

(b) If ergosphere is present but intersects[2] 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.*

The three Killing vectors are related as, $\chi = \xi + \Omega_H \psi$ and Ω_H is called the *angular velocity* of the event horizon.

Since χ is light-like on the horizon, it defines a parameter κ, called the *surface gravity* of the event horizon, by the following equation holding on the event horizon H,

$$\nabla^\mu \chi^2 = -2\kappa\chi^\mu \quad \Rightarrow \quad \chi \cdot \nabla\chi_\mu = \kappa\chi_\mu \tag{8.8}$$

4. *The surface gravity is constant over the horizon*

This is the *zeroth law of black hole mechanics.* This result depends on the stress-tensor satisfying the so called *dominant energy condition.* Constancy of κ allows the interpretation of κ being proportional to the 'temperature'.

The existence of the two Killing vectors ξ, ψ in the general stationary, axisymmetric asymptotically flat space-times give the corresponding conserved quantities, the ADM mass M (7.13), and the angular momentum J (7.14) of the black hole.

At this stage, for the general stationary, axisymmetry, vacuum black holes we have assembled the parameters: area, surface gravity and angular velocity defined at the horizon (two-dimensional) and also the mass and angular momentum defined by Komar integrals evaluated in the asymptotic region. Thanks to the isometries, all of these are constant parameters except the surface gravity which, nevertheless is shown to be constant if matter satisfies dominant energy condition.

5. Vacuum black holes with nearby values of parameters A, κ, Ω, J, M satisfy,

$$\delta M = \frac{\kappa}{8\pi G}\delta A + \Omega\delta J \ .$$

[2]The case where an ergosphere exists but does *not* intersect the horizon is not completely clear [17].

This is the first law of black hole mechanics.

In (5.4.2.1) we had discussed the values of the black hole parameters for the Kerr–Newman family and have also verified the first law. The above is a general result.

The classic black hole uniqueness results establish that starting from the general definitions of stationary black holes in an asymptotically flat and asymptotically strongly predictable space-times, the Kerr family is the *only family* of solutions. The uniqueness results are also extended to inclusion of Maxwell field and leads to the Kerr–Newman family as the corresponding unique family of solutions [56–64]. These are also paraphrased as the statement, *Black Holes have No Hair*.

The uniqueness results for stationary black holes are limited to *four* space-time dimensions and with the naturally occurring long range gravitational and electromagnetic classical fields. In presence of non-abelian gauge fields and as well as interacting scalar fields, the uniqueness results are not completely established. With higher dimensions and non-asymptotically flat space-times lot more work remains to be done. These developments are reviewed in [65].

8.3 Black Hole Thermodynamics

The laws of black hole mechanics, including now the Maxwell fields as well, look very much like the laws of thermodynamics. Here is a table of analogies [17]:

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 thermo-dynamical system, black hole space-times are characterized by a few parameters. Just as for thermo-dynamical 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 [66, 67].

However if taken literally one immediately has a problem. If a black hole has a non-zero temperature, it must radiate. Since the surface gravity is defined for the horizon, we expect the *horizon* to radiate. But by definition nothing can come out of a black hole So how can we reconcile these? Here Hawking made a crucial observation. He noted 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 [17]:

$$\begin{aligned} T &= \frac{\hbar c^3}{8\pi G k_B M_\odot}\left(\frac{M_\odot}{M}\right) \ {}^0K \\ &= 6\times 10^{-8}\left(\frac{M_\odot}{M}\right) \end{aligned} \tag{8.9}$$

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. This is also the cause of the tension between general relativity which allows for black hole horizons and quantum theory which suggests evaporation thereby raising the possibility of pure quantum state evolving into a thermal density matrix - the *information loss problem.*

If the thermodynamic analogy is true, the statistical mechanics cannot be far behind and one way to ascribe micro-states to black hole horizons is to look for a quantum theory of gravity. 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 the entropy being proportional to the area is suggestive of associating *finitely many* states to an *elementary area* of a black hole.

There are very many ways in which one obtains the Bekenstein-Hawking 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-Hawking formula has been derived using the 'isolated horizon' framework (modulo the value of the 'Barbero-Immirzi' parameter being chosen for one black hole), for the so-called non-rotating horizons. String theorists too have 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.

Black holes which began as peculiar solutions of Einstein equations have revealed an arena where general relativity, statistical mechanics and quantum theory are all called in for an understanding.

8.4 Quasi-Local Definitions of Horizons

The various results on black hole mechanics/thermodynamics used event horizon as the *definition* of the black hole. This is unsatisfactory for two reasons. The event horizon definition refers to infinity *and* also needs the entire space-time to be known to identify it. At any spatial slice, an observer would not know if he/she is being engulfed by a surface which will be part of the event horizon! It is much more desirable, both from a conceptual and a practical angle e.g. in a numerical evolution, to characterize a black hole in a more local manner. Indeed such a characterization of black holes is available. For the stationary black holes it is captured by the notion of an *isolated horizon* [66,67] while for a evolving black hole, it is captured by the notion of a *dynamical horizon* [68,69]. Both these notions arose from the notion of *trapping horizons* [70] which are generalizations of the *apparent horizon* [17, 18]. Let us get a glimpse of these and note important results.

We have already defined trapped surfaces (6.17) as two-dimensional, space-like submanifolds such that the expansions of both the orthogonally in-going and out-going null geodesics is negative. These played a role in establishing singularity theorems (6.29, 6.30). These surfaces are also related to the event horizon in a *strongly asymptotically predictable* space-times with $R_{\mu\nu}k^\mu k^\nu \geq 0 \ \forall \ k \cdot k = 0$, namely, any marginally trapped surface is contained in the black hole region B. This property also extends to certain three-dimensional space-like submanifolds [17].

Let Σ be a any asymptotically flat Cauchy surface for $\tilde{V}$ - the region of the unphysical space-time $(\tilde{M}, \tilde{g})$ which is globally hyperbolic - containing the spatial infinity and being space-like there. Let C be a closed, three-dimensional submanifold of $\Sigma \cap M$, with its two-dimensional boundary $\dot{C}$. If the out-going null geodesics orthogonal to $\dot{C}$ have their expansion non-negative, then $\dot{C}$ is called *outer marginally trapped surface* ($\theta \leq 0$) and C is called *a trapped region*. Now not only the boundary $\dot{C}$ but the whole three-dimensional region C is contained in the black hole region, B.

A given Cauchy slice Σ may have several trapped regions. Let $\mathcal{T}$ be the closure of the union of all trapped regions of the Cauchy slice. This is called the *total trapped region* of Σ. Its (topological) boundary, $\mathcal{A} := \dot{\mathcal{T}}$ is called *an apparent horizon* *on Σ*. *If* it so happens that $\mathcal{T}$ is a manifold with boundary, *then* the apparent horizon is an outer, marginally trapped surface with $\theta = 0$. $\mathcal{A}$ is always contained in the event horizon or coincides with it.

In these notions, although event horizon does not play a role in the definition, there is still a weaker reference to the infinity through the asymptotically flat Cauchy slice containing i^0. It is tied to such a slice. In fact it is possible to have Cauchy slices which *do not* have any apparent horizon [71]! Hence we cannot quite use apparent horizon as an alternative characterization of a black hole.

The next effort at a local characterization are the trapping horizons of Hayward [70]. For a trapped surface S, expansions $\theta_\pm$ of both the orthogonally emanating null geodesics are negative. Marginally trapped surfaces allow zero expansion as well. We have thus the possibilities of $\theta_\pm \leq 0$ in different combinations. Furthermore, if we move off the null hypersurfaces $\Delta_\pm$ generated by the in-going/out-going null geodesics, the expansions $\theta_\mp$ may become positive or negative. These possibilities lead to future/past and outer/inner trapping horizons.

One begins with a Penrose's characterization of a trapped surface as one for which the product $\theta_+\theta_- > 0$. If both are positive, it is called a *past trapped* and if both are negative, it is called *future trapped* which are associated with white/black holes respectively. Collect *all points in the space-time* through which at least one trapped surface passes. A connected component of this set constitutes an *inextendible trapped region.* Its boundary is called a *trapping boundary* - not yet a horizon. Note that there is no reference to any asymptotic structures and we also do not know which is 'out-going' or 'in-coming'. Let ℓ and n denote the two null congruences emanating orthogonally from a trapped surface with $\theta_+ := \theta_\ell$ and $\theta_- := \theta_n$[3].

To define a horizon, consider marginally trapped surfaces with $\theta_\ell = 0$ (say). Let H be a three-dimensional submanifold foliated by marginally trapped surfaces with the further property that $\theta_n|_H \neq 0$ *and* the Lie derivative $\mathcal{L}_n\theta_\ell \neq 0$. Its closure $\bar{H}$ is called a *Trapping Horizon.* It is future trapping horizon if $\theta_n < 0$ and past if $\theta_n > 0$ while it is 'outer' if $\mathcal{L}_n\theta_\ell < 0$ and 'inner' if the Lie derivative is positive. The *future outer trapping horizons* provide a quasi-local definition of black holes.

Hayward proves a number of results for the trapping horizons [70]. These are analogous to the theorems for event horizons and we list them below.

The notation and the definitions introduced are based on a 2 + 2 decomposition adapted to double null foliations. We recall the bare minimum notation to state the results, further details are available in [70, 72].

The the space-time is taken to be of the form $S\times\mathbb{R}\times\mathbb{R}$, locally coordinatised by (u, v, x^a), $a = 1, 2$. The local coordinates u, v can be reparametrized and the physical quantities are invariant under these reparametrizations. The metric is parametrized in terms a non-singular, Euclidean signature metric on S, h_{ab}, two 'shift vectors' r, s (two-dimensional) and three 'lapse functions', e^f, a, c accounting for 10 independent metric coefficients. The null foliations require $a = c = 0$. The null normals are denoted as ℓ, n and they are normalized as $\ell \cdot n = -e^f$. Let $\mathcal{L}_+ := \mathcal{L}_{e^{-f}\ell}$, $\mathcal{L}_- := \mathcal{L}_{e^{-f}n}$, $\mathcal{D}$ be the covariant derivative compatible with h_{ab} and μ the corresponding area 2-form on S. Denote,

$$\theta_\pm = \frac{1}{2}h^{ab}\mathcal{L}_\pm h_{ab} \quad , \quad \sigma^\pm_{ab} := h_a{}^c h_b{}^d \mathcal{L}_\pm h_{cd} - h_{ab}\theta_\pm \tag{8.10}$$

$$\omega_a := \frac{1}{2}e^f h_{ab}[\mathcal{L}_-(e^{-f}\ell)]^b \quad , \quad \nu_\pm := \mathcal{L}_\pm f \tag{8.11}$$

[3]The $N_\pm$ of [70] correspond to the ℓ, n here.

For the outer trapping horizons, define the *trapping gravity* (analogue of the surface gravity) as,

$$\kappa \ := \ \frac{1}{2}\sqrt{-e^f \mathcal{L}_n \theta_\ell} \tag{8.12}$$

This is an invariant under the null foliation reparametrizations and is positive (can be allowed to be zero to include degenerate horizons).

For compact trapping horizon, S, define,

$$m \ := \ \sqrt{A/16\pi} \qquad \text{'irreducible energy';} \tag{8.13}$$

$$a \ := \ m\sqrt{\frac{1}{4\pi}\int_S (\omega - \mathcal{D}f)_a(\omega - \mathcal{D}f)^a} \qquad \text{'angular energy';} \tag{8.14}$$

$$q \ := \ m\sqrt{2\int_S \mu \ e^f \ \rho} \qquad \text{'matter energy'.} \tag{8.15}$$

Here is the list of results [70].

1. *Topology:* Outer, marginal surfaces are either spherical (compact) or planar (non-compact). This uses the dominant energy condition and a property of being 'well-adjusted' (so that certain integrals exist) when the surface is non-compact. The topology of inner, marginal surfaces has no such restrictions.

2. *Signature:* Trapping horizons are *null* only if it is instantaneously stationary i.e. $\mathcal{L}_\ell \theta_\ell|_H = 0$ otherwise the outer trapping horizons are spatial (induced metric is Euclidean) while the inner ones are Lorentzian.

3. *Second law:* The area form of future, outer trapping horizons is generically non-decreasing and is constant only if the Horizon is null. This uses the null energy condition. For compact topology, the same holds for the area of the horizon.

4. *Zeroth law:* The trapping gravity has an upper bound for a compact, outer trapping horizons. This upper bound is attained iff the trapping gravity is constant over the horizon. Specifically,

$$\int_S \mu\kappa \ \le \ 4\pi\sqrt{m^2 - a^2 - q^2} \tag{8.16}$$

5. *First law:* The change in the area form along a vector z which is tangent to the horizon and normal to S (normalized if spatial) is determined by the trapping gravity and an energy flux:

$$\kappa \mathcal{L}_z \mu|_H \ = \ 8\pi\Phi_{matter} \ = \ -\mu e^f \theta_n \sqrt{\pi\phi_+ + \frac{1}{32}(\sigma^+)_{ab}(\sigma^+)^{ab}}\Bigg|_H . \tag{8.17}$$

Here ϕ^+ is the $\ell\ell$ component of the matter stress tensor and the 'shear' σ^+_{ab} is defined in eq. (8.10).

A related but different quasi-local formulation of black hole thermodynamics is given by Ashtekar et al. in terms of the *dynamical horizon* [68, 69] and its stationary counterpart, the isolated horizon [66, 67] which is used in the computation of entropy from the quantum micro-states [73].

Dynamical horizons are space-time concepts too as are the trapping horizons. They are stipulated to be space-like and there is no condition the variation of θ_ℓ off the horizon i.e. on $\mathcal{L}_n\theta_\ell$ (though $\theta_n < 0$ holds). The definition is thus tied to only quantities intrinsic to the dynamical horizon. This also suffices to capture an evolving horizon with corresponding laws of mechanics analogous to the thermodynamic ones. Below we summarize basic definitions and results from [68, 69].

We begin with the definition of a dynamical horizon. A *Dynamical Horizon*, H, is a smooth, three-dimensional, space-like submanifold which can be foliated by close 2-manifolds such that on each leaf S, expansion of one of the null normals, l^μ is zero while that of the other null normal is negative: $\theta_\ell = 0,\ \theta_n < 0$. The foliation satisfying these conditions is called *a preferred foliation* and its leaves are called *cross-sections* of the horizon.

The unit time-like normal of H is denoted by $\hat{\tau}^\mu$ while the unit space-like normal to a cross-section which is tangent to H is denoted by $\hat{r}^\mu$. The two null normals to cross-sections are chosen to be $\ell := \hat{\tau}+\hat{r}$ and $n := \hat{\tau}-\hat{r}$ so that $\ell\cdot n = -2$. The induced metric on H is given by $q_{\mu\nu} := g_{\mu\nu} + \hat{\tau}_\mu\hat{\tau}_\nu$ while its extrinsic curvature is given by $K_{\mu\nu} := q_\mu{}^\rho q_\nu{}^\sigma\nabla_\rho\hat{\tau}_\sigma$. The metric compatible covariant derivative on H, and its curvature are denoted by $\mathcal{D}$ and $\mathcal{R}_{\mu\nu}$ etc. Likewise, using $\hat{r}^\mu$, within H we can induce a metric on leaves with the corresponding compatible covariant derivative and curvature as well as extrinsic curvatures of the leaves: $\tilde{q}_{\mu\nu} := q_{\mu\nu} - \hat{r}_\mu\hat{r}_\nu\ ,\ \tilde{K}_{\mu\nu} := \tilde{q}_\mu{}^\rho\tilde{q}_\nu{}^\sigma\mathcal{D}_\rho\hat{r}_\sigma$ etc.

From the very definition of the dynamical horizon and the unit vector $\hat{r}$ normal to the cross-sections, it follows that $\tilde{K} = \tilde{q}^{\mu\nu}\mathcal{D}_\mu\hat{r}_\nu = \frac{1}{2}\tilde{q}^{\mu\nu}\nabla_\mu(\ell_\nu - n_\nu) = -\frac{1}{2}\theta_n > 0$ i.e. the area of cross-section increases monotonically in the direction of $\hat{r}$. On physical grounds we expect this increase to be correlated with the net energy flux entering the horizon. This is the fundamental relation that governs the details of the preferred foliation. A dynamical horizon being space-like, we can use it to do a 3 + 1 decomposition of the Einstein equation. The vector and the scalar constraints, being preserved along any evolution, provides the necessary relations. The constraint equations are:

$$\begin{aligned} 16\pi G\bar{T}_{\mu\nu}\hat{\tau}^\mu\hat{\tau}^\nu &= \mathcal{R} + K^2 - K^{\mu\nu}K_{\mu\nu} \\ 8\pi G q^\alpha_\mu\bar{T}^{\mu\nu}\hat{\tau}_\nu &= \mathcal{D}_\beta(K^{\alpha\beta} - Kq^{\alpha\beta}) \\ \bar{T}_{\mu\nu} &:= T_{\mu\nu} - \frac{1}{8\pi G}\Lambda g_{\mu\nu} \end{aligned} \tag{8.18}$$

The matter energy-momentum current is given by $\bar{T}_{\mu\nu}\hat{\tau}^\nu$ and to define its flux across the horizon, we need to choose a vector field, $\xi^\mu = N\hat{\tau}^\mu + N^\mu$ for a suitable choice of lapse and shift. This should be transversal to the horizon which is space-like and therefore ξ^μ should be time-like or null and it should also be have a component along $\hat{r}^\mu$ to cut across the cross-sections. A natural choice is $\xi^\mu_N = N\ell^\mu$ which corresponds to the choice of shift vector $N^\mu = N\hat{r}^\mu$.

Consider a portion ΔH of H, which is bounded by two cross-sections S_1, S_2 (which are marginally trapped surfaces). The matter-energy flux across this portion is given by,

$$\mathcal{F}_{matter}(\xi_N) \ := \ \int_{\Delta H} d^3v\sqrt{q}\ \bar{T}_{\mu\nu}\hat{\tau}^\mu\xi^\nu_N \tag{8.19}$$

The right-hand side is expressed in terms of linear combinations of the constraints multiplied by the lapse and simplified further using a 2 + 1 decomposition within ΔH. On ΔH introduce an areal radial coordinate R along $\hat{r}$ so that each cross-section with coordinate R, has an area $4\pi R^2$ and two coordinates χ^a along the cross-sections. The Jacobian factor is absorbed by choosing the lapse $N := N_R$ such that $N_R\sqrt{q}d^3v = dR\sqrt{\tilde{q}}d^2\chi$. Putting all these together, one obtains the *area balance equation*,

$$\frac{1}{16\pi G}(R_2 - R_1)\mathcal{I}(S) \ = \ \mathcal{F}_{matter}(\xi_{N_R}) + \mathcal{F}_{grav}(\xi_{N_R}) \quad \text{where,} \tag{8.20}$$

$$\mathcal{F}_{grav}(\xi_{N_R}) \ := \ \frac{1}{16\pi G}\int_{\Delta H} d^3v\sqrt{q}N_R\left(\sigma^{ab}\sigma_{ab} + 2\zeta^a\zeta_a\right)$$

$$\text{and,}\ \sigma_{ab} \ := \ \tilde{q}_a{}^c\tilde{q}_b{}^d\nabla_c\ell_d \quad , \quad \zeta^a \ := \ \tilde{q}^{ab}\hat{r}\cdot\nabla\ell_b, \tag{8.21}$$

and $\mathcal{I}(S)$ is proportional to the Euler invariant of the cross-section (8π for spherical topology). Since the expansion $\theta_\ell = 0$, the change in the horizon radius is also the change in the *Hawking Mass* of the horizon cross-section.

The balance equation implies that if the stress tensor satisfies *dominant energy condition* so that $T\cdot\hat{\tau}\cdot\ell$ is non-negative and the cosmological constant is *non-negative*, then the right-hand side is non-negative and therefore the cross-sections is spherical or toroidal. If $\Lambda < 0$, then any topology is possible.

The quantities σ^2 and ζ^2 are interpreted as contributions of gravitational waves with the ζ^2 being non-zero when the angular momentum is non-zero. It is absent if the dynamical horizon becomes null (or has null portion).

Just as matter flux carries energy across the horizon, so also it carries angular momentum and there is a corresponding balance equation derived from the same constraint equations.

Choose a vector field, ϕ^a, on H such that it is tangential to the cross-sections. Then dotting with the vector constraint and integrating over ΔH, one obtains the *angular momentum balance equation*,

$$J_{S_2}(\phi) - J_{S_1}(\phi) \ = \ \mathcal{J}_{matter}(\phi) + \mathcal{J}_{gravity}(\phi) \qquad \text{where,} \tag{8.22}$$

$$J_S(\phi) \ := \ -\frac{1}{8\pi G}\int_S K_{ab}\phi^a\hat{r}^b\sqrt{\tilde{q}}d^2\chi \quad ,$$

$$\mathcal{J}_{matter}(\phi) \ := \ -\int_{\Delta H} T_{ab}\hat{\tau}^a\phi^b\sqrt{q}d^3v \quad \text{and,}$$

$$\mathcal{J}_{gravity}(\phi) \quad := \quad -\frac{1}{16\pi G}\int_{\Delta H}\left(K^{ab}-Kq^{ab}\right)\mathcal{L}_{\phi}q_{ab}\sqrt{q}d^3v\ .$$

The angular momentum $J_S(\phi)$ is defined here for an arbitrary ϕ tangential to a cross-section and is not required to be a rotational isometry. Hence it is more appropriately called *generalized* angular momentum. The usual angular momentum at infinity is defined using the Arnowitt-Deser-Misner (ADM) framework. We could imagine choosing a Cauchy slice Σ extending to spatial infinity and having an inner boundary at a chosen cross-section S and choosing a new vector field $\bar{\phi}^a$ on Σ which matches with the ϕ^a on S and becoming an axial isometry at infinity. Then the angular momentum at S will have similar expression as above with K_{ab} replaced by $\bar{K}_{ab}$ which is the extrinsic curvature of Σ. The definition thus depends on the slicing and is ambiguous. However, if ϕ^a is *divergence free*, then the angular momentum definition is independent of the slice Σ. It is a result that the ζ^2 *contribution to energy vanishes iff the angular momentum of* S *vanishes for every, divergence-free* ϕ^a *on* S. Restrict now for spherical topology.

We have defined the energy flux and angular momentum flux for the choices of ξ^a, ϕ^a, N^a, the areal radial coordinate and the corresponding choice of the lapse N. Furthermore, noting that for $R_2 = R_1 - \delta R$, the left-hand side of the area balance equation (8.20) is $\frac{\delta A}{8\pi G}\frac{1}{2R}$, we *define* an *effective surface gravity*, $\kappa_R := 2R^{-1}$ and identify the right-hand side of (8.20) as the *change* in the ξ_{N_R}*-energy of the horizon,* $\delta E(\xi_{N_R})$. Now generalizing from ξ_{N_R} to a general time evolution vector field $t^a := N_R\ell^a - \Omega(R)\phi^a$, repeating the steps leading to the area balance equation and using the corresponding identification of effective surface gravity and change in the energy, one gets a combined 'balance' equation,

$$\begin{aligned}\frac{R_2-R_1}{2G} \quad &+ \quad J_{S_2}(\Omega\phi) - J_{S_1}(\Omega\phi) - \int_{\Omega_1}^{\Omega_2} d\Omega J_S(\phi) \\ &= \quad \mathcal{F}_{matter}(t) + \mathcal{F}_{grav}(t) + \mathcal{J}_{grav}(\Omega\phi) \ =: \ \Delta E(t) \quad (8.23)\end{aligned}$$

For infinitesimal $\delta R = R_2 - R_1$, this takes the form,

$$\frac{\kappa_R}{8\pi G}\delta A + \Omega\delta J(\phi) \ = \ \delta E(t) \quad (8.24)$$

This is the *first law of dynamical horizon mechanics.*

The above expressions have been given for the particular choice of areal radial coordinate, but it can be generalized to other coordinates r. There is a first law expression for every choice of t^a, ϕ^a vector fields and the function $\Omega(r)$.

For further details and discussion of existence and uniqueness issues for dynamical horizons, please see [68, 69].

As mentioned before, there is a separate definition of *isolated horizon* to correspond to the equilibrium states of horizons and can be thought of as a generalization of Killing horizons. These are supposed to be solutions of

Einstein equation with stress tensor satisfying the dominant energy condition. Unlike the dynamical horizons, isolated horizons are *null* hypersurfaces Δ with a null normal ℓ. There are three progressively restrictive definitions which are briefly summarized below.

- *Non-Expanding Horizon,* Δ (NEH): requiring only that $\theta_\ell = 0$ and no condition on θ_n. One of its main implications is that the $\mathcal{L}_\ell q_{ab} = 0$ where q_{ab} is the induced degenerate metric on Δ. The intrinsic geometry is 'time' (along ℓ) independent. Any constant scaling of ℓ is also a null normal and all such null normals constitute an equivalence class $[\ell]$. There can be several equivalence classes on a given NEH, Δ. The null normal congruence is expansion, shear and twist free. This in turn implies that there exist a unique, 1-form ω_a on Δ defined by, $V \cdot \nabla \ell^a = (V^b \omega_b)\ell^a \ , \ \forall \ V^a$ tangent to Δ. For each null normal, one defines the *surface gravity*, $\kappa_\ell := \ell^b \omega_b$. It follows that the 1-form ω_a satisfies $d\omega = (\Psi_2 - \bar{\Psi}_2)\epsilon^{(2)}$ where $\Psi_2 = C_{\mu\nu\alpha\beta}\ell^\mu m^\nu \bar{m}^\alpha n^\beta$ is one of the Penrose-Newman components of the Weyl tensor C which turns out to be independent of the null tetrad $(\ell, n, m, \bar{m})$ on a NEH and $\epsilon^{(2)}$ is the natural area 2-form on Δ satisfying $\mathcal{L}_\ell \epsilon^{(2)} = 0 = \ell \cdot \epsilon^{(2)}$.

  An NEH is said to be *non-rotating* if imaginary part of Ψ_2 vanishes.

- *Weakly Isolated Horizon,* $(\Delta, [\ell])$ (WIH): requires that a q_{ab} compatible connection also satisfies $(\mathcal{L}_\ell \mathcal{D}_a - \mathcal{D}_a \mathcal{L}_\ell)\ell^b = 0 \ \forall \ \ell \in [\ell]$. Here $\mathcal{D}$ is the unique, torsion-free, metric compatible derivative on Δ *induced* from ∇ on the space-time. Thus, only *some* components of the induced connection are required to be time independent. It follows that $\mathcal{L}_\ell \omega_a = 0$. Thus follows the *zeroth law of mechanics* for all WIH.

  A WIH with vanishing surface gravity is said to be *extremal*.

- *Isolated Horizon,* $(\Delta, [\ell], \mathcal{D})$ (IH): requires further that *all* components of the induced connection be time independent i.e. in the definition of WIH, replace ℓ^b by an arbitrary V^b tangential to Δ. Every WIH is not necessarily an IH and generically, if a WIH admits IH structure, it is unique.

  Symmetries of a IH are determined by isometries of the induced metric on cross-sections of Δ. If it has rotational isometry i.e. there exist a Killing vector ϕ^a on Δ, the *angular momentum* of a WIH is defined to be $J_\Delta(\phi) := -\frac{1}{8\pi}\int \phi^a \omega_a \ \epsilon^{(2)}$. If there are matter gauge fields, there are further contributions to the horizon angular momentum. The above expression may thus be termed *purely geometrical* (or 'bare') angular momentum of Δ.

  Consider now a definition for mass of an IH. Every constant linear combination of ℓ and ϕ is an isometry of Δ and we may associate a horizon energy with such a Killing vector, $t^a(B, \Omega) := B\ell^a - \ \Omega\phi^a$, where B and Ω are constants. This is explored conveniently in the *covariant* *phase space formalism* i.e. employing the (pre-)symplectic structure on the space of solutions of the field equations. Every vector field on the space-time manifold, induces a vector field on the covariant phase space and this is required to be a Hamiltonian vector field in order to be able to define a *function* on the phase space. Not every vector field $t^a(B,\Omega)$ induces a Hamiltonian vector field on the phase space. Explicit computation shows that the surface gravity, $\kappa(B,\ell) := B\ell^a\omega_a$, and the angular velocity parameter Ω, must be functions of the two quantities defined on Δ, namely, the horizon area a_Δ and the horizon angular momentum J_Δ, satisfying,

$$\frac{\partial\kappa(B,\ell)}{\partial J_\Delta} \;=\; 8\pi G\frac{\partial\Omega}{\partial a_\Delta}$$

This then implies that $\frac{\kappa(B,\ell)}{8\pi G}\delta a_\Delta + \Omega\delta J_\Delta = \delta E_\Delta(B,\Omega,\ell)$ which is the statement of the *first law of mechanics of (weakly) Isolated Horizon.*

As in the case of the dynamical horizons, there are infinitely many first laws. Thanks to the uniqueness theorems for (electro-)vacuum black holes, we have a unique dependence of the surface gravity on area and angular momentum, namely that obtained in the Kerr–Newman solution. Choosing $\kappa(a_\Delta, J_\Delta)$ to be this function, the above integrability conditions can be solved to give the angular velocity and the E_Δ for the isolated horizons.

Finally, imagine that a dynamical horizon 'relaxes' to an isolated horizon so that H and Δ are 'joined' at some boundary. It turns out that the two notions of energy and angular momentum defined on H and Δ agree at the boundary.

This completes our basic summary of the quasi-local generalizations of black holes. There are many other interesting aspects of these and several delicate points which should be seen in the references [66–70].

Chapter 9

Cosmological Space-Times

In section 5.2, we looked at the physically well motivated model space-time for the universe. The metric of this space-time has only *one* function of time, the scale factor and its evolution reveals the first instance of a space-time singularity. Is this an artifact of the presumed very high degree of symmetry? Apart from the fact that the universe is certainly neither exactly homogeneous nor isotropic, the added interest in more general 'cosmological space-times' is for reasons of the issue of singularity. An obvious strategy would be to loosen the degree of symmetry required of the space-time. Thus as a first step, we give up isotropy, but retain homogeneity to get the class of *homogeneous models*. The next steps are to introduce inhomogeneities in just one direction to get for instance, the class of *Gowdy* models and finally to drop homogeneity completely. These mathematically motivated models also serve as testing ground for quantum versions of general relativity. In this chapter, we will discuss the class of homogeneous space-times and briefly describe the Belinskii-Khalatnikov-Lifshitz (BKL) conjecture for approach to a singularity.

The $four$-dimensional, spatially homogeneous space-times have all been classified completely.

Let us recall from the section 5.2 that a space-time is spatially homogeneous if (a) it can be foliated by a 1-parameter family of space-like hypersurfaces, Σ_t and (b) possessing a (Lie) group of isometries such that for each t and any two points $p, q \in \Sigma_t$ there exist an isometry of the space-time metric which maps p to q. The isometry group G is then said to act *transitively* on each of the Σ_t. If the group element connecting p, q is unique, the group action is said to be *simply transitive* (otherwise multiply transitive). Spatially homogeneous space-times can be further divided into two types depending upon whether or not there is any subgroup of isometries which have a simply transitive action.

A spatially homogeneous space-time is be of a *Bianchi type* if the group of isometries contains a subgroup (possibly itself), G^*, which acts simply transitively on Σ_t. If there is no such subgroup, then it is of the *Kantowski–Sachs type*. It turns out that except for the special case of $\Sigma \sim S^2 \times \mathbb{R}$ and $G = SO(3) \times \mathbb{R}$, in all other cases one has a Bianchi type space-time. Interior of the Schwarzschild solution is an example of Kantowski–Sachs type space-time.

Transitive action implies that there must be at least three independent Killing vectors at each point of Σ_t since Σ_t is three-dimensional. But there could be additional Killing vectors which vanish at a point. These Killing vectors generate the *isotropy* (or stability) subgroup, $H \subset G$. Since H will induce a transformation on the tangent spaces to the spatial slices, it must be a subgroup of $SO(3)$ and thus dimension of G can be at the most 6 and at least 3 since the dimension of G^* is always 3. All three-dimensional Lie groups have been classified by Bianchi into 9 types. The classification goes along the following lines [74].

A Lie algebra (or the connected component of a Lie group) is characterized by structure constants ${C^I}_{JK}$ with respect to a basis X_I, satisfying the antisymmetry and Jacobi identity, namely,

$$\begin{aligned} [X_J, X_K] &= {C^I}_{JK} X_I; \;\; {C^I}_{JK} = -{C^I}_{KJ}; \\ 0 &= \sum_{(IJK)} C^N_{IL} {C^L}_{JK}, \quad I, J, K = 1, 2, 3. \end{aligned} \tag{9.1}$$

Using the availability of the Levi–Civita symbols, $\mathcal{E}_{IJK}$, $\mathcal{E}^{IJK}$, $\mathcal{E}_{123} = 1 = \mathcal{E}^{123}$, we can write the structure constants as,

$$ {C^I}_{JK} = \mathcal{E}_{JKL} C^{LI}, \;\; C^{IJ} := M^{IJ} + \mathcal{E}^{IJK} A_K \tag{9.2}$$

Thus, the 9 structure constants are traded for 6 M^{IJ} (symmetric in IJ) and the 3 A_K. This has used only antisymmetry. The Jacobi identity implies, $M^{IJ} A_J = 0$.

Under a change of basis of the Lie algebra, $X_I \to {S_I}^J X_J$, the structure constants too transform linearly. Using these, the symmetric M^{IJ} can be diagonalized by orthogonal transformations and the non-zero eigenvalues can be further scaled to ± 1 i.e. we can arrange, $M^{IJ} = n^I \delta^{IJ}$. The condition $M^{IJ} A_J = 0$ now implies that *either* $\mathrm{A_I} = 0$ (Class A) *or* $\mathrm{A_I} \neq 0$ (class B) in which case M^{IJ} has a zero eigenvalue and we may take the non-zero eigenvector A_I to be along the '1st' axis, i.e. $A_I = a\delta_{I,1}$ and $n^1 = 0$. This leads to,

$$[X_J, X_K] = n^I \mathcal{E}_{IJK} X_I + X_J A_K - X_K A_J \; .$$

In the class A, there are precisely 6 possibilities organized by the *rank of the matrix* M^{IJ} $(= 0, 1, 2, 3)$ and *signature* $(++, +-)$ for rank 2 and $(+++, ++-)$ for the rank 3. The eigenvalues of M^{IJ} can be taken to be $n^I = \pm 1, 0$.

In the class B, the rank of M^{IJ} cannot be 3 and the possibilities are restricted to the ranks 0, 1, 2 and signatures $(++, +-)$ for rank 2. If the rank of M is 0, all three eigenvalues are zero and scaling X_1, we can arrange $a = 1$. For rank 1, taking n_3 to be the non-zero eigenvalue, scaling X_1, X_3 ensures $a = 1$. For rank 2 however, $(n_2 = \pm 1, n_3 = \pm 1)$, no scaling can preserve n_2, n_3 *and* set $a = 1$ (though $a = 1$ is of course possible).

Here is a table of the classification of Riemannian, homogeneous 3-geometries [74]:

Type	a	n_1	n_2	n_3
		Class A		
I	0	0	0	0
II	0	1	0	0
VII_0	0	1	1	0
VI_0	0	1	−1	0
IX	0	1	1	1
VIII	0	1	1	−1
		Class B		
V	1	0	0	0
IV	1	0	0	1
VII_a	a	0	1	1
III	1	0	1	−1
VI_a	a	0	1	−1

When the stability subgroup $H = SO(3)$, one has isotropy in addition to homogeneity i.e. Robertson-Walker space-times. We know that these come in three varieties depending on the constant spatial curvature. The spatially flat case is of type Bianchi I while positively curved case is of type Bianchi IX. The negatively curved case is in class B, type V.

The metrics of the general Bianchi type space-times have at the most 6 degrees of freedom thus constituting what are known as *mini-superspace models*. Just as for the FLRW case, we could fix the spatial metric modulo a scale factor which depends on time, the spatial metrics of these general models *can be put* in the form:

$$ds_3^2 = g_{IJ}e^I_i e^J_j dx^i dx^j, \quad e^I := e^I_i dx^i \text{ satisfy } de^I = \tfrac{1}{2}C^I{}_{JK}e^J \wedge e^K. \quad (9.3)$$

The e^I, $I = 1, 2, 3$ are the so called *Maurer-Cartan* 1-forms on the group manifold G^* and g_{ij} are *constants* on the group manifold which is identified with one of the hypersurfaces of homogeneity, Σ_0, say. The Maurer-Cartan forms are the *unique, Lie algebra valued*, 1-forms invariant under the left action of the group on itself.

A good deal of work on these classes of space-times with or without matter has been done. What is important for us with regards to the singularity issue are the two models: Bianchi-I and Bianchi-IX.

To proceed further, we need to introduce suitable coordinates on the space-time manifold and choose the corresponding metric coefficients. This can be done as follows. At any point $p \in \Sigma_0$, choose a normal and consider the geodesic emanating from it. It will intersect the other surfaces Σ_t's only once due to the foliation property of the Σ's. Choosing a basis in the tangent space of p, parallel transport it along the geodesic. By construction, the transported bases will remain orthogonal in each of the Σ_t. These in turn can be Lie transported on each of the hypersurfaces using the isometries to construct the corresponding spatial metric in the same form as above. The only difference would be that the g_{ij} would now depend on the label t of the hypersurface. We can now change from the foliation label t to the *proper time* τ of the geodesic of the geodesic. This then implies that the *space-time* metric can be taken in the form,

$$\Delta s^2 \;=\; -\Delta\tau^2 + g_{IJ}(\tau) e^I_i e^J_j \Delta x^i \Delta x^j. \tag{9.4}$$

The time τ is called the *synchronous time*. This is the form from which we proceed.

Note that it is always possible to diagonalize $g_{IJ}(\tau)$ at one particular τ, but it cannot be assured that the metric can be taken to be diagonal for all τ. For the *vacuum* case, however, it is consistent to have a diagonal metric for all times.

Bianch-I: The structure constants are zero and the Maurer-Cartan equations allow us to choose dx^I themselves as the invariant 1-forms. By taking linear combinations we can diagonalize the metric to $g_{IJ}(\tau)$ =diag$(a_1(\tau), a_2(\tau), a_3(\tau))$. It is a simple calculation to get the Ricci tensor[1]. The only non-vanishing expressions are:

$$\Gamma^\tau{}_{ii} = a_i\dot{a}_i \quad , \quad \Gamma^i{}_{\tau i} = \frac{\dot{a}_i}{a_i} \qquad\qquad \text{for } i = 1,2,3\ ;$$

$$R_{\tau\tau} = -\sum_{i=1}^{3}\frac{\ddot{a}_i}{a_i} \quad , \quad R_{ii} = -\dot{a}_i^2 + a_i\ddot{a}_i + a_i\dot{a}_i\left(\sum_j \frac{\dot{a}_j}{a_j}\right). \tag{9.5}$$

These are easy to solve. The trivial solution, namely, $\dot{a}_i = 0\ \forall\ i$, is just the Minkowski metric. The evolving solutions show a singular behaviour for the determinant of the spatial metric, namely, it must *vanish* for some value of τ and we choose τ such that it happens at $\tau = 0$. The solution for the three scale factors is then,

$$a_i(\tau) = \hat{a}_i\tau^{p_i}, \quad \sum_i p_i = 1 = \sum_i p_i^2\ . \tag{9.6}$$

The p_i's are constants. The $R_{ii} = 0$ equations give the solution for $a_i(\tau)$ *and* $\sum_i p_i = 1$ while the $R_{\tau\tau} = 0$ equation gives the further constraint $\sum_i p_i^2 = 1$. This is the Kasner solution already obtained in 1925 [75]. The parameter space of p_i's is one-dimensional and we have a one parameter family of evolving space-times.

[1] It is possible to derive the general form of the Einstein tensor for the homogeneous metrics and is available in [17, 74]. For our purposes, the direct computations can be done more easily.

There is one special case where only one p_i equals 1 and others are zero which implies expansion along the corresponding direction, say, the x-direction. It can be seen as a Rindler wedge, by making the coordinate change: $x' = \tau\sinh(x)$, $\tau' = \tau\cosh(x)$. This transforms $-\Delta\tau^2 + \tau^2\Delta x^2 + \Delta y^2 + \Delta z^2 \to -\Delta\tau'^2 + \Delta x'^2 + \Delta y^2 + \Delta z^2$. In the (τ', x') plane, constant τ curves are hyperbolae with $\tau = 0$ degenerating into the pair of 45° lines. The space-time is manifestly Riemann flat and the vanishing of a_1 at $\tau = 0$, is a coordinate singularity. In all other cases, the space-time is non-flat and the singularity due to vanishing determinant as $\tau \to 0$ is a physical singularity.

It can be seen that we must have *one* exponent to be negative and the other two be positive. Therefore while the physical volume of a cell - $(a_1a_2a_3)\times$ the comoving volume - vanishes monotonically as $\tau \to 0$, the shape distorts with two directions contracting while the remaining one expanding, as τ *decreases.*

Bianchi IX: This is the most complex of the Bianchi types and an exact solution is not known. Taking $g_{IJ}(\tau) = \text{diag}(a^2, b^2, c^2)(\tau)$, the vacuum Einstein equation leads to [74]:

$$(abc)\frac{d}{d\tau}\left[\frac{1}{a}(abc)\frac{da}{d\tau}\right] = \frac{1}{2}\left((b^2-c^2)^2 - a^4\right) \quad \text{and cyclic} \tag{9.7}$$

$$bc\frac{d^2a}{d\tau^2} + \text{cyclic} = 0 \tag{9.8}$$

It is convenient to choose a new time parameter η defined through $d\tau := (abc)d\eta$ and also introduce $a := e^{\alpha}$ etc. The equations take the form ($'$ denotes $\frac{d}{d\eta}$),

$$2\alpha_i'' = (a_j^2 - a_k^2)^2 - a_i^4, \quad \sum_i \alpha_i'' = 2\sum_{i<j}\alpha_i'\alpha_j'. \tag{9.9}$$

Notice that if the right-hand sides are zero we get back to the Kasner case ($\eta \sim \ln\tau$ with equality equivalent to $\sum_i p_i = 1.$). This suggests that if the terms on the right-hand sides become close to zero, we can expect approximate Kasner behaviour. However, the right-hand sides of all three equations cannot remain close to zero, since in the backward evolution (η decreasing), two scale factors decrease while one increases. This results in different scale factors increasing and decreasing taking turns. Thus the Kasner pattern of two contracting directions and one expanding one, holds for some duration; the directions get permuted and then the pattern continues again. The rules of following the shifts of Kasner epoch are known and are discussed in detail in [74]. In all of these shifts, the volume of the universe proportional to the product $a_1a_2a_3$, keeps decreasing monotonically as η decreases.

Misner [76] gave a convenient picture of understanding this oscillating behaviour in the approach to singularity in terms of a dynamics of a point moving in a two-dimensional *anisotropy plane.* Introduce the parametrization:

$$\alpha_1 := -\Omega + \beta_+ + \sqrt{3}\beta_-, \quad \alpha_2 := -\Omega + \beta_+ - \sqrt{3}\beta_-, \quad \alpha_3 := -\Omega - 2\beta_+. \tag{9.10}$$

Observe that the volume $\sim a_1a_2a_3 = e^{-3\Omega}$ and the singularity (vanishing volume) is approached as $\Omega \to +\infty$.

Substitution in (9.9) and eliminating $\sum_i \alpha_i''$ we get,

$$\sum_i \alpha_i' = -3\Omega' \quad , \quad \sum_i \alpha_i'^2 = 3\Omega'^2 + 6\beta_+'^2 + 6\beta_-'^2$$

$$\begin{aligned} (\sum_i \alpha')^2 - \sum_i \alpha_i'^2 &= \frac{1}{2}\left(\sum_i a_i^4 - 2(a_1a_2a_3)^2 \sum_i a_i^{-2}\right) \qquad (9.11) \\ \Omega'^2 - \beta_+'^2 - \beta_-'^2 &= \frac{e^{-4\Omega}}{12} V(\beta_+, \beta_-) \qquad \text{where,} \\ V(\beta_+, \beta_-) &:= e^{-8\beta_+} - 4e^{-2\beta_+}\cosh(2\sqrt{3}\beta_-) + \\ & \quad 2e^{4\beta_+}(\cosh(4\sqrt{3}\beta_-) - 1) \qquad (9.12) \end{aligned}$$

Dividing by Ω'^2 we can write the last equation in terms of variation of the anisotropy parameters $\beta_\pm$ with the volume Ω as,

$$1 - \left(\frac{d\beta_+}{d\Omega}\right)^2 - \left(\frac{d\beta_-}{d\Omega}\right)^2 = \left(\frac{e^{-4\Omega}}{12\Omega'^2}\right) V(\beta_+, \beta_-). \qquad (9.13)$$

From the first of the equation (9.9), we also have,

$$\Omega'' = -\frac{1}{6} e^{-4\Omega} V(\beta_+, \beta_-) \qquad (9.14)$$

The equations have the Ω'. We can eliminate it by introducing a new variable, $\Lambda(\Omega) := \Omega'^2$. This gives $\frac{d\ln\Lambda}{d\Omega} = 2\frac{\Omega''}{\Omega'^2}$. Eliminating the double derivative of Ω, we write the $\Omega-$evolution system as,

$$\begin{aligned} 1 - \left(\frac{d\beta_+}{d\Omega}\right)^2 - \left(\frac{d\beta_-}{d\Omega}\right)^2 &= \left(\frac{e^{-4\Omega}}{12\Lambda}\right) V(\beta_+, \beta_-), \\ \frac{d\ln\Lambda}{d\Omega} &= -\left(\frac{e^{-4\Omega}}{3\Lambda}\right) V(\beta_+, \beta_-). \qquad (9.15) \end{aligned}$$

The potential is invariant under a rotation by $2\pi/3$ in the anisotropy plane and, as the $\beta_+ \to -\infty$ behaviour shows, it has exponentially rising 'walls'. The height of a wall at a given location in the anisotropy plane, *decreases* as volume diminishes. Thus the walls can be seen as moving *outward.* The Bianchi-IX evolution can now be qualitatively understood as the bouncing of the system point $(\beta_+, \beta_-)(\Omega)$ against the receding potential walls. The bouncing continues *ad-infinitum* since the speed of the system point is *larger* than the wall speed [76]. Away from the walls the potential can be neglected and the system point follows the Kasner behaviour. Thanks to the shuffling of Kasner epoch interspersed with the reflections, this behaviour has been termed the 'mixmaster' behaviour.

This feature of the approach to the singularity revealed by the vacuum Bianchi-IX model is expected to retain its qualitative behaviour even in the presence of matter as argued in [74]. In general the matter would add non-diagonal pieces of the metric as well and close to the singularity, this results in the changes in the directions of the axes of the scale factors. This is in addition to changes in Kasner epochs relative to a fixed set of axes.

These studies revealed that apart from a power law approach to the singularity as seen in the Kasner solution (scale factors behaving as power of the synchronous time τ), there is another possibility where *no power law* approach is possible. The asymptotic behaviour of the scale factors can be oscillatory and indeed more strongly, chaotic.

Based on these studies, around 1970, Belinski–Khalatnikov–Lifshitz (BKL) [77–79] analysed the vacuum Einstein equation near a presumed space-like singularity. The question posed was: Is there *a general solution* of the equation which is singular? Here singularity was understood as diverging curvature invariants and/or matter invariants (energy density etc.) and *not as any geodesic incompleteness.* By general solution they meant that the solution characterising the behaviour has *four* arbitrary functions on a spatial slice[2].

The asymptotic analysis was carried out with the choice of the synchronous time, with an ansatz of the form,

$$\begin{aligned} g_{IJ}(\tau, x) &= a^2 l_I l_j + b^2 m_I m_J + c^2 n_I n_J \\ \text{with} \quad & a^2 = \tau^{2p_1(x)}\ ,\ b^2 = \tau^{2p_2(x)}\ ,\ c^2 = \tau^{2p_3(x)}, \\ \text{and} \quad & \sum_i p_i(x) \ = \ 1, \quad \sum_i p_i^2(x) \ = \ 1. \end{aligned} \tag{9.16}$$

The directions specified by l_I, m_I, n_I are space dependent too. The solution of the above form is analysed in the limit $\tau \to 0$ as an expansion in powers of τ. This ansatz has 10 unknown functions of the space coordinates: the three components of each of l, m, n vectors and one function in the three p_i's satisfying 2 equations. The Gaussian form of the space-time metric is preserved by making 3 purely spatial coordinate transformations and of course there are 3 constraint equations $R_{\tau i} = 0$. This leaves 4 unrestricted functions of spatial coordinates.

Note that the ansatz contains as a special case the Bianchi-IX metric which is known to be singular as discussed above, and has the oscillatory approach to the singularity. The subsequent detailed analysis shows that as $\tau \to 0$, the spatial derivatives may be neglected making the space-time effectively homogeneous (in a small coordinate patch). The equations themselves then resemble the Bianchi-IX and therefore exhibit the same oscillatory approach to the singularity. This forms the basis for the BKL conjecture informally phrased as:

---

[2]The counting is most clearly seen in the Hamiltonian formulation. There are 12 phase space functions and 4 first class constraints. This leaves 12 - 4 - 4 = 4, freely specifiable function on the 3-manifold. See chapter 11.

Near a presumed space-like singularity, approached as $\tau \to 0$*, the general inhomogeneous solution becomes fragmented into locally homogeneous patches each of which evolves independently, in a manner similar to the vacuum Bianchi-IX space-time.*

This remains a conjecture and continues to be investigated. For a review of numerical approach and status, please see [80] and for a recent critical discussion within the Hamiltonian framework please see [81, 82].

The singularity theorems formulate conditions under which incomplete causal geodesic(s) *must exist* but do not give any further details of the nature of singularity while the BKL approach and conjecture asserts that *general solutions* with space-like singularity *do* exist, have divergent curvature invariants and the approach to singularity is oscillatory. It does not give general conditions under which singularities must exist.

Chapter 10

Gravitational Waves

We have discussed gravitational waves first as an example of non-stationary space-time and then as possible source of energy loss form dynamical matter sources, both in the context of *linearized model.* Historically, the inherent non-linear nature of the field equations as well as the huge coordinate freedom, made it very confusing to reliably assert the 'reality of gravitational waves'. An interesting account of the history related to this confusion may be seen in [83]. In the next section, we will try to gain an appreciation of the issues and their resolutions.

10.1 Conceptual Issues

One of the early and distinctive prediction of general relativity was the *possible existence* of gravitational radiation. The linearized Einstein equation resembles a wave equation and admits plane wave solutions. Furthermore, the analysis of the inhomogeneous equation with bounded matter distribution, reveals that 'accelerated' matter sources radiate energy with power related to the third time derivative of the quadrupole moment of the matter distribution - the quadrupole formula. While this was given by Einstein already in 1916, several doubts arose. Is the radiative solution of the linearized equation a genuine prediction of the exact theory or is it a spurious solution with no corresponding solution of the exact equation to which this is an approximation? Is the ripple character of the solution only an artifact of choice of coordinates? The measure of energy used is not a tensorial quantity (Einstein pseudo-tensor), is it then a reliable measure? Since causality is contingent on the metric and apart from analogy with electromagnetism there is no empirical guidance, which Green function is appropriate to use in the computation of the inhomogeneous solution?

The first issue was tackled and an exact solution of the Einstein equation was given by Einstein and Rosen - now known as the Einstein-Rosen cylindrical waves. So the exact equation does have a wave-like solution, however other issues remained. The ripple character was coordinate dependent (Eddington 1922 - "gravitational waves propagate at the speed of thought") [83]; the measures of energy-momentum pseudo-tensors vanish for this solution; the solution was 'unstable' as revealed by motion of test particles. Once again, the main factors contributing to the issues were the coordinate dependence and lack of definition of localized gravitational energy.

The Bondi-Feynman bead argument at least settled the issue that gravitational waves are capable of 'carrying energy' or at least are capable of doing work on physical system - beads on a rod, suitably oriented would be rubbed against the rod thereby generating heat. The efforts to characterize the gravitational waves, shifted from the metric to the curvature tensor - essentially the Weyl curvature since the Ricci is zero for vacuum solutions [84]. Here, one more point is to be noted. It is not just wave-like solutions 'propagating' that one wants, but more specifically one is looking for the characterization of *radiation emitted by bounded sources* [85]. Such a physical system of bounded sources could also have incoming radiation. This is to be *excluded* in attributing radiation to specific sources. Sachs attempted to capture this by proposing a boundary condition for out-going radiation [86] and concluded that this condition is equivalent to the Weyl tensor exhibiting what is subsequently called the *peeling property* [87], namely, along any null geodesic with an unbounded affine parameter, r, the leading $o(\frac{1}{r})$ curvature has Petrov type N, the next leading has type III, the next has type II, the $o(r^{-4})$ has type I while the order r^{-5} has no relation to the geodesic. This lead to the identification of radiative field as the $o(\frac{1}{r})$ piece which is given by the Weyl component Ψ_4. The Petrov classification is discussed in section 14.8.

Finally, Bondi and his collaborators [43] proceeded to analyse asymptotic fields and identify the radiative parts of the asymptotic fields. They introduced a coordinate system (u, r, θ, ϕ) where u, θ, ϕ label light rays out-going from some fiducial point on the axis of symmetry (they assumed axial symmetry) while r is an affine parameter along the light-like geodesic. This was used to restrict the metric to four functions of three coordinates. The vacuum equations were then simplified and analysed in detail. Further imposition of the 'out-going wave condition' (expansion in inverse powers of r with coefficients functions only of u and θ) and use of further coordinate transformations allowed such metrics to be fully determined in terms of a single function $c(u, \theta)$. Considering the physical situation to start from staticity, have some transients for a finite duration and return again to staticity, they were able to identify a *mass aspect function* which coincided with the mass in the static situation and showed a monotonic decrease when the *news function*, $\partial_u c(u, \theta)$ is non-zero. The entire analysis was at the full non-linear level, however, a linearization was also shown to reproduce the quadrupole formula. The outcome of this analysis and its further refinements (e.g. removing axial symmetry assumption etc.) established that (a) there exists solution of the full non-linear vacuum equation; (b) it can be interpreted to be associated with a bounded, transient source; (c) presence of radiation is unambiguously detected by non-zero news function which also shows mass loss; (d) this mass loss can only be attributed to energy being carried away via gravitational waves.

This established that gravitational radiation does exist with all the physically expected attributes.

How does this help in the *detection* of gravitational waves? The above analysis establishes that general relativity *does* make a physically sustainable prediction of gravitational waves and puts the quadrupole formula, which is used for estimation of expected amplitudes of gravitational waves, on a conceptually reliable footing. The analysis also provides analytical checks on numerical simulations involving strong field regimes (e.g. merger and ring down phases), for instance, by computing the news function for correlating energy loss with back reaction on source motions.

Finally, the observation of orbital decay of the Hulse–Taylor binary pulsar has established that the decay is fully consistent with the energy radiated as per the quadrupole formula. This constitutes an evidence for gravitational radiation *and* also empirically shows that binary systems, even though only under gravitational forces, *can* be a source of gravitational radiation.

10.2 Observational Issues

Unlike the indirect conformation of gravitational waves, which is based on the loss of energy due to gravitational radiation, the approach for direct detection of gravitational waves is based on the *time varying tidal effects* caused by a passing gravitational wave. Consequently, the primary issues are: (i) the types of sources together with the characteristic amplitude and time dependence of their signal and (ii) the choice of test body/detector system and their *sensitivity parameter*. The secondary issues involve the identification of a signal and an estimate of expected detection rate. We will briefly describe these aspects. For a recent review and an excellent textbook please see [27,28].

Sources: As seen in the section (5.3), any matter distribution which is at least quadrupolar and has an accelerated motion is a potential source of generation of gravitational waves. Cataclysmic short duration events such as a supernova and other gravitational collapse produce a *burst* signal while long duration binary systems of compact objects produce *periodic* signals during their in-spiral phase. There are many individually sub-detection level sources which could produce a stochastic background. A cosmic gravitational background is also expected from the very early universe.

Amplitude and frequency estimates: These are based on the quadrupole formula, $h_{ij}^{TT}(t) = \frac{G}{c^6}\frac{2}{r}\frac{d^2}{dt^2}\int_{source}\rho x_i x_j$ which gives the amplitude *at the detector* when the source is at a distance r. Here $\rho = T_{00}$ is the energy density and we have restored the factors of G, c. The amplitude h_{ij} is *dimensionless*. There are three parameters associated with a localized (as distinct from a stochastic background) source - *mass, M*, of the source, a *length scale, L*, associated with the quadrupole and a *time scale, T*, characteristic of the time variation.

Hence on dimensional grounds, we can write for a typical component of the amplitude,

$$h \sim \frac{G}{c^4}\frac{ML^2T^{-2}}{r} \sim 10^{-44}\frac{ML^2T^{-2}}{r}$$

Here r is the distance to the source and we have used the mks units. The M, L, T are not always the total mass or the 'radius' or a period but are some fractions of these. Such geometry dependent numerical factors are absorbed in the M, L, T parameters. For short duration sources such as gravitational collapse, it is easier to estimate the total energy released E, the frequency of the gravitational waves, f, the duration over which the source is observed, T. Then using $\dot{h} \sim hf$ and average power $\sim E/T$, leads to an estimate of the amplitude as, $h \sim \frac{1}{\pi r f}\sqrt{\frac{E}{T}}\sqrt{\frac{G}{c^3}} \sim \frac{10^{-18}}{\pi r f}\sqrt{\frac{E}{T}}$ in MKS units [28].

The supernovae sources are at distances in the Kpc $\sim 10^{19}m$ (in our galaxy) to Mpc (in other galaxies) range. Their estimated event rate is quite low - roughly once in fifty years or so for a Milky Way type galaxy. The properties of supernovae with regards to frequency of the gravitational waves and the energy carried by them, is estimated from simulation which indicate the typical numbers to be $\frac{E}{c^2} \sim 10^{-7}M_\odot \sim 10^{23}$ Kg, $f \sim$ kHz, $T \sim$ millisecond. This leads to an amplitude of about $h \sim 10^{-21}$.

For isolated pulsars as well as binaries of stars and stellar mass black holes, the distance is again in the 10 Kpc - Mpc range. For pulsars, the effective mass parameter would be about $10^{-3}M_\odot \sim 10^{27}kg$, $L \sim 10^4 m$ and $T \sim 10^{-3}s$ leading to $h \sim 10^{-23}$. For long duration, sources of periodic signal, the effective amplitude is actually larger thanks to matched filtering method of signal extraction. If n is the number of cycles of the signal contained in the observation period of T, then the effective amplitude is $h_{eff} \sim \sqrt{n}h$. For a signal of frequency f, observed for time T, the number of cycles is $n = fT$.

The most promising and studied candidates are binary systems. For a mass M spherical object, the radius of last stable circular orbit is about 3 times the Schwarzschild radius. For binaries made up of neutron stars or black holes, the binaries could be quite tight with L closer to the radius of the last stable circular orbit. These are called *coalescing binaries*. For binaries involving white dwarfs or normal stars, the L would be quite large and are called *in-spiraling binaries*. We can eliminate the binary radius by the angular frequency using Kepler's law to get $h \sim \frac{10^{-55}}{r}M^{5/3}\Omega^{2/3}$. For $M \sim M_\odot$ and $\Omega \sim 10^{-4}$, we get $h \sim 10^{-28}$. For the last stable orbit, the angular frequency for a solar mass object would be about $\Omega \sim 10^4$ leading to $h \sim 10^{-22}$. Although the amplitude for white dwarf binaries is quite small, they are nearer and amenable to enhancement through matched filtering. Coalescing binaries of super-massive black holes too are candidates at frequencies of the order of mHz. For coalescing binaries, there is also the possibility of merger into a black hole which then *rings down* to its stationary state. These ringing frequencies, called quasi-normal modes, are characteristic of the black hole parameters.

The amplitudes during this phase turn out to be sizable and vary between about $10^{-21} - 10^{-17}$ even over several hundred mega-parsec distances.

Finally, there are the stochastic gravitational waves which are made up of incoherent superposition of a large number of sources as well an expected *isotropic* component as a relic from the very early universe. Here the study is not by measuring an amplitude, but rather by studying the frequency spectrum of the gravitational energy density or more precisely the quantity, $\Omega_{gw}(f) := \rho_c^{-1}\frac{d\rho_{gw}}{dlogf}$ where $\rho_c := \frac{3c^2H_0^2}{8\pi G}$ is the cosmological critical energy density. The expected frequencies range over $10^{-18} - 10^9$ Hz [27].

The upshot is that the expected amplitude or effective amplitude from various sources is about $h \sim 10^{-21}$ or smaller while the frequencies vary between mHz to kHz for individual sources but over a vast range for stochastic background.

Detection methodology: Since the detection method is based on tidal distortions of bodies, the earliest method proposed by Weber, was to use a *Resonant bar*. The idea is that a strain produced in a system will make the system vibrate with its fundamental frequency. For an aluminium cylinder of length ~ 3 meters and mass of 1000 kg has its resonant frequency in the range of 500–1500 Hz. The amplitude of this vibration will be set by the gravitational wave to be $\sim 10^{-21} \times 10^3$ meters. This is very tiny and is smaller than or comparable to three main sources of triggers - thermal excitations, noise in the amplification process and the quantum uncertainty. Even at low temperatures of tenth of a Kelvin, the rms amplitude of thermal fluctuations is about 6×10^{-18} meters. With a very narrow bandwidth around the fundamental frequency (Q factor of $\sim 10^6$), it is possible to have the duration of the signal to be short (10^{-3} sec.) enough so that the noise amplitude reaches only about a thousandth of its rms value, thereby permitting a signal detection of $h \sim 10^{-20}$. The noise in the amplification process can also be managed for lower frequencies $\sim 10^2$ Hz. The quantum mechanical zero point fluctuation $\sim \sqrt{\hbar/(2M\omega)} \sim 10^{-21}$ meters. So as thermal noise is reduced, the quantum noise begins to challenge. Squeezing of uncertainty in a different observational procedure is a possible option. Apart from the standard bar configuration, spherical resonant bodies have also been used which can have more mass in a smaller volume and also have sensitivity in all directions.

Another type of detector uses light beams between a transmitter and a receiver at different locations and attempts to detect the slight fluctuations in the arrival rates due to the distortion in the *physical path length*[1]. A passing gravitational wave causes the proper length traversed by the light beam to change and hence its arrival time. The rate of light pulses received therefore

[1]The form of the gravitational wave is the simplest in the TT-gauge. This gauge corresponds to a freely falling coordinate frame with the coordinate time being the proper time of the freely falling observer. In this gauge, the spatial coordinates of a particles initially at rest, do not change as can be seen from the geodesic equation. The physical lengths however do change [27].

changes from the rate of emitted pulses. Measuring this rate gives a detection of a gravitational wave [28, 88].

Clearly, this depends on availability of very precise time-stamping. The best clocks with stability of few parts in 10^{16} can detect an amplitude of about 10^{-15}. Pulsars themselves are comparably stable and hence can be used for time-stamping for the emitted pulses. Simultaneous observations of several pulsars over long periods can detect very low frequency ($\sim$ nano Hertz) gravitational waves.

Essentially the same logic holds in interferometric detectors whose two arms change their physical lengths by different amounts producing interference fringes. Many collaborative efforts are built around Michelson-Morley type interferometer using lasers. Since the expected highest frequency is about kHz, the wavelengths are larger than 300 kms. It is impossible to build an interferometer with comparable arm length. The arms of the current Earth based interferometers are in the range of 300 meters (TAMA) and 600 meters (GEO) meters to 4 kilometers (LIGO). It is possible to effectively increase the arm length a hundred fold by making the laser beam make a 100 traversals in a Fabry-Parot cavity before producing fringes. A longer arm length has the advantage that length determination needs to be within about 10^{-16} meters which is smaller than the size of a nucleus! No mirror can be ground to this degree of smoothness. Here the fact that the laser beam has a width means that individual rays reflect from different irregularities on the mirror surface. Averaging the lengths over the beam cross-section, measures the coherent movement and tiny changes in these averages averages can be determined. Being fixed to the Earth, there are many sources of noise e.g. fluctuations in the gravitational field due to seismic shifts and other movements of mass. These are controlled with suspensions and filtered out selecting frequency windows. As mentioned above for the resonant bars, thermal noise is reduced by keeping the mirrors at cryogenic temperatures or by choosing material for suspension fibres. The quantum noise due to Poisson statistics obeyed by the laser photons, called the *photon shot noise* is a limiting noise which is sought to be minimized by using squeezing. For an extensive discussion of possible noise sources and their control or avoidance, please see [27].

Suffice it to say that extracting an unambiguous signal of gravitational waves from some astronomical source from a variety of noises larger than the signal, is a daunting task requiring sophisticated data analysis techniques as well as a ‘bank of templetes’ of expected waveforms for use of matched filtering. The requirements are being met and there is talk of gravitational wave astronomy using data from multiple detectors.

Chapter 11

Field Equation: Evolutionary Interpretation

The field equation stipulates conditions on the space-time geometry as a complete entity. In practice though, we have clear sense of a dynamical evolution of matter. Can this sense be extended to a dynamical view of the space-time itself? In other words, can the set of space-time events be viewed as a sequential arrangement of some three-dimensional entities? We expect such a view to be supported on physical grounds, but does the manifest, space-time covariant mathematical formulation accommodate such a view? This is far from automatic. It turns out that *Einstein equation* does admit such an evolutionary interpretation - a solution space-time *can be* viewed as an evolving *spatial geometry.*

Recall from the discussion of causality and determinism, globally hyperbolic space-times have the property of admitting a Cauchy surface such that data recorded on it has a possibility of determining *completely* the data throughout the entire space-time. These space-times are likely to provide us the clues for seeking an evolutionary view of space-time. We study implications of global hyperbolicity and then reverse the process to get the 'initial value formulation' of the Einstein equation. The notation and presentation follows closely [17].

11.1 The 3 + 1 Decomposition

Let (M, g) be a globally hyperbolic space-time, not necessarily a solution of Einstein equation, and let $T : M \to \mathbb{R}$ be a time function (not unique) which is guaranteed to exist. Let Σ denote a T =constant hypersurface. Let $n_\mu := f\partial_\mu T$ be the unit, time-like normal to Σ, $n \cdot n = -1$. Let $h_{\mu\nu} := g_{\mu\nu} + n_\mu n_\nu$ which satisfies, $h_{\mu\nu}n^\nu = 0$, $h^\mu{}_\alpha h^\alpha{}_\nu = h^\mu{}_\nu$. Thus $h^\mu{}_\nu$ is a projection operator which projects tangent vectors to M to tangent vectors to Σ. On the subspace, $T_p(\Sigma) \subset T_p(M)$, $h_{\mu\nu}$ is a positive definite metric. This is referred to as the *induced metric* on Σ or the *first fundamental form* of Σ. Using these available structures, we carry out a '3 + 1' decomposition i.e. express space-time geometrical quantities in terms of spatial geometrical quantities.

Projection of tensors: Given an arbitrary tensor $T^{\mu_1,\dots,\mu_m}{}_{\nu_1,\dots,\nu_n}$ we define a corresponding *spatial* tensor as,

$$\bar{T}^{\mu_1,\dots,\mu_m}{}_{\nu_1,\dots,\nu_n} \quad := \quad h^{\mu_1}{}_{\alpha_1}\dots h^{\beta_n}{}_{\nu_n} T^{\alpha_1,\dots,\alpha_m}{}_{\beta_1,\dots,\beta_n}$$

Due to the explicit factors of h, contraction of any index with n will given zero. Hence $\bar{T}$ is spatial.

Next, we want to define a covariant derivative operator $\bar{\nabla}$ which will produce a spatial tensor upon acting on a spatial tensor. Observe that if we have a vector X^μ satisfying $n \cdot X = 0$, it does not follow that $\nabla_\mu X^\nu$ satisfies $n \cdot \nabla X^\nu = 0$ or $n_\nu \nabla_\mu X^\nu$. The first type of term involves derivative 'off'-Σ while the second type of term shows that spatial derivative of a spatial tensor need not be spatial. Therefore we need to make explicit projections for both the derivative as well as the differentiated tensor and we define,

$$\bar{\nabla}_\lambda \bar{T}^{\mu_1,\dots,\mu_m}{}_{\nu_1,\dots,\nu_n} \;:=\; h^{\mu_1}{}_{\alpha_1}\dots h^{\beta_n}{}_{\nu_n} h_\lambda{}^\rho \nabla_\rho \bar{T}^{\alpha_1,\dots,\alpha_m}{}_{\beta_1,\dots,\beta_n}$$

The 'bar' on the tensor is put to remind us that the $\bar{\nabla}$ is defined only for spatial tensors which could have been constructed from space-time tensors via projections. Notice that there is one factor of h for each (un-summed) index. This spatial covariant derivative satisfies the same properties of the usual space-time covariant derivative namely, (i) it is linear, (ii) satisfies Leibniz rule, (iii) on functions depending on Σ alone, reduces to ordinary spatial derivative ($\bar{\nabla}_\alpha f = h_\alpha{}^\beta \nabla_\beta f = \partial_\alpha f + n_\alpha n \cdot \partial f =: \bar{\partial}_\alpha f$), (iv) is torsion free (since ∇ is) and (v) $\bar{\nabla}_\lambda h_{\mu\nu} = 0$ i.e. 'metric compatible'. Such a derivative operator on Σ is uniquely defined.

Armed with a spatial covariant derivative operator, we can define the corresponding Riemann tensor from commutator of these derivatives:

$$\begin{aligned}
\bar{\nabla}_\mu \bar{\nabla}_\nu \bar{A}^\lambda \quad &= \quad h_\mu{}^\alpha h_\nu{}^\beta h^\lambda{}_\gamma \nabla_\alpha (\bar{\nabla}_\beta \bar{A}^\gamma) \\
&= \quad h_\mu{}^\alpha h_\nu{}^\beta h^\lambda{}_\gamma \nabla_\alpha (h_\beta{}^\rho h^\gamma{}_\sigma \nabla_\rho \bar{A}^\sigma) \\
&= \quad h_\mu{}^\alpha h_\nu{}^\rho h^\lambda{}_\sigma \nabla_\alpha \nabla_\rho \bar{A}^\sigma + h_\mu{}^\alpha h_\nu{}^\rho h^\lambda{}_\gamma (\nabla_\alpha h^\gamma{}_\sigma) \nabla_\rho \bar{A}^\sigma \\
&\qquad + h_\mu{}^\alpha h_\nu{}^\beta h^\lambda{}_\sigma (\nabla_\alpha h^\rho{}_\beta) \nabla_\rho \bar{A}^\sigma \\
&= \quad h_\mu{}^\alpha h_\nu{}^\rho h^\lambda{}_\sigma \nabla_\alpha \nabla_\rho \bar{A}^\sigma + h_\mu{}^\alpha h_\nu{}^\rho h^\lambda{}_\gamma n_\sigma (\nabla_\alpha n^\gamma) \nabla_\rho \bar{A}^\sigma \\
&\qquad + h_\mu{}^\alpha h_\nu{}^\beta h^\lambda{}_\sigma n^\rho (\nabla_\alpha n_\beta) \nabla_\rho \bar{A}^\sigma \\
&= \quad h_\mu{}^\alpha h_\nu{}^\rho h^\lambda{}_\sigma \nabla_\alpha \nabla_\rho \bar{A}^\sigma + (h_\mu{}^\alpha h^\lambda{}_\gamma \nabla_\alpha n^\gamma) n_\sigma h_\nu{}^\rho \nabla_\rho \bar{A}^\sigma \\
&\qquad + (h_\mu{}^\alpha h_\nu{}^\beta \nabla_\alpha n_\beta) h^\lambda{}_\sigma n^\rho \nabla_\rho \bar{A}^\sigma \\
&:= \quad h_\mu{}^\alpha h_\nu{}^\rho h^\lambda{}_\sigma \nabla_\alpha \nabla_\rho \bar{A}^\sigma + K_\mu{}^\lambda h_\nu{}^\rho n_\sigma \nabla_\rho \bar{A}^\sigma \\
&\qquad + K_{\mu\nu} h^\lambda{}_\sigma n^\rho \nabla_\rho \bar{A}^\sigma
\end{aligned} \tag{11.1}$$

We have defined the *second fundamental form or extrinsic curvature* of Σ, $K_{\mu\nu} := h_\mu{}^\alpha h_\nu{}^\beta \nabla_\alpha n_\beta$, $n_\alpha = f \partial_\alpha T$. Under anti-symmetrization in μ, ν, the last term vanishes, since

$$K_{\mu\nu} - K_{\nu\mu} = h_\mu{}^\alpha h_\nu{}^\beta(\nabla_\alpha n_\beta - \nabla_\beta n_\alpha) = h_\mu{}^\alpha h_\nu{}^\beta(\partial_\alpha n_\beta - \partial_\beta n_\alpha) \ , \text{ and}$$
$$\partial_\alpha n_\beta - \partial_\beta n_\alpha = (\partial_\alpha f)(\partial_\beta T) - (\partial_\beta f)(\partial_\alpha T) = (\partial_\alpha \ln f) n_\beta - (\partial_\beta \ln f) n_\alpha$$

In the middle term, ∇_ρ can be flipped on n_σ and using $K_{\nu\sigma} = h_\nu{}^\rho h_\sigma{}^\beta \nabla_\rho n_\beta = h_\nu{}^\rho \nabla_\rho n_\sigma$, the middle term, after antisymmetrization, can be written as $-(K_\mu^\lambda K_{\nu\sigma} - K_\nu^\lambda K_{\mu\sigma})\bar{A}^\sigma$. Finally, the first term gives the space-time Riemann tensor. Stripping off $\bar{A}^\sigma$, we obtain,

$$\bar{R}^\lambda{}_{\sigma\mu\nu} = h^\lambda{}_\rho h_\sigma{}^\tau h_\mu{}^\alpha h_\nu{}^\beta R^\rho{}_{\tau\alpha\beta} - K^\lambda{}_\mu K_{\nu\sigma} + K^\lambda{}_\nu K_{\mu\sigma} \tag{11.2}$$

This is one of the *Gauss–Codacci* equations which relates the Riemann tensor of M to the Riemann tensor of the first fundamental form, $h_{\mu\nu}$, of a spatial hypersurface and its second fundamental form, $K_{\mu\nu}$. The second of the Gauss–Codacci equations is obtained in a similar manner and relates the spatial covariant derivatives of the extrinsic curvature of Σ to the Ricci tensor of the space-time:

$$\bar{\nabla}_\mu K^\mu{}_\nu - \bar{\nabla}_\nu K^\mu{}_\mu = h_\nu{}^\alpha R_{\alpha\beta} n^\beta \tag{11.3}$$

Consider now the Einstein tensor, $G_{\mu\nu}$ and its projections:

$$\bar{G}_{\mu\nu} := h_\mu{}^\alpha h_\nu{}^\beta G_{\alpha\beta} \ , \ h_\mu{}^\alpha G_{\alpha\beta} n^\beta \ \text{ and } \ G_{\mu\nu} n^\mu n^\nu .$$

Using the definitions, it follows immediately that $R_{\mu\nu\alpha\beta} h^{\mu\alpha} h^{\nu\beta} = 2 G_{\mu\nu} n^\mu n^\nu$. The Gauss–Codacci equations imply,

$$\begin{aligned} h_\mu{}^\alpha G_{\alpha\beta} n^\beta &= h_\mu{}^\alpha R_{\alpha\beta} n^\beta = \bar{\nabla}_\nu K^\nu{}_\mu - \bar{\nabla}_\mu K^\nu{}_\nu \quad \text{and,} & (11.4) \\ \bar{R} &= h^{\alpha\beta} \bar{R}_{\alpha\beta} = h^{\alpha\beta} \bar{R}^\lambda{}_{\alpha\lambda\beta} \\ &= h^{\alpha\beta} \left[h_\alpha{}^\mu h_\beta{}^\nu h^\lambda{}_\rho h_\lambda{}^\sigma R^\rho{}_{\mu\sigma\nu} - K_\lambda{}^\lambda K_{\beta\alpha} + K_\beta{}^\lambda K_{\lambda\alpha} \right] \\ &= h^{\mu\nu} h^{\rho\sigma} R_{\rho\mu\sigma\nu} - (K_\alpha{}^\alpha)^2 + K_{\alpha\beta} K^{\alpha\beta} \\ \therefore \ G_{\mu\nu} n^\mu n^\nu &= \frac{1}{2} \left(\bar{R} + K^2 - K_{\mu\nu} K^{\mu\nu} \right) \quad , \quad K := K_\alpha{}^\alpha & (11.5) \end{aligned}$$

All these are straightforward consequences of global hyperbolicity of (M, g), the Einstein equation has not been used. Now we have the following theorem [17]:

Theorem 11.1 *Let $J_\mu := G_{\mu\nu} n^\nu$ and $\bar{G}_{\mu\nu} := h_\mu{}^\alpha h_\nu{}^\beta G_{\alpha\beta}$. If $\bar{G}_{\mu\nu} = 0$ everywhere and $J_\mu = 0$ on a Σ_{t_0}, then (a) $J_\mu = 0$ everywhere and (b) $G_{\mu\nu} = 0$ everywhere. Furthermore, J_μ has no term involving $(n \cdot \partial)^2$.*

The proof goes through the following steps. $\bar{G}_{\mu\nu} = 0$ everywhere, implies that $-G_{\mu\nu} = n_\mu J_\nu + n_\nu J_\mu + n_\mu n_\nu (n \cdot J)$ everywhere. Therefore $J_\mu = 0$ everywhere will immediately give $G_{\mu\nu} = 0$ everywhere. Next, the contracted Bianchi identity, $\nabla^\mu G_{\mu\nu} = 0$ everywhere, leads to,

$$\begin{aligned} (\nabla \cdot n) J_\nu + n \cdot \nabla J_\nu + J \cdot \nabla n_\nu + n_\nu \nabla \cdot J + n_\nu (n \cdot \nabla)(n \cdot J) & \\ + (n \cdot J) n \cdot \nabla n_\nu + (n \cdot J) n_\nu \nabla \cdot n = 0 & \qquad (11.6) \end{aligned}$$

Contracting with n^ν leads to,

$$\nabla \cdot J + J^\nu n \cdot \nabla n_\nu = 0 \tag{11.7}$$

Both equations are valid everywhere. We can eliminate $\nabla \cdot J$ from the second equation and see that the first equation is of the form $n \cdot \nabla J_\nu = J^\mu X_{\mu\nu}(n)$ and holds everywhere. These are coupled, first order, homogeneous equations for J_μ and evolves it off Σ_t. Hence, if $J_\mu = 0$ on Σ_{t_0} then it is zero everywhere. This establish (a) and (b) of the theorem.

To show that J_μ has no $(n \cdot \partial)^2$ terms, split $\partial_\mu := \partial_\mu^{\|} + n_\mu D,\ D := n \cdot \partial$. Then the only terms that contains D^2 involve two derivatives and are of the form, $\partial_\mu \partial_\nu \sim n_\mu n_\nu D^2$. Now consider a coordinate expression for $G_{\mu\nu}$, keep only the second derivative terms and replace them by $n\ n\ D^2$. These terms have the form,

$$G_{\mu\nu} \sim \left(h_{\mu\nu} h^{\alpha\beta} D^2 g_{\alpha\beta} - h_\mu{}^\alpha h_\nu{}^\beta D^2 g_{\alpha\beta}\right) + \text{ single } D \text{ terms.}$$

The explicit factors of h give the result.

Thus, it is enough that the *six spatial* field equations hold everywhere and the remaining *four* equations $J_\mu = 0$ hold only at one Cauchy surface to imply that the full *ten* equations hold everywhere. In other words, there are redundancies in the Einstein equations. The last sentence of the theorem shows that the equations (11.4,11.5) denote *constraints on the data on* Σ and are *not* evolution equations.

The net conclusion of this analysis is that for the class of globally hyperbolic space-times, the vacuum Einstein equation can be split into *four* equations, $J_\mu = 0$, which are *constraint* equations that need to be imposed on one Cauchy surface and *six* evolution equations, $\bar{G}_{\mu\nu} = 0$. These are sufficient to ensure that the globally hyperbolic space-time is a solution of the vacuum Einstein equation.

The interesting feature now is that the 3 + 1 split can be *reversed* to construct a *space-time* solution of the Einstein equation! This is the Cauchy *Initial Value Theorem* for the Einstein equations.

11.2 Initial Value Formulation

The analysis of the previous section allowed us to identify the constraint equations and the evolution equations, albeit for the class of globally hyperbolic solutions. We now state the theorem which guarantees *existence and uniqueness* of globally hyperbolic *solutions*. We give the theorem as stated in [17].

Theorem 11.2 (Well-Posed Initial Value Problem) *Let Σ be a three-dimensional manifold, g_{ij} a Riemannian metric and K_{ij} a symmetric tensor* *on Σ. Let g_{ij}, K_{ij} satisfy*

$$\bar{\nabla}_i K^i{}_j - \bar{\nabla}_j K^i{}_i \ = \ 0 \quad \textit{and} \quad \bar{R} + (K^i{}_i)^2 - K_{ij}K^{ij} \ = \ 0$$

Then there exist a unique four-dimensional space-time (M, g) and an embedding $\phi : \Sigma \to \phi(\Sigma) \subset M$, such that,

1. $G_{\mu\nu}(g) = 0$.
2. *(M, g) is globally hyperbolic with $\phi(\Sigma)$ as a Cauchy surface for M.*
3. *The first and the second fundamental forms of $\phi(\Sigma)$ coincide with g_{ij}, K_{ij}, respectively, i.e. $g_{ij} = \phi^*(g_{\mu\nu}), K_{ij} = \phi^*(K_{\mu\nu})$.*

   The space-time (M, g) is called the maximal Cauchy development of (Σ, g_{ij}, K_{ij}) while Σ and (g_{ij}, K_{ij}) satisfying the constraint equations on Σ, are called the initial data.
4. *The space-time is unique in the sense that any other (M', g') satisfying the above three conditions, can be embedded isometrically into a subset of (M, g).*
5. *Let (Σ, g, K) and (Σ', g', K') be two initial data sets with (M, g) and (M', g') their maximal Cauchy developments with ϕ, ϕ' the corresponding embeddings. If $f : S \subset \Sigma \to S' \subset \Sigma'$ be a diffeomorphism taking the initial values into each other, then the domain of dependence of $\phi(S)$ is isometrically mapped to the domain of dependence of $\phi'(S')$.*
6. *The space-time metric $g_{\mu\nu}$ depends continuously on the initial values g_{ij}, K_{ij} (Well-posedness property).*

The proof consists of a series of steps relying on certain general existence, uniqueness and well-posedness properties of a class of partial differential equations. We sketch and list these steps and refer the reader to [17] for further details.

We are given Σ and the tensors $\bar{g}_{ij}, \bar{K}_{ij}$ relative to some local coordinates on Σ. Construct a four-dimensional manifold, M as a Cartesian product of an interval I and Σ. Let the local coordinates around a point $p \in \Sigma \subset M$ be denoted as $(t, x^i) \leftrightarrow (x^\mu)$ and arranged so that $t \in I$ and $t = 0$ gives the local portion of Σ. Our task is to show that a four-dimensional metric $g_{\mu\nu}(x)$ can be guaranteed to exist in the chart, with the requisite properties. To this end, we set $g_{ij}(0, \vec{x}) := \bar{g}_{ij}(\vec{x})$ and $\partial_t g_{ij}(0, \vec{x}) := \bar{K}_{ij}(\vec{x})$. The idea is to construct the metric as a *solution* of a type of a partial differential equation with the above as initial conditions. However, the Einstein equation is *not* directly of the required type to guarantee a desired solution. In order to arrive at the required type of equation, consider $H^\lambda(x)$ whose vanishing makes the coordinates to be *harmonic* relative to the metric.

$$H^\lambda := g^{\rho\sigma}\nabla_\rho\nabla_\sigma x^\lambda \ = \ -\frac{1}{2} g^{\alpha\beta} g^{\lambda\tau} \left(\partial_\beta g_{\tau\alpha} + \partial_\alpha g_{\tau\beta} - \partial_\tau g_{\alpha\beta}\right)$$

and

$$R_{\mu\nu} = \frac{1}{2}g^{\alpha\beta}\left(\partial_\alpha\partial_\mu g_{\nu\beta} + \partial_\alpha\partial_\nu g_{\mu\beta} - \partial_\alpha\partial_\beta g_{\mu\nu} - \partial_\mu\partial_\nu g_{\alpha\beta}\right) + \text{a function of } g, \partial g$$

Hence the combination, $R^H_{\mu\nu} := R_{\mu\nu} + g_{\lambda(\mu}\partial_{\nu)}H^\lambda$, has the form,

$$R^H_{\mu\nu} = -\frac{1}{2}g^{\alpha\beta}\partial_\alpha\partial_\beta\ g_{\mu\nu} + \text{a function of } g, \partial g \tag{11.8}$$

The equation, $R^H_{\mu\nu} = 0$ has *requisite form for which local existence, uniqueness and well-posedness properties hold.* The relevant theorems are given in the mathematical background 14, section 14.7. Hence, we are guaranteed a metric in the coordinate neighborhood with the portion of Σ as the Cauchy surface and with the specified initial data. Moreover, $R^H_{\mu\nu} = 0$ implies the vacuum Einstein equation, $R_{\mu\nu} = 0$, provided the coordinates are harmonic coordinates i.e. $H^\lambda(x) = 0$ in the local neighborhood.

We had left the time derivatives $\partial_t g_{0\mu}(x)$ as unspecified. Let us stipulate that these be such that $H^\mu(0, \vec{x}) = 0$. Now one shows that if $R^H_{\mu\nu} = 0$ in the coordinate neighborhood and the first constraint and $H^\mu = 0$ hold on Σ, then, $\partial_t H^\mu(0, \vec{x}) = 0$. The contracted Bianchi identity can be cast as an equation for H^μ which is again of the form that admits a unique solution given the initial data. In particular, for $H^\mu = 0 = \partial_t H^\mu$, it follows that $H^\mu(x) = 0$ in the neighborhood and now the solution of $R^H_{\mu\nu} = 0$ is also a solution of the vacuum Einstein equation. This is the basic step for proving the theorem.

There are a couple of points to be cleared. In applying the theorem 14.2, we have to start with a known solution (the ϕ^0_a of that theorem), namely, the flat space-time. The theorem then guarantees solution for data sufficiently close to the flat space-time. We have to go beyond. Secondly, the constructed solution is only local on Σ.

The first point is taken care of by a scaling trick. Observe that for a sufficiently small coordinate neighborhood of a point $p \in \Sigma$, we can always choose coordinates ($x^\mu(p) = 0$) such that the metric is Minkowskian: $g_{\mu\nu}(p) = diag(-1, 1, 1, 1)$ and $\partial_t g_{\mu\nu}(p) = 0$. Given an initial data, $g_{\mu\nu}(0, \vec{x}), \partial_t g_{\mu\nu}(0, \vec{x})$, not close to 'flat space data', consider new data $g' := \lambda^{-2}g, \partial_t g' := \lambda^{-2}\partial_t g$. Under these constant scalings, the new data satisfy the constraint equations, if the old one does. On this data, make a coordinate transformation, $x'^\mu := \lambda^{-1}x^\mu$. Then we obtain $g'_{\mu\nu}(x') = g_{\mu\nu}(\lambda x')$ and $\partial_{t'} g'_{\mu\nu}(x') = \lambda\partial_t g_{\mu\nu}(\lambda x)$. Clearly as $\lambda \to 0$, the new data approaches the data for flat space-time for which the local solution is guaranteed to exist. This solution, $g^0_{\mu\nu}(x')$ (say), evaluated at $\lambda_0^{-1}x$ is the solution for the original initial data.

For globalising the solution over Σ, one constructs local solutions in overlapping neighborhoods (can be chosen to be finitely many for paracompact Σ). Local uniqueness then permits consistent patching up of the local solutions.

Finally, the existence of maximal Cauchy development is proved by constructing a partial order on the set of all globally hyperbolic solutions with the same data (Σ, g, K) and invoking Zorn's lemma.

The theorem establishes an evolutionary view for the class of globally hyperbolic solutions of the vacuum Einstein equation. It reveals that a subset of the equations are non-dynamical (are constraints on the initial data) which in turn is understood as a consequence of the covariance under arbitrary coordinate transformations. It also leads to possibility of organizing this class of solutions by analysing the solutions of the constraint equations. This forms a basis for numerical solutions which are analytically intractable. More on this in the next chapter on numerical relativity.

11.3 Hamiltonian Formulation (ADM)

We saw that Einstein equation is a second order, 'quasi-linear', hyperbolic, non-linear, partial differential equation (in the harmonic gauge) and admits a Cauchy initial value formulation. It has the further property that it can be put in the form of first order Hamilton equations of motion, albeit with 'first class constraints' in Dirac's terminology. A Hamiltonian formulation provides a phase space view of general covariance and also opens up the possibility of canonical quantization of General relativity. We begin with the Einstein–Hilbert action[1],

$$S[g] \ = \ \int_M d^4x\sqrt{|g|}R(g) \quad , \quad |g| := -\det(g_{\mu\nu}) > 0 \ .$$

Let us quickly verify that Einstein equation follows from extremization of this action. For convenience, we take $g^{\mu\nu}$ as independent variable so that $\delta g_{\mu\nu} = -g_{\mu\alpha}\delta g^{\alpha\beta}g_{\beta\nu}$. Computing $\delta S := S[g^{\mu\nu} + \delta g^{\mu\nu}] - S[g^{\mu\nu}]$ to first order in $\delta g^{\mu\nu}$ we get,

$$\delta S[g] \ = \ \int_M d^4x\sqrt{|g|}\left[\delta g^{\mu\nu}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right) + \nabla_\lambda J^\lambda(\delta g^{\mu\nu})\right] \quad (11.9)$$

Here we have used: $\delta g = gg^{\mu\nu}\delta g_{\mu\nu} = -gg_{\mu\nu}\delta g^{\mu\nu}$ and $g^{\mu\nu}\delta R_{\mu\nu} = \nabla_\lambda J^\lambda$ with,

$$J^\lambda(\delta g^{\mu\nu}) = g^{\alpha\beta}\delta\Gamma^\lambda{}_{\alpha\beta} - g^{\lambda\alpha}\delta\Gamma^\beta{}_{\alpha\beta} \longleftrightarrow J_\lambda = \nabla^\mu\delta g_{\mu\lambda} - g^{\mu\nu}\nabla_\lambda\delta g_{\mu\nu} \quad (11.10)$$

The divergence term is a surface term whose vanishing (or not) will depend upon the boundary conditions specified. The metric fields which will represent physical space-times (i.e. 'solutions of equations of motions') are selected by requiring that arbitrary, first order variation in the action, about a space-time metric, receives contributions only from the boundary values of the metric.

[1] Note that the integrand is a scalar density of weight 1 and hence the action integral is well defined.

This immediately requires the physical metric to satisfy the Einstein equation: $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 0$. A few remarks are in order.

Remark 1: There are several different action principles to 'derive' the Einstein equation. Within the metric formulation of gravity, there are two versions - the one described above in which metric (or its inverse) is the independent variable and is sometimes referred to as the 'second order formulation'. It is possible to use the same action but treat the metric *and* the affine connection ${\Gamma^\lambda}_{\mu\nu}$ as independent variables. The corresponding variational principle is known as the *Palatini formulation.* There are two sets of Euler-Lagrange equations of motion one from $\delta g^{\mu\nu}$ and one from $\delta {\Gamma^\lambda}_{\mu\nu}$. The equation of motion from the variation of the affine connection determines the connection to be the Levi-Civita (torsion-free, metric compatible) connection up to an additive term. The equation of motion from the variation of the metric then gives the Einstein equation regardless of the additive term [89]. This is called a 'first order formulation'. There is another formulation of general relativity, especially when one has *spinorial* matter or *Fermions.* Then one uses tetrad field e^μ_I in place of the metric and a 'spin connection', ${\omega_\mu}^{IJ}$. Once again we have two formulations - one on which the spin connection is determined in terms of the tetrad (and co-tetrad) and a 'Palatini' form wherein both are treated as independent. At the level of classical equations of motion, all different formulations imply the Einstein equation. Corresponding Hamiltonian formulations are available as well, however, we will discuss only the metric formulation with the Levi-Civita connection.

Remark 2: We have not detailed the surface terms. While they are important in the passage to a canonical formulation (they can modify the symplectic structure), they play no role at the level of classical equations of motion. We comment on them at the end of the section.

To obtain a Hamiltonian formulation, we need to identify a 'time' and obtain a '3 + 1' decomposition of the space-time. Only after this is done, we can identify generalized velocities, define the generalized momenta and go over to a Hamiltonian form via the Legendre transform.

Let us assume that our would be space-time manifold is such as to admit a smooth function $T : \mathcal{M} \to \mathbb{R}$ such that the $T = \textit{constant}$ level sets, generate a foliation. Different possible T-functions will generate different foliations. For this to be possible, we must have $\mathcal{M} \sim \mathbb{R} \times \Sigma_3$.

Now choose a vector field $t^\mu\partial_\mu$ which is *transversal* to the foliation i.e. every integral curve of the vector field intersects each of the leaves, transversally. Furthermore, locally in the parameter of the curve, the leaves are intersected once and only once. *Normalize* the vector field so that $t^\mu\partial_\mu T = 1$. This ensures that values of the T-function can be taken as a 'time' parameter which we denote as t.

Fix a leaf Σ_{t_0} and introduce coordinates, $x^i, i = 1, 2, 3$ on it. Carry these along the integral curves of the vector fields, to the other leaves. This sets up a local coordinate system on $\mathcal{M}$ such that the normalized parametrization provides the coordinate t while the integral curves themselves are labelled by the $\{x^i\}$. Note that there is no metric so $\mathcal{M}$ is not yet a space-time. We have only set up a coordinate system.

Choose tensors g_{ij}, N^i, N on each of the leaves in a smooth manner and *define* a space-time metric via the line element:

$$ds^2 \ := \ -N^2 dt^2 + \bar{g}_{ij}\left(dx^i + N^i dt\right)\left(dx^j + N^j dt\right) . \tag{11.11}$$

Choosing $\bar{g}_{ij}$ to be *positive definite* and $N \neq 0$ ensures that the space-time metric $g_{\mu\nu}$ is *invertible*. Its inverse is given by,

$$\begin{aligned} g^{tt} \ &= \ -N^{-2} \ , \ g^{tj} \ = \ N^j N^{-2} \ , \\ g^{ij} \ &= \ \bar{g}^{ij} - N^{-2} N^i N^j \ , \quad \bar{g}^{ik}\bar{g}_{kj} = \delta^i_j \ . \end{aligned} \tag{11.12}$$

We now have a space-time. The space-time metric is defined in terms of 10 independent functions and so there is no loss of generality. It is a convenient parametrization for reasons given below, but alternative parametrizations are possible.

It follows that (i) The induced metric on the leaves is the Riemannian metric $\bar{g}_{ij}$.

(ii) $\tilde{n}_\mu := \partial_\mu T$ is 'normal' to the leaves, in the sense that for any tangent vector $X^\mu\partial_\mu$, to Σ_t, $X^\mu\tilde{n}_\mu = X^\mu\partial_\mu T = 0$. Thanks to the normalization of $t^\mu\partial_\mu$, we have $\tilde{n}_\mu = (1,0,0,0)$.

(iii) $\tilde{n}^\mu := g^{\mu\nu}\tilde{n}_\nu \Rightarrow \tilde{n}^\mu\tilde{n}_\mu = g^{tt} = -N^{-2} < 0$ and therefore the normal is *time-like* and hence the leaves are *space-like*. We take $n_\mu := \epsilon N\tilde{n}_\mu, N > 0,\ \epsilon = \pm 1$ to be the unit time-like normal.

(iv) The original transversal vector field can be decomposed as $t^\mu = an^\mu + \tilde{N}^\mu$ where $\tilde{N}^\mu n_\mu = 0$ and hence $\tilde{N}^\mu$ is tangential to the leaves and $\tilde{N}^0 = 0$. This decomposition refers to N as the *lapse* function and $\tilde{N}^\mu$ as the *shift vector*. Next, $t^\mu n_\mu = -N \Rightarrow a = N$. The sign is chosen so that t^μ and n^μ are *both* future (say) directed time-like vectors. The integral curve equation, $d_t x^\mu = Nn^\mu + \tilde{N}^\mu$ implies for $\mu = i$, $\tilde{N}^i = -Nn^i = -Ng^{it}\epsilon N = -\epsilon N^2(N^{-2}N^i) = -\epsilon N^i$. To identify the N^i with the shift vector (which is spatial) we choose $\epsilon = -1$. Thus, $n_\mu := -N\partial_\mu T$.

The particular parametrization of the space-time metric can be said to be *adapted* to the pre-selected coordinate system.

Since we have the unit time-like normal, n_μ to the hypersurface Σ, we can follow the steps used in the previous sections to express the action in terms of the parametrization. In particular, we note,

$$\begin{aligned} n_\mu \ , \ n^\mu \quad &: \quad n_t = -N \ , \ n_i = 0 \quad , \quad n^t = N^{-1} \ \ , \ n^i = -N^i N^{-1} \\ h_{\mu\nu} \quad &: \quad h_{tt} = \bar{g}_{ij}N^i N^j \ \ , \ \ h_{ti} = \bar{g}_{ij}N^j \ \ , \ \ h_{ij} = \bar{g}_{ij} \\ h_\mu{}^\nu \quad &: \quad h_t{}^t = 0 \ \ , \ \ h_i{}^t = 0 \ \ , \ \ h_t{}^i = N^i \ \ , \ \ h_i{}^j = \delta_i{}^j \ . \end{aligned}$$

Recalling,

$$R = 2(G_{\mu\nu} - R_{\mu\nu})n^\mu n^\nu = \bar{R} + K^2 - K_{\mu\nu}K^{\mu\nu} - 2R_{\mu\nu}n^\mu n^\nu \ .$$

and using the Ricci identity, we express the last term as,

$$\begin{aligned} n^\nu R_{\mu\nu} n^\mu &= n^\nu R^\lambda{}_{\mu\lambda\nu} n^\mu = n^\nu \left[\nabla_\lambda, \nabla_\nu\right] n^\lambda \\ &= (\nabla\cdot n)^2 - (\nabla_\lambda n^\nu)(\nabla_\nu n^\lambda) + \nabla_\lambda(n\cdot\nabla n^\lambda) - \nabla_\nu(n^\nu \nabla\cdot n) \ . \end{aligned} \tag{11.13}$$

Using the definition of the extrinsic curvature, $K_{\mu\nu} := h_\mu{}^\alpha h_\nu{}^\beta \nabla_\alpha n_\beta$ we see that the first two terms are just $K^2 - K_{\mu\nu}K^{\mu\nu}$. The last two terms are divergences which we suppress for the moment.

Lastly, by considering $\delta \ln \det g_{\mu\nu} = g^{\mu\nu}\delta g_{\mu\nu}$ with $\delta g_{\mu\nu}$ induced by $\delta N, \delta N^i$, $\delta\bar{g}_{ij}$, one can see that $\sqrt{|g|} = N\sqrt{\bar{g}},\ \bar{g} := \det\,\bar{g}_{ij}$. Hence, up to divergence terms, the action becomes,

$$S[\bar{g}, N, N^i] = \int_M dt d^3x\ N\sqrt{\bar{g}}\left(\bar{R} - K^2 + K_{\mu\nu}K^{\mu\nu}\right)$$

To make the $\bar{g}, N^i, N$ dependence explicit, we use the explicit expressions for the normal and the projection operators and note,

$$\begin{aligned} K_{\mu\nu} &= h_\mu{}^\alpha h_\nu{}^\beta(\partial_\alpha n_\beta - \Gamma^\gamma{}_{\alpha\beta}n_\gamma) \ = \ h_\mu{}^k h_\nu{}^l(\partial_k n_l + N\Gamma^t{}_{kl}) = \ N h_\mu{}^k h_\nu{}^l \Gamma^t{}_{kl}, \\ \Gamma^t{}_{kl} &= \frac{1}{2N^2}\left(\partial_t \bar{g}_{kl} - \bar{\nabla}_k N_l - \bar{\nabla}_l N_k\right) \qquad \Rightarrow \qquad (11.14) \\ K_{tt} &= N N^k N^l \Gamma^t{}_{kl}\ \ ,\ \ K_{tj} = N N^k \Gamma^t{}_{kj}\ \ ,\ \ K_{ij} = N\Gamma^t{}_{ij} \qquad (11.15) \end{aligned}$$

In the above, we have use $h_\mu{}^t = 0,\ n_l = 0$, and $\bar{\nabla}$ is explicitly defined using the Riemann–Christoffel connection of $\bar{g}_{ij}$. It is clear from these expressions that only time derivatives of $\bar{g}_{ij}$ occur and that to only through the extrinsic curvature. These then are the generalized coordinates while the lapse and shift must be Lagrange multipliers if they occur *linearly* in the action. It follows that,

$$\begin{aligned} \pi^{ij} &:= \frac{\delta\mathcal{L}}{\delta\dot{\bar{g}}_{ij}} \ = \ \sqrt{\bar{g}}\left(K^{ij} - (\bar{g}^{kl}K_{kl})\bar{g}^{ij}\right) \\ &\quad \longleftrightarrow \ \sqrt{\bar{g}}K^{ij} = \pi^{ij} - \left(\frac{\pi^{kl}\bar{g}_{kl}}{2}\right)\bar{g}^{ij} \\ \dot{\bar{g}}_{ij} &= 2N\ K_{ij} + \bar{\nabla}_i N_j + \bar{\nabla}_j N_i \\ &= \frac{2N}{\sqrt{\bar{g}}}\left(\pi_{ij} - \frac{\pi}{2}\bar{g}_{ij}\right) + \bar{\nabla}_i N_j + \bar{\nabla}_j N_i,\ \ \pi := \pi^{ij}\bar{g}_{ij} \end{aligned} \tag{11.16}$$

Using these the canonical Hamiltonian density, $\mathcal{H} := \pi^{ij}\dot{\bar{g}}_{ij} - \mathcal{L}$ becomes,

$$\mathcal{H} \ = \ \sqrt{\bar{g}}\left[N\left\{-\bar{R} + \frac{\pi^{ij}\pi_{ij} - \frac{\pi^2}{2}}{\bar{g}}\right\} + N_i\left\{-2\bar{\nabla}_j(\frac{\pi^{ij}}{\sqrt{\bar{g}}}\right\}\right], \tag{11.17}$$

where we have suppressed the total derivative term, $2\sqrt{\bar{g}}\bar{\nabla}_i(\pi^{ij}N_j/\sqrt{\bar{g}})$. As expected, the lapse and shift appear as Lagrange multipliers whose equations of motion - the coefficients - give the *primary constraints* in Dirac's terminology and the Hamiltonian density is a linear combination of the constraints. The Hamilton's equations of motion are of course equivalent to the Euler-Lagrange equations of motion i.e. to the Einstein equation. For completeness we note the equation for $\dot{\pi}^{ij}$ (boundary terms are ignored) [17]:

$$\begin{aligned} \dot{\pi}^{ij} \; = \; & -N\sqrt{\bar{g}}\left\{\left(\bar{R}^{ij} - \frac{1}{2}\bar{R}\bar{g}^{ij}\right) - \frac{\bar{g}^{ij}}{2\bar{g}}\left(\pi_{kl}\pi^{kl} - \frac{1}{2}\pi^2\right)\right. \\ & \left. + \frac{2}{\bar{g}}\left(\pi^{ik}\pi_k{}^j - \frac{1}{2}\pi\pi^{ij}\right)\right\} + \sqrt{\bar{g}}\left(\bar{\nabla}^i\bar{\nabla}^j - \bar{g}^{ij}\bar{\nabla}^k\bar{\nabla}_k\right)N \\ & + \sqrt{\bar{g}}\bar{\nabla}_k\left(\frac{N^k\pi^{ij}}{\sqrt{\bar{g}}}\right) - \left(\pi^{ik}\bar{\nabla}_k N^j - \pi^{jk}\bar{\nabla}_k N^i\right) \end{aligned} \tag{11.18}$$

The coefficient of the Lapse N is called the *Scalar (or Hamiltonian) constraint* while the coefficient of N_i is called the *vector (or the diffeomorphism) constraint*. The matter-free gravity is thus a Hamiltonian system with phase space coordinatised by a 3-metric (Euclidean signature), g_{ij} and a symmetric tensor field, K_{ij} defined on a three manifold Σ satisfying the scalar and the vector constraints which are *first class constraints* in Dirac's terminology [90]. The space-time description of the initial value formulation has been cast in a phase space formulation, potentially ready for a passage to canonical quantization. This is known as the *Arnowitt-Deser-Misner* (ADM) formulation [91]. The hall mark of general relativity, the space-time covariance, has apparently disappeared. It is not so, the space-time covariance is encoded in the *algebra of constraints*, known as the *Dirac Algebra*:

$$H(N) \; := \; \int_\Sigma d^3x \; N\left(-\sqrt{\bar{g}}\bar{R} + \frac{\pi^{ij}\pi_{ij} - \pi^2/2}{\sqrt{\bar{g}}}\right) \tag{11.19}$$

$$H(\vec{N}) \; := \; \int_\Sigma d^3x \; N_i\left(-2\bar{\nabla}_j\pi^{ij}\right) \tag{11.20}$$

$$\{H(\vec{N}), H(\vec{M})\} \; = \; -H(\vec{L}) \; , \; L^i \; := \; N^j\bar{\nabla}_j M^i - M^j\bar{\nabla}_j N^i \tag{11.21}$$

$$\{H(\vec{N}), H(M)\} \; = \; -H(K) \; , \; K \; := \; N^i \cdot \bar{\nabla}_i M \tag{11.22}$$

$$\{H(N), H(M)\} \; = \; H(\vec{K}) \; , \; K^i \; := \; \bar{g}^{ij}(N\partial_j M - M\partial_j N) \tag{11.23}$$

The detailed demonstration of these facts may be seen in [92, 93].

Suffice it to say that not only Einstein equation admit a dynamical view of space-time as an evolving 3-geometry, this dynamics is a *Hamiltonian* dynamics making general relativity amenable to canonical quantization.

We return to the surface terms now. If we just want to get the equation of motion, then the surface terms could be ignored as they do not affect local equations of motion. However, the idea of a variational principle is to vary over 'all possible fields in a neighborhood of a path'. For this we have to specify what 'all possible' means i.e. specify the space of fields over which the variation is to be considered. The space of fields for the action formulation are suitably smooth *space-time fields* subject to their specification on the space-time boundary. So to begin with the action should be a well-defined function on such a space and it should be stationary with respect to all infinitesimal variations in the vicinity of a potential solution. To ensure that 'all partial derivatives' vanish at the extremum, the variation of the action should depend only on $\delta g^{\mu\nu}$ and all other dependences should cancel, if necessary, by addition of further terms.

Consider the surface terms in the variation of the action under the stipulation $\delta g^{\mu\nu} = 0$ on the boundary ∂M.

The surface term in the Einstein–Hilbert action is explicitly given in (11.10). Taking the boundary of the four-dimensional region on which the action is defined, to be made up of space-like or time-like hypersurfaces, it is easy to see that $J_\lambda n^\lambda = (n^\lambda h^{\mu\nu} - n^\mu h^{\nu\lambda})\nabla_\mu(\delta g_{\nu\lambda})$, where n^λ is the normal to the boundary hypersurface and $h^\mu_{\ \nu} = \delta^\mu_{\ \nu} \pm n^\mu n_\nu$ is the corresponding projection operator. The $\pm$ relates to the space-like/time-like segments. For the variation $\delta g_{\mu\nu} = 0$ on the boundary, $h^{\nu\lambda}\nabla_\lambda \delta g_{\alpha\beta} = 0$ as well on the hypersurface and the surviving term is $J \cdot n = -h^{\alpha\beta} n \cdot \nabla \delta g_{\alpha\beta}$. This is nothing but $-2\delta K^\mu_{\ \mu}$ or the *variation* of the trace of the extrinsic curvature of the hypersurface.

It follows that $S'[g] := S[g] + 2\int_{\partial M} K$, under the variation $\delta g^{\mu\nu}$ vanishing at the boundary, has no boundary contributions and its variation vanishes iff the metric satisfies the Einstein equation.

The Hamiltonian formulation uses the 3 + 1 decomposition and proceeds to identify a *phase space*. The identification of canonical variables is sensitive to the total divergence terms in the action and can lead to quite different canonical formulations[2] giving the same classical equations of motion. Once the symplectic structure (canonical variables) is identified, the *phase space* is defined in terms of appropriately smooth fields on the 3-manifold together with appropriate stipulation of boundary condition. In order to define the Hamilton's equations of motion, the variation of the Hamiltonian $H := \int_\Sigma \mathcal{H}$, over paths in the phase space, should not contain any other contributions from the boundary of the 3-manifold.

For the ADM Hamiltonian, the cases of interests are (a) the 3-manifold Σ being compact without boundary and (b) it being asymptotically flat at spatial infinity. In the former case, there is no boundary while in the latter case an extra term needs to be added and corresponds to the ADM energy of the space-time. For the details, we refer to [17].

The real utility and significance of action formulation(s) is really at the quantum level whether in the path integral approach or the canonical approach. Our focus being the classical level, we conclude this section with these brief remarks.

[2]This is especially so in the tetrad formulation allowing for non-zero torsion. The connection formulation discovered by Ashtekar [94] by a canonical transformation on the ADM phase space, can be obtained from addition of such terms [95].

Chapter 12

Numerical Relativity

Einstein equation is a complicated system of partial differential equations and even though it can be cast as an initial value problem, obtaining solutions for *physically realistic* situations is a very tall order and *Numerical Simulations* are crucially needed. Perhaps the strongest need is felt for extracting wave forms of gravitational waves from compact sources in a variety of possible motions. The numerical simulations however have also unexpectedly revealed the *critical phenomena* in collapse situation and are a tool to explore issues such as cosmic censorship. In this chapter we will describe the basic ingredients of numerical relativity and highlight some of the recent developments. The primary reference is the beautiful review by Luis Lehner [96].

Problems of interest: In the weak field regimes - metric close to Minkowski metric or observation length scales are large compared to the Schwarzschild radius of compact source - systematic perturbative analytical methods such as post-Newtonian expansion exists which are quite reliable. In these regimes too numerical simulations are used e.g. N-body simulations for galaxies, but these are not general relativistic simulations. It is the strong field regimes where numerical methods play a crucial role. This class of problems involve compact stars such neutron stars or black holes, individually or in binaries, formation of horizons (or otherwise) in a gravitational collapse, oscillatory or otherwise approach to cosmological singularities (the Belinskii-Kalatnikov-Lifshitz conjecture) etc. Some of these are primarily of theoretical interest such as exploration of critical collapse, cosmic censorship, BKL singulatities or the general two body problem. Some, however, have practical applications in the astrophysical context, especially the two body problem is strongly motivated by the efforts towards direct detection of gravitational waves. Following [96] we will first describe the numerical approach in somewhat general terms.

The basic problem: The basic problem is of course to solve the Einstein equation numerically. The covariant form of the equation implies that a *physical solution* can be obtained as many different metric coefficients as functions of corresponding (local) coordinates. In arbitrary coordinates, the equation is a local, partial differential equation with no particular type - hyperbolic or parabolic or elliptic. Whichever way a solution, $g_{\mu\nu}(x)$, is obtained, it is a local solution and one attempts to extend it in some 'maximal way'. Already in the discussion of causality and determinism, we noted that not every solution of the equation is physically admissible and the globally hyperbolic solutions are the physically relevant ones. These space-times already have a $\mathbb{R} \times \Sigma_3$ topology and can be viewed (sliced) as evolution of spatial hyper-surfaces in the space-time. In the previous chapter we saw that performing a 'space + time' decomposition and further imposition of coordinate conditions, it is possible to split the equations into elliptic (constraint) and hyperbolic ones for which local existence, uniqueness and well-posedness properties hold. This means that we input (i) a 3-manifold Σ, (ii) two symmetric tensor fields of rank 2 on it namely g_{ij} and K_{ij} one of which is a Riemannian 3-metric, (iii) a lapse function N, (iv) a shift vector N^i and construct a space-time from the evolution equations satisfied by the two tensor fields. There are other methods of viewing the Einstein equation as an evolution of some data specified on 3-manifolds which will be null hypersurfaces in the evolved space-time e.g. the *characteristic value formulation.* We will focus on the Cauchy framework and refer the reader to the references for other approaches [96, 97].

In the previous section, we have already given the 3+1 decomposition as well as obtained the extrinsic curvature K_{ij} in terms of the $\partial_t \bar{g}_{ij}$ (eqn. 11.15). We also had the Hamilton's form of evolution equations. In the numerical approach, it is more customary to present the evolution equations in terms of $\mathcal{L}_n$, $n^\mu = N^{-1}(t^\mu - N^\mu)$ *or* $d_t := \partial_t - \mathcal{L}_{N^\mu} = N\mathcal{L}_n$. Thus the equations are presented as[1],

$$d_t g_{ij} = 2NK_{ij} \tag{12.1}$$

$$d_t K_{ij} = -N(R_{ij} - 2K_i{}^l K_{lj} + KK_{ij}) + \nabla_i \nabla_j N \tag{12.2}$$

These are to be supplied with initial values of g_{ij}, K_{ij} satisfying the constraints (which are satisfied there after),

$$R + K^2 - K^{ij}K_{ij} = 0 \quad , \quad \nabla_j(K^{ij} - Kg^{ij}) = 0 \, . \tag{12.3}$$

and the lapse and shift, which are arbitrary. Prescribing a lapse and shift in some manner corresponds to specification of coordinates. Notice that the lapse appears explicitly in the evolution equation (12.1) while the shift appears when relating the d_t evolution to $\mathcal{L}_t(= \partial_t)$ evolution. Hence their choices have to be made judiciously for a *stable* numerical evolution.

For instance, an obvious choice of lapse and shift would be, $N = 1, N^i = 0$ which corresponds to a Gaussian coordinate system (also known as *geodesic slicing*). A draw back of these coordinates is that geodesics emanating normally from the spatial slice tend to develop caustics thereby limiting the evolution.

Another potential danger, for simulations involving possibility of black hole formation, is developing of singularities. These can be avoided if coordinates can be so chosen as to approach the singularity 'slowly' by making the lapse approach zero in the vicinity. One such choice of slicing is the so called *maximal*

[1] In this section, all quantities are three-dimensional tensors and all over-bars are removed. More standard notation in numerical relativity community is: $N \to \alpha, N^i \to \beta^i, g_{ij} \to \gamma_{ij}$ and $K_{ij} \to -K_{ij}$

slicing i.e. slices on which the local volume is maximized. This is defined by $K = 0$. From the trace of the second evolution equation, it follows that $\nabla^2 N = NK^{ij}K_{ij} + d_t K$ while the trace of the first evolution equation implies $d_t \ln g = 2NK$. Thus for maximal slicing local volume is constant (i.e. $\partial_t g = 0$), if shift is chosen to be zero. This shows the singularity avoidance property. However, it distorts the slicing so much that the spatial gradients start getting larger thereby crashing the simulation. This is an example where the lapse is specified by its own equation (elliptic). Yet another type of specification is imposing conditions on the coordinates directly e.g. the *Harmonic coordinate condition*: $g^{\mu\nu}\nabla_\mu\nabla_\nu x^\lambda = 0$. This was used in the proof of the theorem (11.2). This has problems of developing so called 'coordinate shocks' which are then sought to be alleviated by using a *generalized harmonic coordinate condition*, replacing the right-hand side by a source function H^λ which is further to be supplied with its own evolution equation. There are many variants of these and other slicings too [96].

The coordinate conditions, is just one of the issues to be faced. There are two other main issues. The choice of initial conditions together with treatment of constraints and the treatment of boundary conditions.

Usually, the evolution is taken to be 'free' evolution i.e. the initial data is chosen to satisfy the constraints and constraints are not checked at each step. If the evolved solution fails to satisfy the constraints to within truncation errors, then numerically it would be acceptable to ignore constraints (however the violations could grow). This would be convenient since solving constraint is computationally intensive. If the constraint violation is larger, then simulation could crash. In the context of evolution of black hole binaries, there seem to be two methods of handling this: adopt *generalized harmonic coordinates with constraint damping* or use suitable variants of the so-called *BSSN* evolution scheme together with appropriate gauge choices and treatment of black hole singularities [98, 99].

Choosing initial data itself has two main issues to be faced. The constraint equations do not tell us which components of the metric and the extrinsic curvature or combinations thereof could be specified freely and the remaining ones solved for. The second issue is how to choose solutions of the constraints which capture the physical situation. For example, in the two body problem case, apart from putting in the initial orbit parameters and spins (which itself is non-trivial), one also needs to include the existing gravitational radiation which can only be guessed or tried out with different amounts. The first problem is addressed by extracting a conformal factor from the metric and using a decomposition of the K_{ij} in terms of its transverse, traceless part and a longitudinal part. The most common decomposition is the *York–Lichnerowicz conformal decompositions* [100]. In this decomposition, one chooses *freely* a 'conformal metric' $\tilde{\gamma}_{ij}$ with determinant equal to 1 (say), a traceless symmetric tensor $\tilde{M}_{ij}$ and a scalar K. The desired initial data variables g_{ij}, K_{ij} are defined in terms of these as

$$g_{ij} := \psi^4\tilde{\gamma}_{ij} \quad , \quad K^{ij} := \psi^{-10}\left\{(\tilde{\mathbb{L}}V)^{ij} + \tilde{M}^{ij}\right\} + \frac{1}{3}\psi^{-4}\tilde{\gamma}^{ij}K \qquad (12.4)$$

$$\text{where,} \qquad (\tilde{\mathbb{L}}V)^{ij} := \tilde{\nabla}^i V^j + \tilde{\nabla}^j V^i - \frac{2}{3}\tilde{\gamma}^{ij}\tilde{\nabla}\cdot V$$

The constraint equations now become equations for the function ψ and the vector V^i constituting the longitudinal part of K^{ij}. With such a decomposition, the 4 quantities, ψ, V^i, satisfy standard form of elliptic equations which can be solved for choosing suitable boundary conditions[2]. The free data of course has to be chosen to reflect the physical situation.

This brings us to the issue of boundary conditions. A simulation involving compact bodies has *outer boundary* which is typically asymptotically flat. If the compact body involves a black hole, then there is an *inner boundary* as well which is expected to hide the singularity, assuming censorship holds. Numerical domain cannot be infinite and this is a tricky issue. One natural method would be to use the conformal compactification and specify the boundary conditions consistent with asymptotic flatness at spatial infinity. This however suffers from loss of resolution and 'piling up of gravitational ripples' and usually drives the simulation into instability. Another naive approach would be to make a 'box' around the localized sources by introducing a time-like boundary. However what the appropriate boundary conditions are is not known, all one knows are the fall off conditions as discussed in chapter 7. One way in which the issue is avoided, takes the outer boundary 'far enough' from the localized region containing the 'source' so that the wave form extraction can be achieved within some radius of 'far zone' and region exterior is evolved accepting the loss of resolution. Another strategy is to match the solution with another solution obtained in the characteristic value problem (data on null hypersurface(s)). This is analogous to the method of extending a solution and is called the *Cauchy Characteristic Matching*. Suffice it to say that there is no single, clear-cut and computationally viable strategy.

Inner boundaries arise potentially in the context involving black holes which contain singularities and horizons. One way to avoid emergence of divergences due to singularities is to exclude from the grid a portion where singularities could arise. Natural questions are how to identify the excision region and what boundary conditions are to be provided at the excision boundary. Obvious place to introduce an inner boundary is somewhere on or inside an apparent horizon (since event horizon cannot be known in advance). From the analytical studies of black holes we know that once inside the event horizon, no null rays can come out. So there would always exist inner boundaries where *all* null rays of the solution would be directed inwards i.e. out-going from the computational domain. Since no 'cause' inside the excised region can have any effect in the computational domain, no boundary conditions need be specified!

---

[2]The transverse parts, $\tilde{Q}^{ij}$ can be defined either with respect to g_{ij} or the $\tilde{\gamma}_{ij}$ i.e. $\nabla_j\tilde{Q}^{ij} = 0$ or $\tilde{\nabla}_j\tilde{Q}^{ij} = 0$ [100].

It remains to locate apparent horizon by detecting outer, marginally trapped surfaces defined in the definition (6.17). Tracking apparent horizon however is computationally expensive.

Another way of dealing with evolution in presence of singularities is to replace a black hole singularity by a *puncture*. Roughly speaking, a black hole's interior is replaced by a transformed copy of its exterior with the spatial infinity represented by a 'point' or a *puncture*. For instance, consider the Schwarzschild worm hole which has two copies of exterior Schwarzschild space-time and joined at the $r = 2m$ sphere. A coordinate transformation, from the standard areal radial coordinate $r \to \rho, r(\rho) := \frac{(2\rho+m)^2}{4\rho}$ allows the two copies to be covered by $\rho \in (\infty, m/2]$ and $\rho \in [m/2, 0)$. In the (ρ, θ, ϕ) coordinate system, the $\rho = 0$ corresponds to the spatial infinity of one of the copies and the full range correspond to the exterior region which is what is relevant in the evolution. The behaviour of the metric at spatial infinity is irregular but understood. Thus, for numerical evolution, a black hole is represented not by any apparent horizon, but by a puncture (spatial infinity) in the initial slice. How the slice is evolved in the vicinity of the puncture(s) i.e. choice of lapse and shift and their evolution controls the stability of the evolution. If the shift is made to vanish at a puncture, it is a 'fixed' puncture while allowing shift to be non-zero gives a 'moving puncture' evolution. This is one of the ingredients that made the breakthrough in evolving binary systems of black holes in 2005 [101].

We have so far assumed that the system of evolution equation (and variables) is the one that followed the standard ADM form. There are alternative schemes of evolution that have been successful. One such is the *BSSN (Baumgarte-Shapiro-Shibata-Nakamura)* scheme [102, 103]. This is similar to the York–Lichnerowicz conformal decomposition mentioned above. Its basic variables are the conformal metric $\tilde{\gamma}$ with determinant 1, the trace of the extrinsic curvature, K, the trace-free part of the conformally scaled extrinsic curvature, $\tilde{A}_{ij} := \psi^{-4}(K_{ij} - \frac{1}{3}g_{ij}K)$ *and* the connection variable $\tilde{\Gamma}^i := \tilde{\Gamma}^i{}_{jk}\tilde{\gamma}^{jk} = -\,\partial_j\tilde{\gamma}^{ij}$. The evolution equations for $\psi, \tilde{\gamma}, \tilde{\Gamma}^i$ follow from their definitions while those for $K, \tilde{A}_{ij}$ come from the Einstein equation. The lapse and shift remain freely prescribable. In actual implementations constraints are used to eliminate certain terms to simplify the equations incorporating constraints partially at least. The treatment of $\tilde{\Gamma}^i$ as an independent variables allows the system to be cast as a hyperbolic system with suitable choices of gauge conditions [99].

Generalized Harmonic condition with constraint damping is another evolution scheme which has been successful too. We have already noted that the drawbacks of the Harmonic coordinate condition are alleviated by introducing the source function, $H^\lambda := g^{\mu\nu}\nabla_\mu\nabla_\nu x^\lambda$. The usual scalar and vector constraints can be expressed in terms of $C_\mu := g_{\mu\nu}(H^\nu - \Box x^\nu)$. Hence vanishing of C_μ which is same as imposing the generalized harmonic condition, implies vanishing of the usual constraints. The evolution scheme now expresses the

Einstein equation in terms of the usual variables and a multiple of the constraints C_μ is added which gives it the adjective of 'constraint damping'. An evolution equation for the source function must be specified to complete the equation system [99, 104].

The break through in achieving a long term stable evolution of black hole binaries is based on the BSSN evolution scheme with moving punctures (shift is allowed to be non-zero but the lapse is arranged to vanish near the punctures). Prior to 2005, the evolution used to crash before completing barely one orbit. After the break through however, over ten orbits are possible which include merger phase and an overlap with in-spiral.

Further comments and details may be seen in [98, 99].

This concludes our brief sketch of some of the basic issues that arise in a numerically solving the Einstein equation. There are many technical variations and clever implementations which are too specialized for this book and the reader is directed to the references.

Chapter 13

Into the Quantum Realm

From the Newtonian conception of gravity as a universal and instantaneously acting force to the Einsteinian one wherein the Newtonian gravity is seen as a manifestation of the curvature of the space-time involved a revision of the space and time to a changeable, merged entity called a space-time. It extended the scope of gravitational phenomena from solar system scale to cosmological scale, triggered further instabilities in the stellar equilibria to suggest formation of the black holes, accommodated an expanding universe providing a mechanism for formation of structures at various scales and gave an independent dynamical status (degrees of freedom) to the space-time geometry via the brand new prediction of gravitational waves. This transition came about by the challenge thrown by special relativity at the Newtonian gravity with its accidental equality of inertial and gravitational masses. The relativistic gravity faces its own new challenges. What are these challenges?

General relativity predicts the black hole solutions - space-times with horizons containing trapped surfaces. Within the classical framework, the singularity theorems (6.29) imply existence of a singularity inside a black hole horizon implying breakdown of general relativistic model of a space-time. Likewise, an everywhere expanding universe too imply a breakdown of the relativistic space-time, *without necessarily implying infinite gravity (tidal forces).* Thus, the singularity theorems challenge the relativistic model of the space-time arena, *without any input from the quantum.*

The same solutions with horizons also obey certain laws of their behaviour with a curious formal similarity with the laws of thermodynamics. This could have remained an intriguing curiosity, but Hawking showed that these horizons have the ability to accentuate the *quantum fluctuations* enough to make them radiate like a hot body. This suggests a mechanism for a possibly *complete evaporation* of the black hole or possibly leaving behind a *remnant.* The distinction between evaporation of the by-now-proverbial piece of coal and that of a black hole horizon is that the coal does not hide its contents at any stage while a black hole horizon does all through the evaporation process. This entails two possibilities: (a) a black hole may be formed from the collapse of a pure quantum state and at the end of evaporation, one has only a thermal state - such an evolution cannot be unitary[1]; (b) one may send in a pure state into a black hole and wait for its complete evaporation to retrieve it. But this

[1] A small size remnant is not large enough to retain the information.

is not possible, leading once again to the loss of 'information'. The former is a process including formation of a black hole with its subsequent evaporation and it is not unitary while the latter implies non-unitarity of a scattering process in the presence of an evaporating black hole.

Usual quantum theory does not allow for such non-unitary processes while classical general relativity, by admitting black holes, implies existence of such processes. This is a potential conflict between classical general relativity and the quantum theory.

There is a third and more elementary challenge from the quantum theory. Recall the determination of the space-time metric from section (4.4). It is premised on experimental determination of local light-cones. The wave nature of any particle-like probe will render such a determination a little 'fuzzy'. Thus quantum theory will directly interfere with the operational meaning of any space-time metric at *sufficiently small length/time scales.*

There are different interpretations of these challenges as well as different suggestions of meeting them. Broadly, one interpretation can be described as 'gravity is an emergent phenomenon' while the second one interprets it as an indication that 'gravity needs a quantum extension'. We briefly comment on these alternatives below.

13.1 Gravity Is "Emergent"

This is a view which is a mixture of many ideas with a common thread that gravitational interaction need not be a 'fundamental' interaction and by implication, need not be 'quantized' at all. The strongest hint of this is inferred from the black hole thermodynamics[2].

Once a black hole has formed and settled in, all memory of its formation process is lost except for the mass, angular momentum and charge. Furthermore the membrane paradigm establishes a detailed analogy of their horizons with surfaces endowed with certain mechanical and electromagnetic properties. These two together contained a hint that space-time may be some kind of a fluid with Einstein equation being the hydro-dynamical equation [107]. This idea got a further boost in the work of Jacobson [108] which argued that

[2] Historically, Sakharov was the first to suggest that *dynamics* of space-time metric may be determined by quantum fluctuations of matter on a Lorentzian manifold [105, 106]. The motivation was from the astrophysical possibility of non-zero cosmological constant while the rationale was from the 1-loop effective action for a background geometry from quantized matter on that background. This is completely independent of any black hole thermodynamics and implicitly assumes a background to be not too far from Minkowski space-time. General covariance and a derivative expansion is sufficient to generate both the cosmological constant and the Einstein–Hilbert action as leading terms. Away from such weak fields we would not even know how to quantize matter fields on an arbitrary Lorentzian manifold.

it is possible to interpret Einstein equation as an equation of state from a thermo-dynamical system identified by a *suitable causal horizon* whose existence in a neighborhood of every space-time point is assured. The rationale is based on the inputs that (a) a causal horizon makes some of the microscopic degrees of freedoms (molecules of a thermo-dynamical system) unobservable and energy flux across it can be thought of as 'heat'; (b) the *quantum* entanglement across a horizon can account for the entropy of the system 'behind' the horizon and (c) the quantum fluctuations are seen as thermal fluctuations by a local Rindler observer hence a temperature can be taken to be the local Unruh temperature. However, not every causal horizon can give a thermodynamic system in equilibrium. Its stipulation is done locally in an inertial frame at a space-time point p such that the generators of the causal horizon have zero expansion and shear at the point p. This allows the area change to be linked with the heat flow to first order within the neighborhood. *Assuming* that the entropy is proportional to the area of the cross-section of the horizon, the equation of state of such a system is precisely the Einstein equation, including the possibility of a cosmological constant. The *solution space-times* thus correspond to *equilibrium states* of the unknown micro-constituents of the system. From this perspective, the gravitational waves can be thought of as sound waves in a gas, propagated by the large number of 'collisions' of the micro-constituents, while the singular solutions simply indicate deviation from the equilibria. This is a persuasive argument with a sense of generality since it is applied near every point and is free from black hole horizons.

The thermodynamic interpretation of gravity is pushed further by Padmanabhan along with his collaborators [109]. They argue that just as there exist local inertial observers so do there exist local Rindler observers and thanks to the local Unruh temperature, in these observer's view, space-time is a thermodynamical entity. Padmanabhan advances the criteria that if such a view of a Rindler observer is to be viable, then there should be a variational principle (not just equation of state) which has a direct thermodynamical analogy. Furthermore, the analogy should continue to hold for other possible field equations which require different 'actions' and hopefully lead to some new insights. A variational principle based on null hypersurfaces (in a given background Lorentzian geometry) is proposed and the viability criteria are checked, at least for Lovelock type actions which include the Einstein–Hilbert action. In particular, such a variational principle leads to an equation of the form $(G_{\mu\nu} - T_{\mu\nu})n^\mu n^\nu = 0 \; \forall \; n \cdot n = 0$. It is noted that such an equation is invariant under addition of a constant to the Lagrangian density since it adds a cosmological constant term which is killed by the null vector norm. This is useful since it removes the ambiguity in the value of the cosmological constant by shifts in the matter actions. This approach emphasizes that using (suitably defined) surface and bulk degrees of freedom in a region, Einstein's equations *can be* written in a thermodynamic language instead of in a purely geometric language, [110].

The main argument here seems to be that since the Einstein equation can be obtained *without treating the metric as a fundamental dynamical variable*, gravity described by the Einstein equation is an emergent phenomenon[3].

There are other ideas centered around the black horizons. For instance, Laughlin regards black hole horizons (but not any other causal horizons?) as a place where the classical general relativity breaks down in analogy with critical surfaces in quantum phase transitions, again suggesting that gravity is not fundamental [112].

All these approaches, presume the framework of Lorentzian manifold and what is taken as 'emerging' are the space-times *satisfying the Einstein equation.* An implication is usually made that *it is inappropriate to quantize the* metric *which does not represent underlying microscopic degrees of freedom.* These views while acknowledging the existence of 'different' microscopic degrees of freedom, do not contain any clues as to what those might be or how they may be 'discovered'. Although local Rindler observers exist within the classical Lorentzian space-times, a quantum input is brought in via the Unruh temperature.

There has also been a proposal wherein not just the dynamics, but kinematics of general relativity - smooth Lorentzian manifold - is also envisaged to be emergent. The *Causal Set Theory* approach [113–115] falls into this category which recovers continuum geometry from a statistical sprinkling of finite set of points with built-in causal relations as a partial order.

13.2 The Quantum Gravity Paradigm

Independent of the information loss issue, the existence of black hole entropy itself is taken as a clue to the existence of micro-structure to classical horizon and in the light of classical black holes being pure geometry objects, the presumed micro-structure is taken to be 'atoms of geometry'. The quantum geometry of LQG provides enough structure to explain the black hole entropy in this manner.

String theory explains the entropy in a different manner. It posits that microscopically, the (extremal) black holes are actually collections of D-branes whose states can be counted. Changing the string coupling then brings us to a regime wherein these states can be seen as states of conventional geometrical black holes. The supersymmetry plays a crucial role by preserving the state counting through the process of changing the coupling.

Together with the breakdown of general relativity indicated by the singularity theorems, these aspects strongly suggest a need for 'quantizing gravity'.

[3] There is also a notion of 'emergent gravity' (and 'emergent gauge theory') in the context of AdS/CFT or gauge-gravity duality, which is within a quantum framework and is of a very different nature. This may be seen in the contribution of Horowitz and Polchinski in [111].

In the next subsection we recall some of the salient points of the two main approaches to quantum gravity - Loop Quantum Gravity and String Theory. For more comprehensive views, please see [111].

13.2.1 String Theory: The Unification Paradigm

Historically, string theory arose out of the s-matrix approach to strong interactions [116–120]. Characteristically it predicted a massless, helicity-2 state in its spectrum while no such hadron is known. This led authors of [121, 122] to suggest that the theory is perhaps more appropriately regarded as being applicable at the Planck scale and thus be interpreted as a theory of gravity. It is primarily a *quantum* theory of a string (or strings) propagating in some embedding space-time i.e. an open or close, two-dimensional world sheet, embedded in some $D+1$-dimensional space-time. In its classical formulation, it is a 1+1 dimensional field theory with fields taking values in a $D+1$-dimensional target manifold. The fields defined on the target manifold serve as 'coupling constants' of the two-dimensional field theory. The theory is invariant under world-sheet diffeomorphisms and Weyl scalings of the world-sheet metric. In its quantum form, preservations of these 'gauge invariances' puts severe constraints on its properties. The most characteristic ones are: (1) a critical dimension - 26 for strings with only bosonic fields and 10 for strings including fermionic fields with supersymmetry; (2) bosonic strings have *tachyons* for both the open and closed strings but more crucially the closed strings have a massless helicity-2 state, graviton, in its spectrum; (3) the fermionic strings display supersymmetric spectra, no tachyons and again a graviton in the closed superstring spectra; (4) the spectra are very rich and tightly controlled (mass and spins being correlated) and so are the the scattering amplitudes; (5) the consistency conditions are so strong that the allowed Yang-Mills groups are limited two just two - $SO(32)$ and $E_8 \times E_8$ [123].

These features lent support to the unification paradigm - a single framework to 'understand' all known (and possibly yet to be known) interactions together with the participating entities, perturbative gravity being automatically included by the demand of consistency of the theory. It also provided a novel way to view space-time fields and their equations of motion as arising from the (super)conformal invariance of suitable world-sheet quantum field theories. The perturbative approach provided an appealing picture of interactions being constrained by the geometrical joining and splitting of strings. The perturbative niceties were soon seen to be inadequate thanks to the non-summability of the string perturbation theory [124] paving the way for non-perturbative sectors of the theory together with various dualities linking 'different' string theories [125, 126].

As far as *gravity* is concerned, the string theory contains the following features. First, perturbative gravity is automatically incorporated. Second, the non-renormalizable ultraviolet divergences that plague the perturbative field theory are supposed to be effectively absent thanks to the modular invariance properties of the amplitudes. Third, the non-perturbative sector contains solitonic excitations such as the Dp-branes (Dirichlet boundary condition along p-directions) [127] which provide a microscopic description of black holes, at least of the extremal variety and lead to their Bekenstein-Hawking entropy. Fourth, there are possible mechanism such as the 'fuzz ball' picture [128] which provides a replacement of the traditional black hole horizon thereby avoiding the 'information loss' conflict. Fifth, the T-duality suggests a possibility that as the cosmological singularity is approached in a particular presentation, the would-be-singular presentation can be related to a non-singular one thereby 'resolving' the cosmological (FLRW) singularity. For a 25-year perspective on string theory, please see [129].

It is perhaps fair to say that lessons from string theory for a modification of Einsteinian gravity are that quantum mechanically gravity should not be viewed in isolation but as packaged with matter. It is also conceivable that the underlying quantum world may accessed via a *collection of perturbative string theories* appropriate for different regimes, connected nevertheless together thanks to the various dualities. String theory claims to be one such (and currently only such) package with the consistency of the framework being conditional on extra space-time dimensions and supersymmetry.

13.2.2 Loop Quantum Gravity: The Background Independence Demand

General covariance is an essential property of relativistic gravity and this is in conflict with any prescribed background. Let us note a few points to appreciate this statement. General covariance refers to form invariance under *general coordinate transformations* whose most explicit articulation is in the framework of differentiable manifolds. The adjective 'general' refers to arbitrary, appropriately C^k and usually C^∞, invertible change of local coordinates. 'Covariance' then is most economically expressed by using tensorial objects as basic variable as well as any local differential equations they may be required to be satisfied (e.g. the Einstein equation). For contrast, the covariance under Lorentz transformations limits the permitted transformation to *linear coordinate transformations* forming the group $O(1,3)$. The physical reason for requiring general covariance has been the postulated equivalence of observers in arbitrary relative motion, regardless of any equation the metrics may satisfy. Now imagine we have some tensor field(s) prescribed ab initio on a manifold on which we wish to study other dynamical tensorial fields, including a Lorentzian metric. The prescribed fields are fixed as a *background* on which we may anchor locations and time stamps. Any such background, immediately limits the set of permitted coordinate transformations to those that leave the background unchanged and we get only a limited covariance. There are *no tensorial fields* which are invariant under arbitrary coordinate transformations and hence general covariance is in conflict with prescription of any background fields. In this sense, general covariance is synonymous with requirement of independence of *any* background fields, manifestly or otherwise. This applies not just to the gravitational (metric) field but also to all matter fields coupled to gravity. Loop Quantum Gravity (LQG) is an approach to constructing a *quantum* theory of gravity maintaining background independence manifest.

As Einstein reasoned in his 'Hole Argument', general covariance is also in conflict with a *deterministic interpretation* of the dynamical equations unless a solution and its transforms under space-time diffeomorphisms are regarded as physically the same. This results in the dynamics being *constrained* - the clearest expression of which is obtained in the canonical form as a system with first class constraints. This again is true in presence of non-gravitational fields as well. In a quantum theory, this 'gauge redundancy' of the tensorial fields must be properly taken into account. Note that classically, this gauge redundancy only masks sometimes, the diffeomorphism equivalence of apparently different solution. Quantum mechanically, an incorrect accounting of this redundancy could miss or over-count equivalence classes of classical solutions jeopardizing the semiclassical correspondence or lead to an inconsistency. Background independence is thus crucially relevant at the quantum level.

Let us follow the canonical form since Loop Quantum Gravity primarily developed in this form (although a covariant form has been developing in terms of the Spin foams [130, 131]). We have a theory with first class constraints. Its *reduced phase space*, space of orbits generated by the constraints, on the constraint submanifold, is not known well enough to construct a quantum theory directly. The alternative is to construct the quantum theory in two (or more) steps. First construct a Hilbert space ignoring the constraints, called the kinematical Hilbert space, and then identify the *physical states* as solutions of the constraint operators defined on it. Typically, the solutions are non-normalizable (or are distributional) and one needs to define a new inner product on this space of solutions [132].

In constructing the kinematical Hilbert space, typically a measure on the configuration space of the system is needed to define an inner product. Since we want to get rid of the gauge redundancies, we want the inner product to be diffeomorphism invariant. We can't use any background field, let alone the metric which is a dynamical variable itself. This is in contrast with the usual case of non-gravitational systems where we *can* choose the usual background metric as there is no requirement of general covariance. Here in lies the main difficulty in quantizing a theory with general covariance.

The space of Euclidean 3-metrics seems too unwieldy to admit any measure let alone a 3-diffeomorphism invariant measure. In the mid-eighties, Ashtekar gave an alternative formulation of general relativity in terms of gauge connections [94] which has since been presented in the so-called 'real SU(2)' formulation [133, 134]. In this alternative formulation, the phase space of gravity is same as that of a Yang-Mills gauge theory with gauge group $SU(2)$. Thus apart from the 3-diffeomorphism redundancies, we also have the usual gauge actions under SU(2). The naive configuration space variables are the $A_a^i(x)$

fields on a 3-manifold with 'i' being the adjoint index of SU(2) and 'a' being the space index, both taking 3 values each. The space of gauge connections is unsuitable because of two reasons: under SU(2) transformations it transforms inhomogeneously and *there is no known diffeomorphism invariant measure on this space*[4]. Lattice gauge theories, employing non-perturbative approach, use *Wilson loops or holonomies* as elementary variables which *do* transform homogeneously [135]. Furthermore, *on the space of holonomies, there is a unique, diffeomorphism invariant measure*! Using holonomies as basic variables together with the unique, *Ashtekar-Lewandowski measure* on the space of functions thereof, constitutes *Loop Quantization* [136, 137]. This gives the kinematical Hilbert space in a *background independent manner* and the first step of Dirac quantization is successfully achieved. The subsequent steps of obtaining a physical Hilbert space together with physical observables have not been satisfactorily concluded. The familiar background space-times are now to be understood as appropriate *semiclassical states* of the quantum theory and these too are yet to be obtained in a satisfactory manner. All these steps - complete Dirac quantization with semiclassical states has been obtained in the physically relevant, toy version for homogeneous, isotropic cosmology with a particular form of resolution of the isotropic Big Bang singularity [138, 139]. One of the implications of loop quantization is that the spectrum of any area operator [140] is *discrete*. This has been used in the context of the isolated horizons to obtain the black hole entropy by state counting [73, 141].

One of the key possibility uncovered by this approach is that the spatial geometry need not be a continuum Riemannian geometry in the sense that the spectra of geometrical operators [140, 142, 143] can be discrete and is also *non-commutative* [144]. At a microscopic level, the classical space-time picture may be completely superseded by some discrete structure, LQG providing a specific one. For further details of LQG, please refer to [145, 146].

We would like to mention that there are other more radical approaches which potentially remove a dependence on the background structures such as a manifold and its topology. These include the *Causal Sets* mentioned above and *Causal Dynamical Triangulation* [147]. The CDT approach in particular hints at the possibility of a changing 'dimensionality' of the space—from 2 at the micro-level to 4 at the macro-level [148].

[4]These issues do not arise in the quantization of usual Yang-Mills theories in Minkowski space-time because one typically works in a perturbation theory which needs a gauge fixing and there is *no* requirement of diffeomorphism invariance.

Chapter 14

Mathematical Background

14.1 Basic Differential Geometry

In this chapter we take the opportunity to introduce the hierarchy of structures leading to the desired Riemannian geometry. The idea is to do this in stages to see what structures enable us to do what. Only basic ideas are discussed. There are several excellent books available [149–151] as well the parts of [17, 18].

14.2 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 a set) of subsets of a given set, and 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 e.g. *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 *empty set.* 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 generalize the familiar calculus, we need to suitably generalize 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.

Convergence: 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,

continuity: 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) $|\vec{x} - \vec{y}| \leq |\vec{x} - \vec{z}| + |\vec{z} - \vec{y}|$ (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 $\epsilon-$neighborhood 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 $\epsilon-$neighborhood 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 generalisation 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 (i) $X, \Phi \in T$, (ii) arbitrary unions of members of T is contained in T and (iii) 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.*

For practice, re-write the definition of limit of a sequence in a topological space.

There are three basic properties of topological spaces, namely: (i) connectedness and local connectedness; (ii) separability and (iii) compactness and local compactness. These may be seen in [152].

<u>Remark:</u>

- 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*, <u>every</u> 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 ϵ-neighborhoods.

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

We can immediately define mappings between two topological spaces: $f : (X,T) \to (X',T')$. 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. $f : X \to Y$ 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 $\mathbb{R}^N$.

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

14.3 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 generalize 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 neighborhood of p, together with a homeomorphism $\phi_\alpha : u_\alpha \to O_\alpha \subset \mathbb{R}^n$ 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 neighborhoods 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 *differentiable manifold* or *manifold* for short.

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 e.g. 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 at $x(p)$ (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 differentiable. Furthermore, f being smooth will be meaningless, unless the coordinate transformation is 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.

[1]Functions which are 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.

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 $\mathbb{R}^n$.

Just as there are different topologies on a given set, there can be several different differential structures on the same topological space e.g. 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 of the same dimension 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 γ 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 $f : M \to \mathbb{R}$,

$$\left.\frac{d}{dt} f\right|_{\gamma} := \lim_{\epsilon \to 0} \frac{f(\gamma(\epsilon)) - f(\gamma(0))}{\epsilon} \tag{14.1}$$

Using a chart, (u_α, ϕ_α), 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,

$$\left.\frac{d}{dt} f\right|_{\gamma} = \lim_{\epsilon \to 0} \frac{f_\alpha(x^i(\gamma(\epsilon))) - f_\alpha(x^i(\gamma(0)))}{\epsilon} \qquad \text{But,}$$

$$x^i(\gamma(\epsilon)) - x^i(\gamma(0)) \approx \epsilon \left.\frac{dx^i}{dt}\right|_{t=0}$$

$$\therefore \left.\frac{d}{dt} f\right|_{\gamma} := \lim_{\epsilon \to 0} \frac{f_\alpha(x^i(\gamma(0)) + \epsilon \frac{dx^i}{dt}) - f_\alpha(x^i(\gamma(0)))}{\epsilon}$$

$$= \lim_{\epsilon \to 0} \frac{\epsilon \frac{dx^i}{dt} \frac{\partial f_\alpha}{\partial x^i}}{\epsilon}$$

$$= \left. \frac{dx^i}{dt} \right|_\gamma \frac{\partial}{\partial x^i} f_\alpha \qquad \forall \;\; f : M \to \mathbb{R} \tag{14.2}$$

The (14.1) 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.(14.2). 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

$$\left\{\frac{dx^i}{dt}\right\} \frac{\partial}{\partial x^i} = \left\{\frac{dx^i}{dt}\right\} \left\{\frac{\partial y^j}{\partial x^i}\right\} \frac{\partial}{\partial y^j} = \left\{\frac{dy^j}{dt}\right\} \frac{\partial}{\partial y^j} \qquad \text{or,}$$

$$X^i \frac{\partial}{\partial x^i} = X^i \left\{\frac{\partial y^j}{\partial x^i}\right\} \frac{\partial}{\partial y^j} = Y^j \frac{\partial}{\partial y^j}$$

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* which is an element 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 section (14.6). 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,

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

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 independent meaning.

Being elements of a vector space, tensors of the same rank at a given point, can be added and multiplied by real (or complex) numbers. 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 on 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 operation called *wedge product* is defined. Its definition is included in the section 14.6.

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.

14.4 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,

$$\begin{aligned}
\frac{\partial A'^{\,i}(y)}{\partial y^j} &= \frac{\partial x^k}{\partial y^j}\frac{\partial}{\partial x^k}\left(\frac{\partial y^i}{\partial x^l}A^l(x)\right) \\
&= \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\frac{\partial A^l}{\partial x^k} + \frac{\partial x^k}{\partial y^j}\frac{\partial^2 y^i}{\partial x^k \partial x^l}A^l \qquad (14.3)
\end{aligned}$$

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.

$$\begin{aligned}
\nabla'_j A'^{\,i} &:= \frac{\partial A'^{\,i}(y)}{\partial y^j} + \Gamma'^i{}_{jk}A'^k \\
&= \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\frac{\partial A^l}{\partial x^k} + \frac{\partial x^k}{\partial y^j}\frac{\partial^2 y^i}{\partial x^k \partial x^l}A^l \;+\; \Gamma'^i{}_{jk}\frac{\partial y^k}{\partial x^l}A^l
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\left(\frac{\partial A^l}{\partial x^k} + \Gamma^l{}_{km}A^m\right) + \\
&\qquad \left[\left(\frac{\partial x^k}{\partial y^j}\frac{\partial^2 y^i}{\partial x^k \partial x^m} \ + \ \Gamma'^i{}_{jk}\frac{\partial y^k}{\partial x^m} - \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\Gamma^l{}_{km}\right)A^m\right] \\
&= \frac{\partial x^k}{\partial y^j}\frac{\partial y^i}{\partial x^l}\nabla_k A^l + 0 \qquad (14.4)
\end{aligned}
$$

Thus we deduce that,

$$
\Gamma'^i{}_{jk}(y(x)) \ := \ \frac{\partial y^i}{\partial x^l}\frac{\partial x^m}{\partial y^j}\frac{\partial x^n}{\partial y^k}\Gamma^l{}_{mn}(x) + \frac{\partial y^i}{\partial x^l}\frac{\partial^2 x^l}{\partial y^j \partial y^k} \qquad (14.5)
$$

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 next section, 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 $y^i(x) := x^i + \frac{1}{2}a^i{}_{jk}x^j x^k + o(x^3)$. This implies that the inverse transformation is $x^i(y) = y^i - \frac{1}{2}a^i{}_{jk}y^j y^k + o(y^3)$. 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.

By exactly analogous reasoning it can be checked that partial derivatives of a scalar is a tensor of rank (0,1) without any affine connection modification while for tensor of rank (0,1), affine connection term is needed. The definition: $\nabla_j B_i := \frac{\partial B_i}{\partial x^j} - \Gamma^k{}_{ji}B_k$ constructs a tensor of rank (0,2).

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,

$$\begin{aligned}
\nabla_l \nabla_k B_j &= \partial_l(\nabla_k B_j) - \Gamma^m_{lk}\nabla_m B_j - \Gamma^m_{lj}\nabla_k B_m \\
&= \left\{\partial_l\partial_k B_j - \Gamma^m_{kj}\partial_l B_m - \Gamma^m_{lk}\partial_m B_j - \Gamma^m_{lj}\partial_k B_m\right\} \\
&\quad + \left\{-\partial_l \Gamma^n{}_{kj} + \Gamma^m_{lj}\Gamma^n{}_{km} + \Gamma^m_{lk}\Gamma^n{}_{mj}\right\} B_n \qquad (14.6) \\
\therefore\ [\nabla_l, \nabla_k] B_j &= \left\{\partial_k \Gamma^i{}_{lj} - \partial_l \Gamma^i{}_{kj} + \Gamma^i{}_{km}\Gamma^m_{lj} - \Gamma^i{}_{lm}\Gamma^m_{kj}\right\} B_i \\
\text{or}\ \ [\nabla_l, \nabla_k] B_j &= -R^i{}_{jlk} B_i \qquad \text{with} \qquad (14.7) \\
R^i{}_{jkl}(\Gamma) &:= \partial_k \Gamma^i{}_{lj} - \partial_l \Gamma^i{}_{kj} + \Gamma^i{}_{km}\Gamma^m_{lj} - \Gamma^i{}_{lm}\Gamma^m_{kj} \qquad (14.8)
\end{aligned}$$

The terms in the first braces, all involving derivatives of the tensor field, and the last term in the second braces in eq. (14.6) are *symmetric in* $k \leftrightarrow l$ and hence drop out in the commutator of the covariant derivatives in the next equation. Equation (14.7) is known as the 'Ricci Identity' and eq. (14.8) defines the *Riemann Curvature tensor.*

Remarks:

- There is an alternative notation to denote partial and covariant derivatives, namely, $\partial_j T \Leftrightarrow T_{\ ,j}$ and $\nabla_j T \Leftrightarrow T_{\ ;j}$.
- It is straightforward to verify that

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

  and the commutator on higher rank tensors goes index-by-index.
- From eq.(14.7), 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. 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 neighborhood.*

  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*, $\sum_{(klm)} \nabla_m R^i{}_{jkl} = 0$. Here $\sum_{(ijk)}$ means sum over cyclic permutations of the indices.

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

- 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.*

Solution of $X \cdot \nabla T_{\|} = 0$ *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, $X \cdot \nabla X^i_{\|} = 0$. In general, $X_{\|} \neq X$. 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.

To summarize: 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].

---

[3]There is a slightly general definition of affine geodesics, namely, that $X_{\|} \propto X$ which implies $X \cdot \nabla X^i = \xi X^i$. However by re-parametrizing 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.*

[4]There are other notions of derivatives producing tensors e.g. 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.

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

14.5 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 diagonalized. 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 $n_+ = n-1, n_- = 1$. Not all manifold admit Lorentzian metrics, the next section 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*. 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,

$$\Gamma^i_{\ jk}(g) \ := \ \frac{1}{2} g^{il} \left(g_{lj,k} + g_{lk,j} - g_{jk,l} \right) . \tag{14.9}$$

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

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 summarized in the next (summary) section.

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 neighborhood 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 neighborhood is controlled by the curvature tensor.

14.6 Summary of Differential Geometry

This is a summary of basic definitions which also serves to state some of the conventions[5]. We consider only real manifolds and the Einstein summation convention is used throughout.

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**.

[5]In the main text, we have $n = 4$ and $Torsion = 0$.

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.

   It is taken to be **oriented** i.e. the Jacobian determinant of the coordinate transformations in all overlapping charts is *positive*.

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, $\left\{\frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^n}\right\}$ 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 $\left\{\frac{\partial}{\partial x^i}\right\}$ is denoted as $\left\{dx^1, \cdots, dx^n\right\}$ 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:

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

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

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

   A **Tensor of rank (r, s) at $p \in M$** 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:

$$\begin{aligned} T &= \sum T^{a_1\cdots a_r}{}_{b_1\cdots b_s} E_{a_1\cdots a_r}{}^{b_1,\cdots,b_s} \qquad \Longleftrightarrow \\ T^{a_1\cdots a_r}{}_{b_1\cdots b_s} &= T(E^{a_1}, \cdots, E^{a_r}, E_{b_1}, \cdots, E_{b_s}) \end{aligned}$$

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:

$$\begin{aligned} (T')^{i_1\cdots i_r}{}_{j_1\cdots j_s}(x') &= \frac{\partial (x')^{i_1}}{\partial x^{m_1}} \cdots \frac{\partial (x')^{i_r}}{\partial x^{m_r}} \frac{\partial x^{n_1}}{\partial (x')^{j_1}} \cdots \frac{\partial x^{n_s}}{\partial (x')^{j_s}} \times \\ & \qquad (T)^{m_1\cdots m_r}{}_{n_1\cdots n_s}(x) \end{aligned}$$

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):

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

In terms of components:

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

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:

$$\begin{aligned} T(\omega_1,\cdots,\omega_{r-1};X_1,\cdots,X_{s-1}) &:= \\ &T(\omega_1,\cdots,E^a,\cdots,\omega_{r-1};X_1,\cdots,E_a,\cdots,X_{s-1}) \quad \Leftrightarrow \\ &\quad T^{a_1,\cdots,a_{r-1}}{}_{b_1,\cdots,b_{s-1}} := T^{a_1,\cdots,c,\cdots,a_{r-1}}{}_{b_1,\cdots,c,\cdots,b_{s-1}} \end{aligned}$$

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, $0 \leq k \leq n$ 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 is a vector space, denoted as Λ^k and has the dimension ${}^nC_k$.

Denote by Λ the *direct sum* of all of 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 $p+q \leq n$. 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 basic **Tensor Algebra** that we 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, $df(X) = X(f) \;\; \forall\, X \in T_p(M)$. 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,

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

its exterior derivative is then defined by,

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

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

(c) $d^2\omega = 0 \quad \forall \;\; \omega \in \Lambda$ (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.

12. **Levi–Civita Symbol $\mathcal{E}_{i_1 \cdots i_n}$:**

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

This allows us to write,

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

For any non-singular matrix $J^i{}_j$, we have the useful relation,

$$(\det J)(J^{-1})_{i_1}^{k_1} \ldots (J^{-1})_{i_m}^{k_m} \mathcal{E}_{k_1 \ldots k_m i_{m+1} \ldots i_n}$$
$$= J_{i_{m+1}}^{k_{m+1}} \ldots J_{i_n}^{k_n} \mathcal{E}_{i_1 \ldots i_m k_{m+1} \ldots k_n} \ \forall\, 0 \le m \le n \ .$$

For $m = 0$, it is just the definition of a determinant.

13. **Volume Form, Orientation, Densities:** Recall that Λ^n is one dimensional. A non-zero n-form $\omega \in \Lambda^n$ at p, 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* Λ^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* $[\omega]$ if $\omega(E_1, \cdots, E_n) > 0$.

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

Volume forms provide an alternative definition of orientability. 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 has been made. This definition of orientability is equivalent to the one given in terms of the sign of the Jacobian determinant of coordinate transformations in the overlapping charts.

For a volume form μ, let its coordinate component $\mu_{i_1 \ldots i_n} := v(x) \mathcal{E}_{i_1 \ldots i_n}$. Under a coordinate transformation $x \to x'$ and the definition $\mu'_{i_1 \ldots i_n} := v'(x') \mathcal{E}_{i_1 \ldots i_n}$, we get that $v'(x') = \det\left(\frac{\partial x}{\partial x'}\right) v(x)$.

It is easy to see that $\tilde{T}_{i_1 \ldots i_m} := \mathcal{E}_{i_1 \ldots i_m i_{m+1} \ldots i_n} T^{i_{m+1} \ldots i_n}$ transforms as a tensor of rank *(0, m)* with *an additional factor* of det $\left(\frac{\partial x}{\partial x'}\right)$. Such quantities are called *tensor densities of weight 1, rank (0, m).*

We could like wise define Levi–Civita symbol with upper indices (notational convenience only) and construct tensor densities of *rank (m, 0) with weight -1.* As an aside, we note that had we allowed orientation reversing transformations as well, then the additional factor could be either the Jacobian determinant or its absolute value. The tensor densities transform by the absolute value of the Jacobian determinant while those which transform by the determinant are called *pseudo-tensors.*

14. **Integration:** Integration on a manifold is a generalization of the Riemann integral of functions of N variables over bounded domains of $\mathbb{R}^N$. The crucial property of these integrals is their invariance under the *change of variables.* On a manifold, using charts, we can import the Riemann integral onto an open set of the manifold. The change of chart induces a change of variables in the local integration and unless the Jacobian determinant of the coordinate transformations is cancelled against an inverse Jacobian coming for the integrand, we cannot have the integral to be chart independent. The only entity on a manifold that has such a property is the component of an n-form as noted above.

    Thus, to define an *oriented integral of an n-form ω, on an oriented n-manifold*, choose an atlas and on each of its charts, define

$$\int_{u_\alpha} \omega := \int_{\phi^\alpha(u_\alpha)} d^n x \; \omega_{1\ldots n}(x) \ ,$$

    where the right-hand side is the Riemann (or Lebesgue) integral over a domain in $\mathbb{R}^n$and the coordinate basis is positively oriented.

    In the overlap of charts, the two expressions will match thanks to the Jacobian determinant being cancelled by its inverse coming from the component of ω. Thus on n-dimensional manifolds *only* n-forms can be integrated meaningfully. These are locally given by,

$$\int_M \omega := \int dx^1 \wedge \cdots \wedge dx^n \; \omega_{1\cdots n} := \sum_i \int_{\phi^i(u_i)} d^n x \; \omega_{1\cdots n} \ .$$

    Technically, in order to make sense of the sum, the manifold is required to be paracompact [17]. (This property is also needed of existence of metric tensor).

    Since the sole component of an n-form can also be viewed as *scalar density of weight 1*, an equivalent statement is that the integrand must be a scalar density one object for its integral to be well defined.

15. **Stoke's Theorem:** This arises in the context of a manifold with boundary. To define a manifold with boundary, we go back to the definition

of a chart and allow the ϕ_α to map u_α into open sets of $\frac{1}{2}\mathbb{R}^n$. The '1/2 $\mathbb{R}^n$' is defined by the restriction $x^1 \leq 0$ (say). Thus $x^1 = 0$ hyperplane is the boundary of $\frac{1}{2}\mathbb{R}^n$ and is itself a manifold of dimension $(n-1)$. The set of points of u_α which are mapped to points in the boundary of $O_\alpha \subset \frac{1}{2}\mathbb{R}^n$, constitute the boundary points of u_α. To define a manifold with boundary, we consider atlases which include charts which have boundary points. Defining compatible atlases etc., we arrive at the definition of a manifold, M, with boundary. The boundary ∂M consists of the boundary points of the charts with boundaries. ∂M in turn is a (n - 1) dimensional manifold *without boundary.*

If M is an oriented manifold, then there is a volume form μ, defining its orientation. Since ∂M is a submanifold of M, at points on the boundary, $T_p(\partial M)$ is an $n-1$ dimensional subspace of $T_p(M)$. Let $X_1, \ldots, X_{n-1}$ be a basis for $T_p(\partial M)$. Let $Y \in T_p(M)$ be linearly independent from the $X's$. Since $X's$ are also in $T_p(M)$, we can choose Y to be such that $\mu(Y, X_1, \ldots, X_{n-1}) > 0$ and *define* $\mu_{ind}(X_1, \ldots, X_{n-1}) := \mu(Y, X_1, \ldots, X_{n-1})$. This defines a volume form on ∂M and the corresponding orientation on the boundary is called the *induced orientation* on the boundary. With these definitions, we have the *Stoke's theorem*,

$$\int_M d\eta \ = \ \int_{\partial M} \eta \ ,$$

where η is an $(n-1)$ form.

This theorem plays a role in the conservation laws discussed in the text.

16. **Mapping of Manifolds:** Let M, N be two manifolds of dimensions m, n, respectively. Let $\phi : M \to N$ be a map which is smooth i.e. locally, ϕ is represented as $(x^1, \ldots, x^m) \to (y^1, \ldots, y^n)$ such that the functions $y^\alpha(x^i)$ are smooth. Such a map allows us to push-forward and pull-back tensors defined on the two manifolds. For example, given $f : N \to \mathbb{R}$, we define its *pull-back*, $(\phi^* f)(p) := f(\phi(p))$, a function on M. Locally, $(\phi^* f)(x) = f(y = \phi(x))$. From this we define a *push-forward* of a tangent vector $X \in T_p(M)$ to a tangent vector $\phi_* X \in T_{\phi(p)} N$ as: $[\phi_* X](f)|_{\phi(p)} := X(\phi^*(f))|_p \ \forall \ p \in M$. Locally, $(\phi_* X)^\alpha(\phi(p)) = \frac{\partial y^\alpha}{\partial x^i}|_p X^i(p)$. This in turn allows us to *pull-back* co-tangent vector $\omega \in T^*_{\phi(p)}(N)$ to a co-tangent vector $(\phi^* \omega) \in T^*_p(M)$, as: $(\phi^* \omega)(X) := \omega(\phi_* X) \ \forall \ X \in T_p(M)$ and locally, $(\phi^* \omega)_i(p) := \frac{\partial y^\alpha}{\partial x^i}|_p \omega_\alpha(\phi(p))$. This generalizes to tensors of higher ranks with contravariant rank tensors pushed forward while covariant rank tensors pulled-back. Also, the corresponding ranks are preserved in these maps. Mixed rank tensors have no such natural relation.

    When ϕ is a diffeomorphism $(m = n)$, its inverse is also a diffeomorphism and using both of these we can either push forward all tensors or pull back all tensors. We can consider continuous families of diffeomorphisms and consider their infinitesimal forms. Since we can have a tensor at a point p, and a tensor at a point q, pulled back to p, we can take their difference, divide by the infinitesimal parameter and take the limit of vanishing parameter to define the *Lie derivative* of the tensor: $\mathcal{L}_X T|_p := \lim_{\epsilon \to 0} \epsilon^{-1}(\phi^*_\epsilon T - T)$. Here the one parameter family of diffeomorphisms is taken to generate a vector field X. In the next item we give the algebraic form of the definition. This notion plays a role in defining *symmetries* of specific tensor fields. Thus if a diffeomorphism ϕ leaves a metric tensor invariant i.e. $(\phi^* g)(p) = g(p)$, the diffeomorphism is said to be an isometry.

17. **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. We have used: $\langle \omega, X\rangle := \omega(X)$.

    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: $\mathcal{L}_X \omega = i_X d\omega + d(i_X \omega), \quad \forall \quad \omega \in \Lambda^k, k = 0, \cdots, n$; It follows that $d\mathcal{L}_X \omega = \mathcal{L}_X d\omega$, i.e. the Lie-derivative and the exterior derivatives commute.

18. **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) $\nabla_{fX+gY}(T) = f\nabla_X(T) + g\nabla_Y(T) \ \forall$ *functions* f, 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, $\nabla_X\langle\omega, Y\rangle = \langle\nabla_X\omega, Y\rangle + \langle\omega, \nabla_X Y\rangle$;

$\nabla_X T$ 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 (18c). 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 (18c).

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.

Changing to a different coordinate basis and using the definition of the corresponding components, the transformation law for the components of the affine connection can be deduced and it can be verified that 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: $\nabla_{\partial_i} A := A^j{}_{;i}\partial_j$.

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

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.*

19. **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,

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

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 γ.** 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 (has the same components).

Thus, by definition, *a tensor parallelly transported along X* satisfies: $X \cdot \nabla T_{||} = 0$. A non-zero covariant derivative thus measures 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 not be equal to the vector field itself, $X_{||} \nsim X$. 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 *Geodesic!non-affinely parametrized* affine geodesics.

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

20. **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: $T(f\omega, gX, hY) =$

$fghT(\omega, X, Y) \quad \forall \quad functions f, g, h$. The stipulated properties of ∇_X are crucial for this proof.

It follows 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.

For a symmetric connection,

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

21. **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:

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

It follows 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:

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

22. **Cartan Structural Equations:** The definitions associated with an affine connection imply certain *relations* which can be conveniently used as alternative definitions of the curvature and the torsion tensors relative to an arbitrary basis. To see this recall (and define) for generic bases, E_a, E^a:

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

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

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**):

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

In the previous items, 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, then 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 well known identities: the cyclic identity and the Bianchi identity by simply taking the exterior derivative of these equations.

23. **The Cyclic Identity:**

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

Specializing to coordinate bases and using the explicit definitions of wedge products, covariant derivatives etc., 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.

24. **The Bianchi Identity:**

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

In coordinate bases, 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.

25. **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.

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

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

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

26. **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 throughout 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 (32).

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

27. **The Metric Tensor:** A *symmetric tensor field 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 diagonalized 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**; otherwise it is generically called **Pseudo–Riemannian**. When the signature is $\pm(n - 2)$, the metric is said to be **Lorentzian**. In our convention, the signature is $(n - 2)$ i.e. $n_- = 1$ and $n_+ = n - 1$.

    Basic *existence results: [18]*

    (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.

28. **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 $\Psi : M \to \mathbb{R}$.

29. **Extra Operations Available Due to a Metric Tensor:** A non-degenerate metric gives us both g_{ab} and g^{ab} which sets up a *canonical* isomorphism between the tangent and the cotangent spaces which extends to tensors of higher ranks through the operations of *raising and lowering of indices.*

(a) A non-degenerate metric determines a unique affine connection through the compatibility condition, $\nabla_k g_{ij} = 0 \;\; \forall \;\; i, j, k$:

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

For the zero-torsion case, the connection is given only by the first term and is called the *Riemann–Christoffel Connection* or the *metric connection.* It is given completely in terms of the metric. 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 30.

(b) *Invariant Volume Form*:

From transformation of the metric it follows that $\sqrt{|\text{det} g_{ij}|}$ transforms as,

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

Hence, on a (pseudo-)Riemannian manifold we have a natural tensor density of weight 1, namely the $\sqrt{|\text{det} g|}$.

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

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

We also define the *Levi-Civita densities* of weights 1 and -1 respectively as,

$$\epsilon_{i_1 \dots i_n} := \sqrt{|g|}\mathcal{E}_{i_1 \dots i_n} \quad , \quad \epsilon^{i_1 \dots i_n} := \frac{1}{\sqrt{|g|}}\mathcal{E}^{i_1 \dots i_n}$$

(c) *Hodge Isomorphism:* 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).

The *Hodge Isomorphism (or Hodge * operator)* is defined as: $* : \Lambda^k \to \Lambda^{n-k}$ such that

$$\alpha \wedge (*\beta) := (\alpha, \beta)\mu_g \quad \forall \quad \alpha \in \Lambda^k$$

It follows,

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

The local expressions for components of $*\beta$:

$$(*\beta)_{i_1 \cdots i_{n-k}} = \frac{1}{k!}(-1)^{k(n-k)}\epsilon_{i_1 \cdots i_{n-k} j_1 \cdots j_k}\beta^{j_1 \cdots j_k}$$

(d) *Co-differential:* On k-form fields we defined the *exterior differential* $d : \Lambda^k \to \Lambda^{k+1}$. With a non-degenerate metric tensor available, the *Co-Differential* δ is defined as: $\delta : \Lambda^k \to \Lambda^{k-1}$,

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

On a $k-$form,

$$(\delta\omega)_{i_1, \cdots, i_{k-1}} := (-1)^k g^{ij}\nabla_i\omega_{i_1, \cdots, i_{k-1}, j} = -\ g^{ij}\nabla_i\omega_{j, i_1, \cdots, i_{k-1}}$$

where ∇ is the torsion-free, metric compatible, covariant derivative (it could be replaced by ∂_i due to anti-symmtrisation).

It follows that $\delta^2\omega = 0 \ \forall\ \omega \in \Lambda$.

ω is said to be *Co-closed* if $\delta\omega = 0$;
It is *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, $\Delta := d\delta + \delta d$. Evidently it maps k-forms to k-forms.

(e) On the space of smooth k-$form\ fields$ one defines a bilinear, symmetric, non-degenerate quadratic form:

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

For *Riemannian* manifolds without boundary,

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

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 an *orthogonal decomposition*, the *Hodge Decomposition*, for any k-form as:

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

30. **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:

$$\begin{aligned} R_{ijkl} &= -R_{ijlk} && \text{From definition ;} \\ \sum_{(jkl)} R_{ijkl} &= 0 && \text{Cyclic identity;} \\ \sum_{(klm)} R_{ijkl;m} &= 0 && \text{Bianchi identity;} \end{aligned}$$

$$\begin{aligned} (\nabla_k\nabla_l - \nabla_l\nabla_k)T^{i_1\cdots i_m}{}_{j_1\cdots j_n} &= +\sum_{\sigma=1}^{m} R^{i_\sigma}{}_{jkl}T^{i_1\cdots j\cdots i_m}{}_{j_1\cdots j_n} \\ &\quad +\sum_{\sigma=1}^{n} R^{i}{}_{j_\sigma kl}T^{i_1\cdots i_m}{}_{j_1\cdots i\cdots j_n} \end{aligned}$$

(Ricci identity)

Further Symmetry Properties:

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

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, e.g. R_{ijil} with $i \neq j, i \neq l$, and $j \neq l$ (ii) two pairs of indices are equal e.g. 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,

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

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,

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

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.

31. **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,

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

32. **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 two-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:

$$\begin{aligned} a^i &= u^j \nabla_j (u^k \nabla_k X^i) = u \cdot \nabla (X \cdot \nabla u^i) \qquad ([X,u] = 0) \qquad (14.10) \\ &= X^j u \cdot \nabla(\nabla_j u^i) + (\nabla_j u^i) u \cdot \nabla X^j \\ &= X^j u^k \nabla_k \nabla_j u^i + (\nabla_j u^i) X \cdot \nabla u^j \\ &= X^j u^k \nabla_j \nabla_k u^i - R^i{}_{kjl} u^k X^j u^l + (X \cdot \nabla u^j) \nabla_j u^i \\ &= (X \cdot \nabla)(u \cdot \nabla u^i) - R^i{}_{kjl} u^k X^j u^l \quad \text{Or,} \\ a^i &= -R^i{}_{jkl} u^j X^k u^l \qquad \text{the } \textit{Deviation Equation.} \qquad (14.11) \end{aligned}$$

14.7 Theorems on Initial Value Problem

We list here a set of theorems from [17] for the convenience of a self-contained reading.

Initial value problem is the manner in which we are accustomed to thinking from the experience in particle mechanics which is governed by ordinary differential equations. Partial differential equations in $(t, x^1, \ldots, x^n)$ introduces new features - the initial data consists of functions of x^i's which is a 'infinite amount of data'. So some regularity properties need to be stipulated on the initial data and these have to be preserved by the evolution in t. When we have equations from relativistic theories (finite speed of propagation of information), the initial data should influence only the data in its 'forward light cone' while the evolved data should have correlations induced from its past light cone only. If a solution is exist and is uniquely determined, we also need to have the *well-posedness property*: the solution (evolved data) depends *continuously* on the initial data in a suitable sense. With these in mind, we list the theorems.

Cauchy–Kowalewski theorem: This applies for partial differential equations in $n+1$ variables, $(t, x^1, \ldots, x^n)$, which are of the forms,

$$\partial_t \phi_a(t, x^i) = F_a(t, x^i \ ; \ \phi_b, \partial_t \phi_b \ ; \ \partial_i \phi_b, \partial^2_{ij} \phi_b) \quad , \quad a, b = 1, \ldots, m. \tag{14.12}$$

where the F_a's are *analytic functions* of their arguments. Let $f_a(x^j), g_a(x^j)$ be *two analytic functions*.

Theorem 14.1 (Cauchy–Kowalewski)
There exist an open neighborhood of a hypersurface Σ_0 *and a* unique, analytic *solution of the equation (14.12) such that* $\phi_a(t_0, x^j) = f_a(x^j) \ , \ \partial_t \phi_a(t_0, x^j) = g_a(x^j)$.

While existence and uniqueness for analytic data is assured by the theorem, it does not assure the well-posedness property for any choice of reasonable topology [17]. The requirement of analytic data means that initial data cannot be changed even in a small neighborhood on Σ_0 without affecting the data everywhere. So to formulate a notion of causal propagation, we need to relax the analyticity requirement. However, for smooth data, the Cauchy–Kowalewski analysis does not prove even existence.

An example where we do have local existence, uniqueness, causal propagation and well-posedness is the massive Klein–Gordon equation in Minkowski space-time: $(\partial_t^2 - \partial_i^2 + m^2)\phi = 0$. Firstly, thanks to the background Minkowski space-time, we do have the notion of relativistic causality. We also have the domains of dependences of subsets of space-like hypersurfaces. The proof of the properties uses the conservation of the stress-tensor, the existence of time translation Killing vector and the fact that the stress tensor satisfies the dominant energy condition. First it is established that if for a given smooth data on a portion $S_0 \subset \Sigma_0$ solutions exists in the domain of dependence $D^+(S_0) \cap J^-(\Sigma_1)$ where Σ_1 is another constant t-hypersurface in the future of Σ_0, then the solution is unique. The same step also establishes that a change in the initial data outside S_0, cannot affect the solution within $D^+(S_0)$. Next, using Sobolev norms, the property of well-posedness is established. Finally, using the continuity property, *existence of a smooth solution* is proved. The details crucially use the *linear* nature of the equation and the Lorentzian nature of the metric which makes it a wave equation. For a Euclidean metric we have an elliptic equation and it does not have well-posedness property.

The next generalization is to a *general hyperbolic equation* on a globally hyperbolic space-time. Let (M, g) be a globally hyperbolic space-time. A *second order, linear partial differential equation* is said to be *hyperbolic* iff it can be expressed in the form,

$$g^{\mu\nu}\nabla_\mu\nabla_\nu\phi + A^\mu\nabla_\mu\phi + B\phi + C = 0 \; . \tag{14.13}$$

This has well-posed initial value formulation for initial data $(\phi, n^\mu\nabla_\mu\phi)$ on any smooth, space-like Cauchy surface Σ. Here, ∇ is any derivative operator on the space-time, n^μ is normal to the Cauchy surface and the coefficients g, A, B, C are all smooth. While it is not possible to construct a conserved stress tensor satisfying dominant energy condition, it is possible to construct a 'stress tensor' satisfying the dominant energy condition with a non-zero but bounded divergence and this suffices to construct a proof.

This is further generalized to a *system* of linear, second order, hyperbolic equations with weaker differentiability requirements [18].

The final generalization relevant for us is a theorem of Leray.

A system of m, second order, partial differential equations for m unknowns ϕ_a, on a manifold M is said to be *quasi-linear, diagonal, second order, hyperbolic system* if it can be put in the form,

$$g^{\mu\nu}(x; \phi_b, \nabla_\alpha\phi_b)\ \nabla_\mu\nabla_\nu\phi_a \;=\; F_a(x; \phi_b, \nabla_\alpha\phi_b) \; , \tag{14.14}$$

where g is a smooth Lorentzian metric on M and F_a are *smooth* functions of its arguments.

Theorem 14.2 (Main Theorem)

Let ϕ^0_a be any solution of (14.14) and let $g^{\mu\nu}_0 := g^{\mu\nu}(x; \phi^0_a, \nabla_\alpha\phi^0_a)$. Suppose, $(M, g^{\mu\nu}_0)$ is globally hyperbolic. Let Σ be a smooth Cauchy surface for this space-time. The the initial value formulation for the equation is well-posed in the sense:

For initial data, sufficiently close to the initial data for the solution ϕ^0_a (induced on Σ), there exist an open neighborhood $O \supset \Sigma$, such that the equation system has a solution ϕ_a in O with $(O, g^{\mu\nu}(x; \phi_a, \nabla_\alpha\phi_a))$ being globally hyperbolic. The solution is unique and propagates causally i.e. if the initial data for two solutions ϕ'_a and ϕ_a coincide on $S \subset \Sigma$, then the solutions coincide *on $O \cap D^+(S)$. Finally, the solution depends continuously on the initial data in the same sense as it does for the Klein–Gordon equation* [17, 18].

The theorem guarantees existence, uniqueness, causal propagation and well-posedness *only* for initial data close (in a suitable sense) to the initial data induced by a solution *given to exist*. This is used in establishing existence, uniqueness and well-posedness property for the Einstein equation in the main text.

14.8 Petrov Classification

The non-trivial vacuum solutions have non-zero Weyl tensor. So a classification of such solutions is naturally done in terms of the *algebraic classification* of the Weyl tensor. At any point in the space-time, the Weyl tensor $C_{\mu\nu}{}^{\rho\sigma}$ can be viewd as a 6×6 matrix acting on the space of 2-forms which is six-dimensional. Any square matrix has a Jordan canonical form such that the number of Jordan blocks and their dimensions are uniquely determined. Giving a list of the dimensions of the Jordan blocks goes under the name of *Segre classification*. Special to four-dimensions is the split of the (complexified) vector space of 2-forms into self-dual and anti-self-dual subspaces which are three-dimensional and the Segre classification can be applied to the three-dimensional complex matrix representing the Weyl tensor. This results in the *Petrov classification*. There are several alternative ways of obtaining a classification which are discussed and summarized in the thesis of Carlos Batista [153]. A method which is convenient and at the same time also shows a simplification of the Weyl tensor, is based on the use of null tetrad and their behaviour under Lorentz transformations [29].

We introduced null tetrad in the context of gravitational waves while describing their helicities. Given an arbitrary choice of a null tetrad $(\ell, n, m, \bar{m})$, $\ell \cdot n = -1$, $m \cdot \bar{m} = 1$, the *Weyl Scalars* are defined as,

$$\begin{aligned}
\Psi_0 &= C_{\mu\nu\rho\sigma}\ell^\mu m^\nu \ell^\rho m^\sigma \quad , \quad \Psi_1 = C_{\mu\nu\rho\sigma}\ell^\mu n^\nu \ell^\rho m^\sigma \\
\Psi_2 &= C_{\mu\nu\rho\sigma}\ell^\mu m^\nu \bar{m}^\rho n^\sigma \quad , \quad \Psi_3 = C_{\mu\nu\rho\sigma}\ell^\mu n^\nu \bar{m}^\rho n^\sigma \\
\Psi_4 &= C_{\mu\nu\rho\sigma}n^\mu \bar{m}^\nu n^\rho \bar{m}^\sigma \quad , \quad \text{All are complex.}
\end{aligned} \tag{14.15}$$

While these are scalars under general coordinate transformations, they change under Lorentz transformations which transform one null tetrad into another one. The Lorentz transformations are grouped into three classes: (i) those which leave ℓ invariant (one complex parameter a), (ii) those which leave n invariant (one complex parameter b) and (iii) those which scale ℓ, n by λ, λ^{-1} and rotate $m, \bar{m}$ by $e^{\pm i\theta}$ respectively. The idea of the classification is to see *which and how many of these scalars can be made zero by a suitable choice of the null tetrad.*

Starting with a non-zero Ψ_4 (if necessary by making a class (i) transformation), consider a general class (ii) transformation which leaves Ψ_4 invariant. Under this,

$$\Psi_0'(b) = \Psi_0 + 4b\Psi_1 + 6b^2\Psi_2 + 4b^3\Psi_3 + b^4\Psi_4 \tag{14.16}$$

$$\Psi_1'(b) = \frac{1}{4}\frac{d\Psi_0'}{db} \quad , \quad \Psi_2'(b) = \frac{1}{3}\frac{d\Psi_1'}{db} \quad , \quad \Psi_3'(b) = \frac{1}{2}\frac{d\Psi_2'}{db} \quad . \tag{14.17}$$

The highest order polynomial in b is the Ψ_0'. It can always be set to zero by maximally 4 transformations of class (ii). If b_i is a root, then the transformed ℓ, $\ell' := b^* m + b\bar{m} + b^* b n$ is called a *Principal Null Direction* (PND) of the Weyl tensor. Clearly, there exist at least one PND (unless Weyl tensor is zero) and at the most 4 PNDs. Having found a PND, we can look for further transformations which will make additional scalars zero without affecting the previously arranged values. The classification results from the way the roots of the equation $\Psi_0'(b) = 0$, coincide or be distinct. This is discussed in detail in [29] and we will just summarize the results.

Type	Distinct PNDs and Roots	Vanishing Weyl Scalars
I	4 (Algebraically general)	$\Psi_0 = \Psi_4 = 0$
II	3 (One double root)	$\Psi_0 = \Psi_1 = \Psi_4 = 0$
D	2 (Two double roots)	$\Psi_0 = \Psi_1 = \Psi_3 = \Psi_4 = 0$
III	2 (One triple root)	$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_4 = 0$
N	1 (One quadruple root)	$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = 0$
O	0 (Conformally flat)	Weyl tensor vanishes.

As examples, we note that all known black hole solutions are of Type D, the gravitational radiation is of Type N while the FLRW space-time (non-vacuum) is of Type O. Further examples may be found in [154].

Epilogue

The primary focus of this book has been to appreciate the wide class of phenomena contained within classical general relativity. The mere extension from the Minkowskian space-time of special relativity to an arbitrary Lorentzian space-time, introduces the new feature of gravitational frequency shift and clock rates. Although tiny, these are large enough to affect the accuracy of the global positioning systems. The tiny effect of 'bending of light' is large enough to reveal gravitational lensing by galaxies. The possibility of space-time geometry being changeable, accommodates an expanding universe which is one of the key ingredients in the precipitation of different matter species and formation of structures. In all these, the specific Einstein equation provides the quantitative control on relating the geometry and the matter distribution. Although GR plays a marginal role in the properties of stable stars, it does predict the possibility of a complete gravitational collapse thereby suggesting existence of black holes (or a naked singularity). The brand new phenomenon of gravitational radiation generates a further instability in binary systems. The scales of the phenomena range from the planetary to the cosmological. Surprisingly, in the very early universe, when the length scales get closer to the Planck scale, the theory itself suggests its inadequacy. The book has discussed these aspects.

There are several other topics which are left out - some due to the space and time bounds on the book and some by choice. A topic of practical applications is the post-Newtonian formalism and its cousins. Relatively recently, effective field theory methods have been developed for application to motion of extended bodies and the gravitational radiation. These are more specialized and important computational tools and I would have liked to introduce them at the same level as numerical relativity. Another topic left out is the complete gravitational collapse including the critical phenomena first discovered by Choptuik in the spherically symmetric gravitational collapse of massless scalar field. The initial data is divided in two regions: one corresponding to collapse to a black hole and the other to a dispersal. There are universal critical exponents as the boundary is approached and the boundary solution has naked singularity. These could not be included due to the space and time constraints.

I have left out the various action formulations and the ensuing canonical forms. These have been and are important in the formal structure of GR especially as a preparation to its quantization. That is a shift from the main focus of the book which I felt is beyond the scope of the book.

This is a good place for a pause.

Bibliography

[1] A. Einstein. *The meaning of relativity.* Routledge, 2003.

[2] S. Weinberg. *Gravitation and cosmology: Principle and applications of general theory of relativity.* John Wiley and Sons, Inc., New York, 1972.

[3] R. V. Pound and G. A. Rebka. Gravitational red-shift in nuclear resonance. *Physical Review Letters*, 3:439–441, 1959.

[4] R. V. Pound and G. A. Rebka. Apparent weight of photons. *Physical Review Letters*, 4:337–341, 1960.

[5] R. V. Pound and J. L. Snider. Effect of gravity on nuclear resonance. *Physical Review Letters*, 13:539–540, 1964.

[6] F. Rohrlich. The principle of equivalence. *Annals of Physics*, 22:169–191, 1963.

[7] D. G. Boulware. Radiation from a uniformly accelerated charge. *Annals of Physics*, 124:169–188, 1980.

[8] J. A. Wheeler, K.S. Thorne, and C.W. Misner. *Gravitation.* San Francisco, 1973.

[9] L. I. Schiff. Possible new experimental test of general relativity theory. *Physical Review Letters*, 4:215–217, 1960.

[10] C. W. F. Everitt, D. B. DeBra, B. W. Parkinson, J. P. Turneaure, J. W. Conklin, M. I. Heifetz, G. M. Keiser, A. S. Silbergleit, T. Holmes, J. Kolodziejczak, et al. Gravity probe b: Final results of a space experiment to test general relativity. *Physical Review Letters*, 106:221101, 2011.

[11] W. G. Dixon. Dynamics of extended bodies in general relativity. I. momentum and angular momentum. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 314:499–527, 1970.

[12] W. G. Dixon. Dynamics of extended bodies in general relativity. II. Moments of the charge-current vector. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 319:509–547, 1970.

[13] W. G. Dixon. Dynamics of extended bodies in general relativity. III. Equations of motion. *Philosophical Transactions for the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 59–119, 1974.

[14] W. G. Dixon. The definition of multipole moments for extended bodies. *General Relativity and Gravitation*, 4:199–209, 1973.

[15] J. Ehlers and E. Rudolph. Dynamics of extended bodies in general relativity center-of-mass description and quasirigidity. *General Relativity and Gravitation*, 8:197–217, 1977.

[16] S. M. Carroll. *Spacetime and geometry. An introduction to general relativity.* Addison-Wesley, 2004.

[17] R. M. Wald. *General relativity.* University of Chicago Press, Chicago, 1984.

[18] S. W. Hawking and G. F. R. Ellis. The large scale structure of the universe. *CUP, Cambridge*, 1973.

[19] C. M. Will. The confrontation between general relativity and experiment. *Living Reviews in Relativity*, 9(3), 2006.

[20] T. Sauer. A brief history of gravitational lensing. *Einstein Online*, 4:1005, 2010.

[21] J. Wambsganss. Gravitational lensing in astronomy. *Living Reviews in Relativity*, 1(12), 1998.

[22] V. Perlick. Gravitational lensing from a spacetime perspective. *Living Reviews in Relativity*, 7(9), 2004.

[23] S. Weinberg. *Cosmology.* Oxford University Press, Oxford, 2008.

[24] A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419–421, 1965.

[25] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. Cosmic black-body radiation. *The Astrophysical Journal*, 142:414–419, 1965.

[26] A. Linde. *Inflationary cosmology.* Springer, 2007.

[27] M. Maggiore. *Gravitational waves: Volume 1: Theory and experiments.* Oxford University Press, Oxford, 2007.

[28] B. S. Sathyaprakash and B. F. Schutz. Physics, astrophysics and cosmology with gravitational waves. *Living Reviews in Relativity*, 12(2), 2009.

[29] S. Chandrasekhar. *The mathematical theory of black holes.* Clarendon Press, Oxford, 1983.

[30] R. Narayan. Black holes in astrophysics. *New Journal of Physics*, 7:199, 2005.

[31] A. Celotti, J. C. Miller, and D. W. Sciama. Astrophysical evidence for the existence of black holes. *Classical and Quantum Gravity*, 16:A3, 1999.

[32] G. Date. On a static solution of the Einstein equation with incoming and outgoing radiation. *General Relativity and Gravitation*, 29(8):953–971, 1997.

[33] R. P. Geroch. Global structure. In S. W Hawking and W. Israel, editors, *General relativity: An Einstein centenary survey.* Cambridge University Press, Cambridge 1979.

[34] P. S. Joshi. Global aspects in gravitation and cosmology. *Int. Ser. Monogr. Phys.*, 87, 1993.

[35] Y. Choquet-Bruhat, C. de Witt-Morette, and M. Dillard-Bleick. *Analysis, manifolds and physics.* North Holland Publishing Company, Amsterdam, 1978.

[36] A. Borde and A. Vilenkin. Eternal inflation and the initial singularity. *Physical Review Letters*, 72:3305–3308, 1994.

[37] S. M. Scott and P. Szekeres. The abstract boundary - a new approach to singularities of manifolds. *Journal of Geometry and Physics*, 13:223–253, 1994

[38] J. M. M. Senovilla. Singularity theorems and their consequences. *General Relativity and Gravitation*, 30:701–848, 1998.

[39] A. Ashtekar and R. O. Hansen. A unified treatment of null and spatial infinity in general relativity. i. universal structure, asymptotic symmetries, and conserved quantities at spatial infinity. *Journal of Mathematical Physics*, 19(7):1542–1566, 1978.

[40] A. Ashtekar. Asymptotic structure of the gravitational field at spatial infinity. In Alan Held, editor, *General relativity and gravitation: one hundred years after the birth of Albert Einstein.* Plenum Press, New York, 1980.

[41] A. Ashtekar and J. D. Romano. Spatial infinity as a boundary of spacetime. *Classical and Quantum Gravity*, 9(4):1069, 1992.

[42] J. Frauendiener. Conformal infinity. *Living Reviews in Relativity*, 7(1), 2004.

[43] H. Bondi, MGJ Van der Burg, and AWK Metzner. Gravitational waves in general relativity. VII. waves from axi-symmetric isolated systems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 269:21–52, 1962.

[44] R. Geroch. Asymptotic structure of space-time. In *Asymptotic Structure of Space-Time*, pages 1–105. Springer, 1977.

[45] R. Penrose. Zero rest-mass fields including gravitation: asymptotic behaviour. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 284:159–203, 1965.

[46] A. Ashtekar. *Asymptotic quantization: based on 1984 Naples lectures.* Bibliopolis, 1987.

[47] R. Sachs. Asymptotic symmetries in gravitational theory. *Physical Review*, 128(6):2851, 1962.

[48] R. Geroch and J. Winicour. Linkages in general relativity. *Journal of Mathematical Physics*, 22(4):803–812, 1981.

[49] A. Komar. Covariant conservation laws in general relativity. *Physical Review*, 113:934–936, 1959.

[50] M. T. Anderson et al. On the structure os asymptotically de Sitter and anti-de Sitter spaces. *Advances in Theoretical and Mathematical Physics*, 8:861–893, 2004.

[51] D. Christodoulou. Reversible and irreversible transformations in black-hole physics. *Physical Review Letters*, 25:1596–1597, 1970.

[52] R. Penrose. Gravitational collapse: The role of general relativity. *General Relativity and Gravitation*, 34:1141–1165, 2002.

[53] R. M. Wald. Gravitational collapse and cosmic censorship. *arXiv*, gr-qc/9710068, 1997.

[54] R. M. Wald. *Quantum field theory in curved spacetime and black hole thermodynamics.* University of Chicago Press, 1994.

[55] M. Heusler. *Black hole uniqueness theorems*, volume 6, *Cambridge Lecture Notes in Physics.* Cambridge University Press, 1996.

[56] W. Israel. Event horizons in static vacuum space-times. *Physical Review*, 164(5):1776, 1967.

[57] W. Israel. Event horizons in static electrovac space-times. *Communications in Mathematical Physics*, 8(3):245–260, 1968.

[58] B. Carter. Axisymmetric black hole has only two degrees of freedom. *Physical Review Letters*, 26(6):331–333, 1971.

[59] P. Hajicek. General theory of vacuum ergospheres. *Physical Review D*, 7:2311, 1973.

[60] B. Carter. Black hole equilibrium states. In C. De Witt and B. S. De Witt, editors, *Black holes*, pages 57–214. Gordon and Breach, 1973.

[61] S. Hawking. The event horizon. In C. de Witt and B. S. De Witt, editors, *Black holes*, pages 5–34. Gordon and Breach, New York, 1973.

[62] D. C. Robinson. Classification of black holes with electromagnetic fields. *Physical Review D*, 10:458–460, 1974.

[63] D. C. Robinson. Uniqueness of the Kerr black hole. *Physical Review Letters*, 34:905, 1975.

[64] P. O. Mazur. Proof of uniqueness of the Kerr–Newman black hole solution. *Journal of Physics A: Mathematical and General*, 15:3173, 1982.

[65] P. T. Chruciel, J. Lopes Costa, and M. Heusler. Stationary black holes: Uniqueness and beyond. *Living Reviews in Relativity*, 15(7), 2012.

[66] A. Ashtekar, C. Beetle, and S. Fairhurst. Isolated horizons: a generalization of black hole mechanics. *Classical and Quantum Gravity*, 16(2):L1, 1999.

[67] A. Ashtekar, C. Beetle, and S. Fairhurst. Mechanics of isolated horizons. *Classical and Quantum Gravity*, 17(2):253, 2000.

[68] A. Ashtekar and B. Krishnan. Dynamical horizons: energy, angular momentum, fluxes, and balance laws. *Physical Review Letters*, 89(26):261101, 2002.

[69] A. Ashtekar and B. Krishnan. Isolated and dynamical horizons and their applications. *Living Reviews in Relativity*, 7(10):2000, 2004.

[70] S. A. Hayward. General laws of black-hole dynamics. *Physical Review D*, 49:6467–6474, 1994.

[71] R. M. Wald and V. Iyer. Trapped surfaces in the schwarzschild geometry and cosmic censorship. *Physical Review D*, 44:R3719–R3722, 1991.

[72] S. A. Hayward. Dual-null dynamics of the Einstein field. *Classical and Quantum Gravity*, 10(4):779, 1993.

[73] A. Ashtekar, J Baez, A Corichi, and K. Krasnov. Quantum geometry and black hole entropy. *Physical Review Letters*, 80(5):904, 1998.

[74] L. D. Landau and E. M. Lifshitz. *The classical theory of fields: Volume 2 (course of theoretical physics series) L. D. Landau, E. M. L.* In Butterworth-Heinemann, 1980.

[75] E. K. Solutions of the Einstein equations involving functions of only one variable. *Transactions of the American Mathematical Society*, 27:155–162, 1925.

[76] C. W. Misner. Mixmaster universe. *Physical Review Letters*, 22:1071–1074, 1969.

[77] V. A. Belinskii, I. M. Khalatnikov, and EM Lifshttz. Oscillatory approach to a singular point in the relativistic cosmology. *Advances in Physics*, 19:525–573, 1970.

[78] V. A. Belinskii, I. M. Khalatnikov, and E. M. Lifshitz. A general solution of the Einstein equations with a time singularity. *Advances in Physics*, 31:639–667, 1982.

[79] V. A. Belinkski. On the cosmological singularity. *arXiv*, arXiv:1404.3864, 2014.

[80] M. Henneaux, D. Persson, and P. Spindel. Spacelike singularities and hidden symmetries of gravity. *Living Reviews in Relativity*, 11(1), 2008.

[81] A. Ashtekar, A. Henderson, and D. Sloan. Hamiltonian general relativity and the Belinskii–Khalatnikov–Lifshitz conjecture. *Classical and Quantum Gravity*, 26:052001, 2009.

[82] A. Ashtekar, A. Henderson, and D. Sloan. Hamiltonian formulation of the Belinskii-Khalatnikov-Lifshitz conjecture. *Physical Review D*, 83:084024, 2011.

[83] D. Kennefick. *Traveling at the speed of thought: Einstein and the quest for gravitational waves.* Princeton University Press, 2007.

[84] F. A. E. Pirani. Invariant formulation of gravitational radiation theory. *Physical Review*, 105:1089–1099, 1957.

[85] A. Trautman. Radiation and boundary conditions in the theory of gravitation. *Bulletin of the Academy of Polon. Sci. Cl. III*, 6:407–412, 1958.

[86] R. Sachs. Gravitational waves in general relativity. VI. the outgoing radiation condition. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 264:309–338, 1961.

[87] R. K. Sachs. Gravitational waves in general relativity. VIII. Waves in asymptotically flat space-time. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 270:103–126, 1962.

[88] B. F. Schutz. Gravitational waves notes for lectures at the Azores School on Observational Cosmology September 2011. *Observational Cosmology*, 2011.

[89] N. Dadhich and J. M. Pons. On the equivalence of the Einstein–Hilbert and the Einstein-Palatini formulations of general relativity for an arbitrary connection. *General Relativity and Gravitation*, 44:2337–2352, 2012.

[90] P. AM Dirac. *Lectures on quantum mechanics.* Yeshiva University Press, New York, 1964.

[91] R. Arnowitt, S. Deser, and C. W. Misner. Republication of The dynamics of general relativity. *General Relativity and Gravitation*, 40:1997–2027, 2008.

[92] S. A. Hojman, K. Kuchař, and C. Teitelboim. Geometrodynamics regained. *Annals of Physics*, 96:88–135, 1976.

[93] C. Teitelboim. How commutators of constraints reflect the spacetime structure. *Annals of Physics*, 79:542–557, 1973.

[94] A. Ashtekar. New variables for classical and quantum gravity. *Physical Review Letters*, 57(18):2244–2247, 1986.

[95] G. Date, R. K. Kaul, and S. Sengupta. Topological interpretation of Barbero-Immirzi parameter. *Physical Review D*, 79:044008, 2009.

[96] L. Lehner. Numerical relativity: A review. *Classical and Quantum Gravity*, 18:R25, 2001.

[97] J. Winicour. Characteristic evolution and matching. *Living Reviews in Relativity*, 15(2), 2012.

[98] F. Pretorius. Evolution of binary black-hole spacetimes. *Physical Review Letters*, 95:121101, 2005.

[99] F. Pretorius. Binary black hole coalescence. In *Physics of relativistic objects in compact binaries: From birth to coalescence*, pages 305–369. Springer, 2009.

[100] G. B. Cook. Initial data for numerical relativity. *Living Reviews in Relativity*, 3(5), 2000.

[101] S. Brandt and B. Brügmann. Black hole punctures as initial data for general relativity. *Physical Review Letters*, 78:3606, 1997 [arXiv:gr-qc/9703066].

[102] M. Shibata and T. Nakamura. Evolution of three-dimensional gravitational waves: Harmonic slicing case. *Physical Review D*, 52:5428, 1995.

[103] T. W. Baumgarte and S. L. Shapiro. Numerical integration of Einstein's field equations. *Physical Review D*, 59:024007, 1998.

[104] H. Friedrich. On the hyperbolicity of Einstein's and other gauge field equations. *Communications in Mathematical Physics*, 100:525–543, 1985.

[105] A. D. Sakharov. Vacuum quantum fluctuations in curved space and the theory of gravitation. In *Soviet Physics Doklady*, volume 12, page 1040, 1968.

[106] M. Visser. Sakharov's induced gravity: a modern perspective. *Modern Physics Letters A*, 17(15n17):977–991, 2002.

[107] K. S. Thorne, R. H. Price, and D. A. MacDonald. *Black holes: The membrane paradigm.* Yale University Press, 1986.

[108] T. Jacobson. Thermodynamics of spacetime: the Einstein equation of state. *Physical Review Letters*, 75(7):1260, 1995.

[109] T. Padmanabhan. Thermodynamical aspects of gravity: new insights. *Reports on Progress in Physics*, 73:046901, 2010.

[110] T. Padmanabhan. General relativity from a thermodynamic perspective. arXiv preprint arXiv:1312.3253, 2013.

[111] D. Oriti. *Approaches to quantum gravity: Toward a new understanding of space, time and matter.* Cambridge University Press, 2009.

[112] R. B. Laughlin. Emergent relativity. *International Journal of Modern Physics A*, 18:831–853, 2003.

[113] R. D. Sorkin. Causal sets: Discrete gravity. In *Lectures on Quantum Gravity*, pages 305–327. Springer, 2005.

[114] F. Dowker. Causal sets as discrete spacetime. *Contemporary Physics*, 47(1):1–9, 2006.

[115] J. Henson. The causal set approach to quantum gravity. In D. Oriti, editor, *Approaches to quantum gravity: Towards a new understanding of space, time and matter*, pages 393–413. Cambridge University Press, 2009.

[116] Y. Nambu. Quark model and the factorization of the Veneziano model. In R. Chand, editor, *Proceedings of the International Conference on Symmetries and Quark Models*. Gordon and Breach, NY, 1970.

[117] T. Eguchi and K. Nishijima. *Broken symmetry: Selected papers of Y Nambu*, volume 13. World Scientific, 1995.

[118] H. Nielsen. An almost physical interpretation of the integrand of the n-point Veneziano model. In *15th International Conference on High Energy Physics*, 1970.

[119] D. B. Fairlie and H. B. Nielsen. An analogue model for KSV theory. *Nuclear Physics B*, 20(3):637–651, 1970.

[120] L. Susskind. Harmonic-oscillator analogy for the Veneziano model. *Physical Review Letters*, 23:545–547, 1969.

[121] T. Yoneya. Quantum gravity and the zero-slope limit of the generalized Virasoro model. *Lettere al Nuovo Cimento*, 8:951–955, 1973.

[122] J. Scherk and J. H. Schwarz. Dual models for non-hadrons. *Nuclear Physics B*, 81(1):118–144, 1974.

[123] M. B. Green and J. H. Schwarz. Anomaly cancellations in supersymmetric d = 10 gauge theory and superstring theory. *Physics Letters B*, 149(1):117–122, 1984.

[124] D. J. Gross and V. Periwal. String perturbation theory diverges. *Physical Review Letters*, 60:2105–2108, 1988.

[125] A. Sen. Unification of string dualities. *Nuclear Physics B-Proceedings Supplements*, 58:5–19, 1997.

[126] C. Vafa. Lectures on strings and dualities. *arXiv preprint hep-th/9702201*, 1997.

[127] J. Polchinski. Tasi lectures on d-branes. *arXiv preprint hep-th/9611050*, 1996.

[128] S. D. Mathur. The fuzzball proposal for black holes: An elementary review. *Fortschritte der Physik*, 53:793–827, 2005.

[129] S. Mukhi. String theory: A perspective over the last 25 years. *Classical and Quantum Gravity*, 28:153001, 2011.

[130] C. Rovelli. *Quantum gravity*. Cambridge University Press, 2004.

[131] A. Perez. The spin-foam approach to quantum gravity. *Living Reviews in Relativity*, 16(3), 2013.

[132] A. Ashtekar, J. Lewandowski, D. Marolf, J. Mourao, and T. Thiemann. Quantization of diffeomorphism invariant theories of connections with local degrees of freedom. *Journal of Mathematical Physics*, 36:6456–6493, 1995.

[133] J Fernando and G. Barbero. Real Ashtekar variables for Lorentzian signature space-times. *Physical Review D*, 51:5507–5510, 1995.

[134] G. Immirzi. Real and complex connections for canonical gravity. *Classical and Quantum Gravity*, 14:L177, 1997.

[135] C. Rovelli and L. Smolin. Loop space representation of quantum general relativity. *Nuclear Physics B*, 331:80–152, 1990.

[136] A. Ashtekar and J. Lewandowski. Representation theory of analytic holonomy c*-algebras. In *Knots and Quantum Gravity*, volume 1, page 21, 1994 [arXiv:gr-qc/9311010].

[137] J. Lewandowski, A. Okołów, H. Sahlmann, and T. Thiemann. Uniqueness of diffeomorphism invariant states on holonomy–flux algebras. *Communications in Mathematical Physics*, 267:703–733, 2006.

[138] A. Ashtekar, M. Bojowald, and J. Lewandowski. Mathematical structure of loop quantum cosmology. *Advances in Theoretical and Mathematical Physics*, 7:233–268, 2003.

[139] A. Ashtekar and P. Singh. Loop quantum cosmology: A status report. *Classical and Quantum Gravity*, 28:213001, 2011.

[140] A. Ashtekar and J. Lewandowski. Quantum theory of geometry: I. Area operators. *Classical and Quantum Gravity*, 14:A55, 1997.

[141] A. Ashtekar, J. C. Baez, and K. Krasnov. Quantum geometry of isolated horizons and black hole entropy. In *Advances in Theoretical and Mathematical Physics*, 1999 [arXiv:gr-qc/0005126].

[142] A. Ashtekar and J. Lewandowski. Quantum theory of geometry ii: Volume operators. Arxiv preprint gr-qc/9711031, 1997.

[143] T. Thiemann. A length operator for canonical quantum gravity. *Journal of Mathematical Physics*, 39:3372–3392, 1998.

[144] A. Ashtekar, A. Corichi, and J. A. Zapata. Quantum theory of geometry: Non-commutativity of Riemannian structures. *Classical and Quantum Gravity*, 15:2955, 1998.

[145] A. Ashtekar and J. Lewandowski. Background independent quantum gravity: A status report. *Classical and Quantum Gravity*, 21:R53, 2004.

[146] T. Thiemann. *Modern canonical quantum general relativity*. Cambridge University Press, 2007.

[147] J. Ambjørn, J. Jurkiewicz, and R. Loll. Nonperturbative Lorentzian path integral for gravity. *Physical Review Letters*, 85:924–927, 2000.

[148] J. Ambjørn, J. Jurkiewicz, and R. Loll. The spectral dimension of the universe is scale dependent. *Physical Review Letters*, 95:171301, 2005.

[149] S. Kobayashi and K. Nomizu. *Foundations of differential geometry*, 2 volumes. Interscience, New York, 1963.

[150] R. Abraham, J. E. Marsden, and T. S. Ratiu. *Manifolds, tensor analysis, and applications*, volume 75. Springer, 1988.

[151] S. Helgason. *Differential geometry and symmetric spaces*, volume 12. American Mathematical Soc., 1964.

[152] G. F. Simmons and J. K. Hammitt. *Introduction to topology and modern analysis.* McGraw-Hill, New York, 1963.

[153] C. Batista. On the pursuit of generalizations for the Petrov classification and the Goldberg-Sachs theorem. arXiv, arXiv:1311.7110, 2013.

[154] H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers, and E. Herlt. *Exact solutions of Einstein's field equations.* Cambridge University Press, 2003.

Physics

General Relativity: Basics and Beyond familiarizes you with the basic features of the theory of general relativity as well as some of its more advanced aspects. Employing the pedagogical style of a textbook, it includes essential ideas and just enough background material needed for you to appreciate the issues and current research.

The first five chapters form the core of an introductory course on general relativity. The author traces Einstein's arguments and presents examples of space-times corresponding to different types of gravitational fields. He discusses the adaptation of dynamics in a Riemannian geometry framework, the Einstein equation and its elementary properties, and different phenomena predicted or influenced by general relativity.

Moving on to more sophisticated features of general relativity, the book presents the physical requirements of a well-defined deterministic framework for non-gravitational dynamics and describes the characterization of asymptotic space-times. After covering black holes, gravitational waves, and cosmological space-times, the book examines the evolutionary interpretation for the class of globally hyperbolic space-times, explores numerical relativity, and discusses approaches that address the challenges of general relativity.

Features

- Emphasizes the physical ideas and motivations of the theory while refraining from exhaustive details
- Covers the mathematical aspects that are important in understanding the scope and limitations of the theory, such as the mathematical model for space-time and basic physical quantities related to space-time measurements
- Explores current research topics, including the quasi-local generalization of black holes, the challenge of directly detecting gravitational waves, and the nature of cosmological singularities
- Discusses an emergent gravity viewpoint and the main approaches to a quantum theory of gravity
- Provides the necessary background information, including differential geometry

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K15233
ISBN: 978-1-4665-5271-5
90000
9 781466 552715
www.crcpress.com